Anomaly Detection in Hyperspectral Imagery Based on Kernel ICA Feature Extraction

Feng Mei[1], Chunhui Zhao[1], Liguo Wang[1] and Hanjun Huo[2]
[1]College of Information and Communication Engineering,
Harbin Engineering University, Harbin, Heilongjiang, 150001, China
[2]College of Electrical & Information Engineering,
Heilongjiang Institute of Science and Technology, Harbin, Heilongjiang, 150027, China
meifeng12@163.com, zhaochunhui@hrbeu.edu.cn, wangliguo@hrbeu.edu.cn,
huohanjun@yahoo.com.cn

Abstract

A kernel-based independent component analysis algorithm, which combines kernel principal component analysis (KPCA) and independent component analysis (ICA) is proposed for anomaly detection in hyperspectral imagery. The conventional RX anomaly detector suffers from high false alarm rates and low probability of detection. In this paper, KPCA is performed on a feature space to whiten data and fully mine the nonlinear information between spectral bands. Then, ICA seeks the projection directions in the KPCA whitened space for making the distribution of the projected data mutually independent. Finally, RX detector is performed on the projected data to locate the anomaly targets. The kernel ICA algorithm extracts the nonlinear independent components along with the dimensional reduction, and improves the performance of RX detector in hyperspectral data. Numerical experiments are conducted on real hyperspectral imagery collocted by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). Using receiver operating characteristic (ROC) curves, the results show the improved performance and reduction in the false-alarm rate.

1. Introduction

Recently, the advances in hyperspectral sensors with high spectral and spatial resolution have led to an increased interest in exploiting spectral imagery for anomoly and target detection. Given the availability of spectral libraries for a wide range of materials, detection algorithms that exploit a known target signature have been widely investigated. But the performance of known signature detectors degenerate due to the difficulties come from complete spectral libraries establishment, accurate spectral calibration, and the availability of reliable atmospheric data to convert reflectance values to the radiance spectra [1]. Such difficulties can be overcome by using an anomaly detector [1-5]. In spectral anomaly detection algorithms, materials that have a significantly different spectral signature from their surrounding background clutter pixels are identified as spectral anomalies, and the spectral signatures are used to detect anomalies embedded within the background clutter with a very low signal-to-noise ratio (SNR) [3]. In spectral anomaly detectors, no prior knowledge of the target spectral signature is utilized or assumed.

A widely used anomaly detector for multispectral data is the RX algorithm presented in [2], [5]. Extending the RX algorithm to hyperspectral imagery suffers from two major limitations. Firstly, in many environments, the local normal model provides an inadequate representation of the underlying distribution [1], [6]. Secondly, the RX algorithm is computationally expensive, since the RX detector requires evaluating the inverse of the sample covariance matrix of the hyperspectral data.

Based on the conventional RX algorithm, Kwon et al. proposed a nonlinear RX algorithm that is called kernel RX [3]. By the kernel method, this algorithm adequately mines the high-order correlation between spectral bands and shows good nonlinear performance for anomaly detection. However, this algorithm is greatly time consuming. Yanfeng Gu et al. proposed a High-Order statistics based selective kernel principal component analysis (SKPCA) algorithm for anomaly detection [7]. The algorithm mined the high-order correlation between spectral bands, but the extracted features were only irrelevant in feature space, and the background was not suppressed sufficiently.

To reduce the dimensionality of hyperspectral imagery data for RX detector, and exploit the nonlinear information lying in the massive amounts of hyperspectral data simultaneously, in this paper, we propose a kernel-based Independent Component Analysis (or KICA) for anomaly detection in hyperspectral imagery. In this paper,

978-0-7695-3497-8/08 $25.00 © 2008 IEEE
DOI 10.1109/IITA.2008.98

the original hyperspectral data was nonlinearly mapped into a feature space (which could be potentially of infinite dimensionality) consisting of possibly the original spectral bands and many nonlinear combinations of the spectral bands of the original spectral signature. Then we can perform ICA in the feature space and obtain nonlinear independent componments in the input space, correspondingly. F. R. Bach introduced the conception of Kernel Independent Component Analysis [8] in 2002 and L. Bai et al. applied it for spectral recognition of celestial objects [9] in 2006, but they did not bring forward the two-steps analysis flow for KICA explicitly, and the the kernel function in the algorithm is only for reforming the canonical correlation derived from polynomial kernel function.

