ON THE USE OF ICA FOR HYPERSPECTRAL IMAGE ANALYSIS

A. Villa[⋆,⋄], *J. Chanussot*[⋆], *C. Jutten*[⋆], *J. A. Benediktsson*[⋄] *and S. Moussaoui*[†]

[⋆]GIPSA-lab, Signal & Image Dept., Grenoble Institute of Technology - INPG, France.
[⋄]Dept. of Electrical and Computer Engineering, University of Iceland, Iceland.
[†]Institut de Recherche en Communications et Cybernetique de Nantes, France

ABSTRACT

Independent component analysis (ICA) is a very popular method that has shown success in blind source separation, feature extraction and unsupervised recognition. In recent years ICA has been largely studied by researchers from the signal processing community. This paper addresses a more in-depth study on the use of this method, applied to hyperspectral images used for remote sensing purposes. In a first part, source separation is addressed. Since the independence of sources is usually not verified in hyperspectral real data images, ICA, if used alone, is not a suitable tool to unmix sources. We propose a hierarchical approximation for the use of ICA as a pre-processing step for a Bayesian Positive Source Separation method. In a second part, the use of ICA for dimensionality reduction is studied in the frame of hyperspectral data classification. Experimental results show the effectiveness of ICA when used for hyperspectral image pre-processing for the two considered applications.

I. INTRODUCTION

Hyperspectral images are composed of hundreds of bands with a very high spectral resolution, from the visible to the infra-red region. The wide spectral range, coupled with an always increasing spatial resolution, allows to better characterize materials and gives the ability to pinpoint chemical species laying on the observed surface, making hyperspectral imagery suitable for source separation and classification process. On the other side, the huge amount of high dimensional data increases the computational load and can also degrade the results of the classification process, due to the curse of dimensionality [1]. Many techniques have been proposed in the last years to overcome these problems.

Recently, Independent Component Analysis (ICA) has received attention because of its wide range of potential applications [2]–[8]. The goal of ICA is to recover independent sources given only sensor observations that are unknown linear mixtures of the unobserved independent source signals. In contrast to correlation-based transformations such as Principal Component Analysis (PCA), ICA not only decorrelates the signals (2nd-order statistics) but also reduces higher-order statistical dependencies, attempting to make the signals as independent as possible. In the recent past, ICA has been proposed as a tool to unmix hyperspectral data [5]. ICA allows each source to be automatically extracted from the observation of linear combinations of these sources. To correctly retrieve the sources, ICA needs the assumption of their statistical independence. This assumption is unfortunately usually not verified in the case of real data. This makes ICA efficient for removing artifact but not suitable for segmentation of hyperspectral images [6].

ICA can also be used as an alternative approach to Principal Component Analysis for dimensionality reduction. In order to obtain the generating factors, ICA is designed not to search the principal components, which allows to represent the maximum of the return dispersion, but the more independent factors which can linearly generate the returns. The algorithm includes higher (than second order) statistics, thus it seems to be an attractive candidate for dimensionality reduction. ICA has proven good performances with respect to PCA in various fields, such as object recognition and geoscience applications. ICA has also been applied to hyperspectral data for dimensionality reduction for source separation and data compression [7].

In this paper, we propose a more in-depth study of ICA for hyperspectral image analysis, in the case of source separation and classification application.

II. INDEPENDENT COMPONENT ANALYSIS

ICA consists in finding a linear decomposition of observed data into statistically independent components. Given an observation model:

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

where $\mathbf{x}$ is the vector of the observed signals, $\mathbf{A}$ a scalar matrix of the mixing coefficients and $\mathbf{s}$ the vector of the source signals, ICA finds a separating matrix $\mathbf{W}$ such that:

$$\mathbf{y} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s}, \tag{2}$$

where $\mathbf{y}$ is a vector of independent components. ICA looks for a linear representation that maximizes a non-gaussianity measure, or minimizing an objective function. A commonly

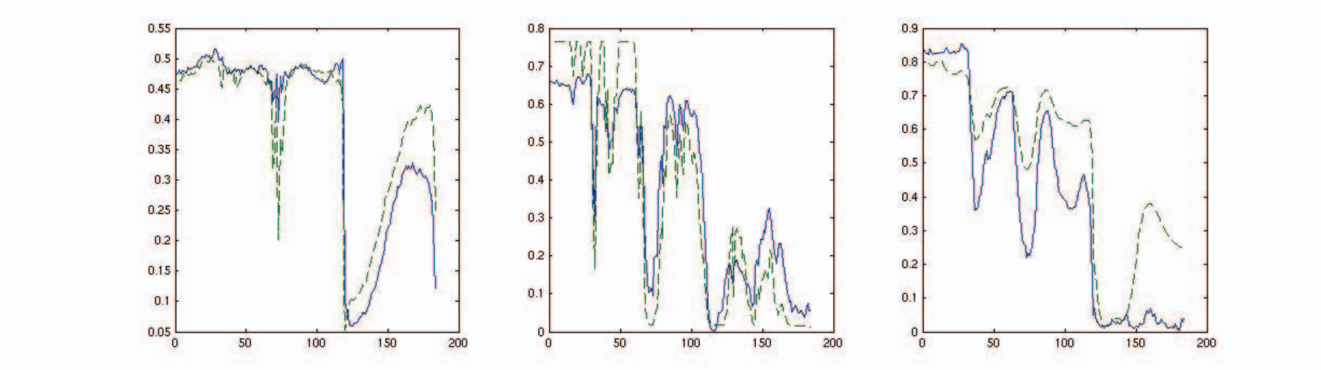

Fig. 1. Comparison between retrieved (solid line) and reference spectra (dotted line), of dust, CO_2 ice and water ice.

objective function used in ICA algorithms is the mutual information of vector $\mathbf{y}$:

$$I(\mathbf{y}, \mathbf{W}) = \sum_i H(y_i) - H(\mathbf{y}) \tag{3}$$

where $H(y_i)$ and $H(\mathbf{y})$ are the entropy of random variable y_i and of random vector $\mathbf{y}$, respectively. More details about the general framework of ICA can be found in [2]–[4]

III. ICA FOR SOURCE SEPARATION

When applied to hyperspectral imaging, the two basic assumptions of ICA are that each pixel is a linear mixture of the endmember signatures weighted by the corresponding abundance fractions and that the sources are independent. Concerning hyperspectral data, only the first assumption can be satisfied, while the second assumption is not valid, due to physical constraints on the acquisition process. Because of this, ICA fails when used to retrieve endmembers. In this paper we propose to use ICA as a pre-processing step before applying a Bayesian Positive Source Separation algorithm (BPSS). The basic idea is to use spatial ICA to obtain a rough classification of the pixels, which allows selection