This paper is organized as follows. Section 2 porvides an introduction to the conventional ICA method. In Section 3 we show the derivation of kernal ICA for hyperspectral imagery dimensional reduction. Section 4 contains a brief introduction to RX-detector for anomaly detection in the extracted features. Experimental results and analysis are provided in Section 5. Finally, concluding remarks are given in Section 6.

2. Independent Component Analysis

Independent Component Analysis is proposed for Blind Source Separation (BSS) by Herault and Jutten in 1988 [10]. It is widely used in fields like voice and medical signal analysis and image processing.

With the assumption of the existence of M independent source signals $\mathbf{s} = [s_1, s_2, \ldots, s_M]^T$, and the observations of J mixtures $\mathbf{x} = [x_1, x_2, \ldots, x_J]^T$, the classical model of ICA is

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \tag{1}$$

where $\mathbf{A}$ is a $J \times M$ mixing matrix. The purpose of the ICA is to find a demixing matrix $\mathbf{W}$ to separate the signal source vector into a set of sources which are statistically independent. So the independent components can be estimated as

$$\mathbf{y} = \mathbf{W}\mathbf{x}. \tag{2}$$

where $\mathbf{y}$ is the estimation of $\mathbf{s}$. It can be seen from (1) and (2) that $\mathbf{y}$ is the linear combination of $\mathbf{s}$,

$$\mathbf{y} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s} = \mathbf{V}\mathbf{s}, \tag{3}$$

where $\mathbf{V} = \mathbf{W}\mathbf{A}$.

3. Kernel Independent Component Analysis

To fully exploit highorder correlation between spectral bands, the input data is mapped into an implicit feature space $\boldsymbol{F}$ via a nonlinear map function $\boldsymbol{\Phi}$,

$$\mathbf{x} \in \boldsymbol{R}^{\boldsymbol{J}} \xrightarrow{\Phi} \Phi(\mathbf{x}) \in \boldsymbol{F}, \tag{4}$$

where $\mathbf{x}$ is a random vector in the input sapce $\boldsymbol{R}^{\boldsymbol{J}}$, and $\Phi(\mathbf{x})$ is its image in the feature space $\boldsymbol{F}$. The feature space $\boldsymbol{F}$ potentially can be of much higher (could be infinite) dimensionality to its conterpart $\boldsymbol{R}^{\boldsymbol{J}}$.

Now we need to find a demixing matrix $\mathbf{W}_\Phi$ in the feature space to reconstruct the nonlinear independent components

$$\hat{\mathbf{s}} = \mathbf{W}_\Phi \Phi(\mathbf{x}). \tag{5}$$

For the classical ICA, to make the problem of ICA estimation simpler and better conditioned, a useful preprocessing strategy is to first whiten the observed variables utilizing PCA transform. Similarly, we can perform KPCA in the feature space $\boldsymbol{F}$ for data whitening to exploit highorder correlation between spectral bands. And then, ICA is performed on the whitened data in the feature space to extract the nonlinear independent components. Based on this idea, we will develop a kernel ICA algorithm for anomaly detection in hyperspectral imagery.

3.1. Whitening data in feature space by KPCA

Let each input spectral signal consisting of J spectral bands be denoted by $\mathbf{x} = [x_1, x_2, \ldots, x_J]^T$. Define $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_1, \ldots, \mathbf{x}_N]$ of N observation vectors to be the observation matrix, and the image of $\mathbf{X}$ via the nonlinear mapping function $\boldsymbol{\Phi}$ in the feature space is defined as $\mathbf{X}_\Phi = [\Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \ldots, \Phi(\mathbf{x}_N)]$. we assume that we are dealing with centered data, $\sum_{j=1}^{N} \Phi(\mathbf{x}_j) = 0$, the covariance matrix of $\mathbf{X}_\Phi$ in $\boldsymbol{F}$ takes the form

$$\hat{\mathbf{C}}_\Phi = \frac{1}{N} \mathbf{X}_\Phi \mathbf{X}_\Phi{}^T. \tag{6}$$

$\hat{\mathbf{C}}_\Phi$ is a real symmetric matrix and its eigenvector decomposition or spectral decomposition as given by

$$\hat{\mathbf{C}}_\Phi = \mathbf{V}_\Phi \boldsymbol{\Lambda}_\Phi \mathbf{V}_\Phi^T, \tag{7}$$

where $\Lambda_\Phi = diag([\lambda_{\Phi 1}, \lambda_{\Phi 2}, \ldots, \lambda_{\Phi N}])$ is a diagonal matrix consisting of the nonzero eigenvalues and $\mathbf{V}_\Phi = [\mathbf{v}_1^\Phi, \mathbf{v}_2^\Phi, \ldots, \mathbf{v}_N^\Phi]$ is a matrix whose columns are the eigenvectors of $\hat{\mathbf{C}}_\Phi$ in the feature space, correspondingly.

Due to the high dimensionality of the feature space , it is computationally infeasible to directly implement any algorithm in the feature space. However, the embarrassment can be avoided by utilizing kernel-based learning algorithms which use an effective kernel trick to implement dot products in feature space by employing kernel functions [11]. The kernel representation for the dot products in the feature space $\boldsymbol{F}$ is expressed as

$$k(\mathbf{x}_i, \mathbf{x}_j) = <\Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j)> = \Phi(\mathbf{x}_i)\cdot\Phi(\mathbf{x}_j)\,. \quad (8)$$

Using the kernel trick (8), it allows us to implicitly compute the dot products in $\boldsymbol{F}$ without mapping the input vectors into it. We define the Gram matrix

$$\mathbf{K} = \mathbf{X}_\Phi^T \mathbf{X}_\Phi\,, \quad (9)$$

its elements can be determined by virtue of the given kernel function, i.e.,

$$(\mathbf{K})_{ij} = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j) = k(\mathbf{x}_i, \mathbf{x}_j)\,. \quad (10)$$

As $\mathbf{K}$ is a real symmetric matrix and its eigenvector decomposition as given by

$$\mathbf{K} = \mathbf{\Omega}\mathbf{\Lambda}_K \mathbf{\Omega}^T\,, \quad (11)$$

where $\mathbf{\Omega} = [\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_N]$ is a matrix whose columns are the eigenvectors of $\mathbf{K}$ and $\Lambda_K = diag([\lambda_{K1}, \lambda_{K2}, \ldots, \lambda_{KN}])$ is is a diagonal matrix consisting of the nonzero eigenvalues, correspondingly.

For (7) and (11), from the derivation in [11], we can find these conclusions that $\hat{\mathbf{C}}_\Phi$ and $\mathbf{K}$ are of the same eigenstructures, i.e.,

$$\mathbf{\Lambda}_K = N\mathbf{\Lambda}_\Phi\,, \quad (12)$$

$$\mathbf{V}_\Phi = \mathbf{X}_\Phi \mathbf{\Omega}\,. \quad (13)$$

Let $\mathbf{x}$ be a test point, with an image $\Phi(\mathbf{x})$ in $\boldsymbol{F}$. We need compute projections onto the eigenvectors $\mathbf{v}_k^\Phi$ in $\boldsymbol{F}$,

$$(\mathbf{v}_k^\Phi \cdot \Phi(\mathbf{x})) = \sum_{i=1}^{N} \omega_{ik} (\Phi(\mathbf{x}_i)\cdot\Phi(\mathbf{x}))\,, \quad (14)$$

where $(k = 1, 2, \cdots, N)$. We define the sphering operator $\mathbf{P}^T = \mathbf{\Lambda}_\Phi^{-\frac{1}{2}} \mathbf{V}_\Phi^T$, and vector $\mathbf{x}$ after sphering transformation as $\mathbf{z}$, then

$$\begin{aligned}\mathbf{z} &= \mathbf{P}^T \Phi(\mathbf{x}) \\ &= \sqrt{N}\mathbf{\Lambda}_K^{-1}\mathbf{\Omega}^T [k(\mathbf{x}_1, \mathbf{x}), k(\mathbf{x}_2, \mathbf{x}), \ldots, k(\mathbf{x}_N, \mathbf{x})]^T \\ &= \sqrt{N}\mathbf{\Lambda}_K^{-1}\mathbf{\Omega}^T \mathbf{k}_x\end{aligned} \quad (15)$$

Kernel method allows us to implicitly compute the dot products in $\boldsymbol{F}$ without mapping the input vectors into it. The mapping function does not need to be identified, and we never need the mapped patterns explicitly, too. What we need to do is find a appropriate kernel function. In a certain sense, the particular kernel used then implicitly determines the mapping function Φ and the space $\boldsymbol{F}$ of all possible features [11].

Now, let us go back to the problem of centering data. we assumed that the data were centered in the feature space; however, we cannot center the data in the high-dimensional feature space because we do not have any knowledge about the nonlinear mapping Φ. Fortunately, a slight modification of the following process can achieve this. Centered Gram matrix $\widetilde{\mathbf{K}}$ and $\widetilde{\mathbf{k}}_x$ can be represented by $\mathbf{K}$ and $\mathbf{k}_x$ as following:

$$\widetilde{\mathbf{K}} = \mathbf{K} - \mathbf{K}\mathbf{I}_N - \mathbf{I}_N\mathbf{K} + \mathbf{I}_N\mathbf{K}\mathbf{I}_N\,, \quad (16)$$

$$\widetilde{\mathbf{k}}_x = \mathbf{k}_x - \mathbf{K}\mathbf{i}_N - \mathbf{I}_N\mathbf{k}_x + \mathbf{I}_N\mathbf{K}\mathbf{i}_N\,, \quad (17)$$

where $(\mathbf{I}_N)_{ij} := 1/N$ is a $N \times N$ matrix, and $\mathbf{i}_N$ is the $N \times 1$ column vector with all entries equal to $1/N$.

3.2. Independend component extraction in feature space

After whitening, the task of KICA is to find the separating matrix $\mathbf{W}$ in the KPCA-transformed feature space to recover the nonlinear independent components $\hat{\mathbf{s}}$ from $\mathbf{z}$ defined in (15).

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{z}\,, \quad (18)$$

Intuitively speaking, the key to estimating the ICA model is nongaussianity. Among those different criteria proposed to measure source nongaussianity (or independency), negentropy is well justified by statistical theory. With simpler approximations of negentropy, Hyvarinen A et al. proposed a FastICA [12]. In contrast to many ICA basing on other nongaussianity measurement, the convergence of FastICA is cubic, and the independent components can be estimated one by one. Applying FastICA in the KPCA-transformed feature space, we can obtain the nonlinear independent componments finally.

4. RX Anomaly Detector

The RX anomaly detector introduced by Reed and Yu [2] has become the benchmark for hyperspectral anomaly detection because of its natural assumption that neither the target spectrum nor the covariance matrix of the background clutter need to be known. It is based on comparing the difference between the test spectrum and the spectra of the immediate background samples. It is similar to the Mahalanobis distance measure and is given by

$$\mathrm{RX}(\mathbf{r}) = (\mathbf{r} - \hat{\boldsymbol{\mu}}_b)^T \hat{\mathbf{C}}_b^{-1} (\mathbf{r} - \hat{\boldsymbol{\mu}}_b) \underset{\mathrm{H}_0}{\overset{\mathrm{H}_1}{\gtrless}} \eta\,, \quad (19)$$

where $\mathbf{r}$ is a test sample, $\hat{\boldsymbol{\mu}}_b$ is the estimated background clutter sample mean, $\hat{\mathbf{C}}_b$ is the background covariance matrix estimated from the reference background clutter data, and η is a threshold of the test. If $\mathrm{RX}(\mathbf{r}) \geq \eta$, then the assumption with the target present is valid; otherwise, the assumption with the target absent is adopted.

5. Experiments and result analysis

In this section, the proposed algorithm was compared with other feature extraction methods (ICA and SKPCA). Moreover, they were evaluated by virtue of anomaly detection. Real hyperspectral imagery (HIS) collected by the AVIRIS was used in the experiments. Due to water absorption and low signal-to-noise ratio (SNR), only 126 of those bands in the wavelength range from 0.4 to 1.8 μm are actually used here. There are 38 panels of anomalous targets in the image scene. The first band image of the data and the truth case of anomalous targets distribution are shown in Figure 1. (a) and (b), respectively.

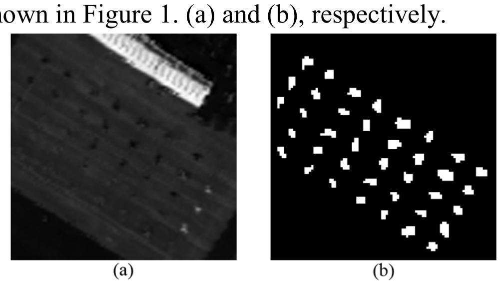

Figure 1. First band image and real targets distribution.

In this paper, we use a Gaussian Radial Basis Function (RBF) kernel which takes the form

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(\frac{-\|\mathbf{x}-\mathbf{y}\|^2}{c}\right), \tag{20}$$

where $c > 0$ is a critical kernel parameter representing the width of the Gaussian kernel. This parameter must be chosen properly so that the RBF function can full exploit the data variations. In this paper, the value of c was experimentally determined and set to 68.

For SKPCA, as describeed in [7], it is performed on the original hyperspectral data, and then, the average local singularity (LS) is defined based on the high-order statistics in a local sliding window, which is used as a measure for selecting the most informative nonlinear component for anomaly detection. Finally, the selected component with maximum average LS is used as input for anomaly detectors. ICA (FastICA method adopted here) just extracts the independent components from the original hyperspectral data for RX detector.

For convenience of comparison, we select the first nine components transformed on the original hyperspectral data by ICA, SKPCA and KICA, respectively. And then, the RX detector is applied on these transformed data respectively. We use the local dual-window [3] for the local covariance estimations for the conventional RX-algorithms. The sizes of the inner window and outer window used for the local covariance matrix estimations are 3×3 and 11×11 pixel areas, respectively.

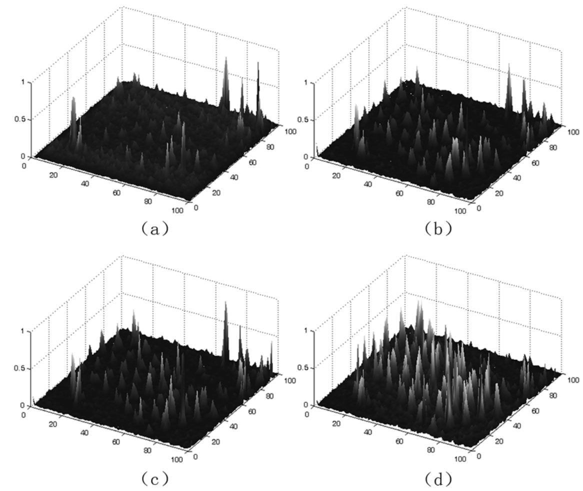

Figure 2. 3-D plot of detection results using (a) RX on original HSI, and RX after feature extraction by (b) ICA,(c) SKPCA,(d) KICA, respectively

The detection results using RX detector after feature extraction by ICA, SKPCA and KICA are shown in Figure 2. (b), (c), and (d). The detection result using RX directly on HSI is shown in Figure 2. (a). From the result comparisions, we get that the performance of KICA is the best of the four algorithms. Because SKPCA and KICA exploit highorder correlation between spectral bands, the background in Figure 2. (c) and (d) are supressed more sufficiently than in Figure 2. (b). And due to mutually independence among the components extracted by KICA, the anomaly information in Figure 2. (d) is concentrated better than SKPCA in Figure 2. (c). Because of the inability in processing efficiently the massive amounts of original hyperspectral data and exploiting highorder correlation between spectral bands, the performance of the conventional RX detector applying to the original hyperspectral data is the worst of the four in Figure 2. (a).

The receiver operating characteristics (ROC) curves representing detection probability P_d versus false-alarm rates P_f were also generated to provide quantitative performance comparison. The ROC curves obtained by RX, ICA, SKPCA and KICA for the AVIRIS data are shown in Figure 3. It can be found that the detection performance of KICA and SKPCA are better than ICA, and RX is the worst. Note the KICA significantly outperformed the SKPCA and ICA at lower false-alarm rates. The detection probability of SKPCA is closer to

KICA as long as the false-alarm rate increases. It means that the nonlinear mapping the original data into the high dimensinal feature space and extracting the highorder correlation between spectral bands significantly improved the performance of the conventional RX detector.

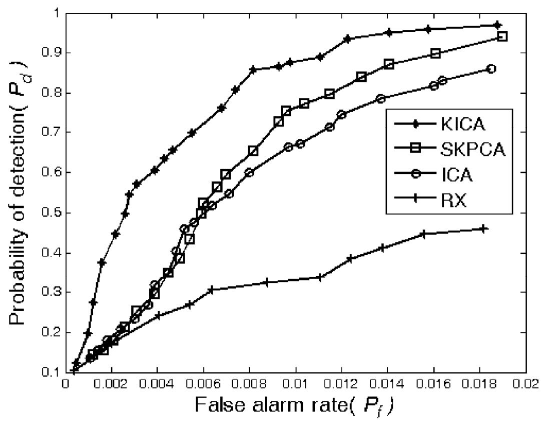

Figure 3. ROC curves obtained by KICA, SKPCA, ICA and RX

6. Conclusions

In this paper, a kernel independent component analysis algorithm has been proposed as an nonlinear independent components extraction method for anomaly detection in hyperspectral imagery. The high-order correlation was exploited via mapping the original hyperspectral data into a higher dimensional feature space, and nonlinear independent components are extracted by ICA performed on the whitened data in the feature space. The experimental results prove that the KICA significantly improves the performance of the conventional RX detector.

Acknowledgement

This work was supported in part by the the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20060217021), and by the Natural Science Foundation of Heilongjiang Province, China (Grant No. ZJG0606-01).

References

[1] A. Banerjee, P. Burlina and C. Diehl, "A support vector method for anomaly detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.,* vol. 44, pp. 2282-2291, 2006.

[2] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. 38, pp. 1760-70, 10. 1990.

[3] H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.,* vol. 43, pp. 388-397, 2005.

[4] L. He, Q. Pan, W. Di and Y. Li, "Anomaly detection in hyperspectral imagery based on maximum entropy and nonparametric estimation," *Pattern Recog. Lett.,* vol. 29, pp. 1392-1403, 2008.

[5] X. Yu, L. E. Hoff, I. S. Reed, An Mei Chen and L. B. Stotts, "Automatic target detection and recognition in multiband imagery: a unified ML detection and estimation approach," *IEEE Trans. Image Process.,* vol. 6, pp. 143-56, 01. 1997.

[6] 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 Process. Mag., vol. 19, pp. 58-69, 01. 2002.

[7] Y. Gu, Y. Liu and Y. Zhang, "A selective KPCA algorithm based on high-order statistics for anomaly detection in hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters,* vol. 5, pp. 43-47, 2008.

[8] F. R. Bach and M. I. Jordan, "Kernel independent component analysis," *Journal of Machine Learning Research,* vol. 3, pp. 1-48, 01/01. 2003.

[9] L. Bai, A. Xu, P. Guo and Y. Jia, "Kernel ICA feature extraction for spectral recognition of celestial objects," in *2006 IEEE Conference on Systems, Man, and Cybernetics,* 2006, pp. 3922-6.

[10] C. Jutten and J. Herault, "Independent components analysis (INCA) versus principal components analysis," in *Fourth European Signal Processing Conference,* 1988, pp. 643-6.

[11] B. Scholkopf, A. Smola and K. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.,* vol. 10, pp. 1299, 1998.

[12] A. Hyvarinen and E. Oja, "Independent component analysis: algorithms and applications," Neural Networks, vol. 13, pp. 411-30, 05. 2000.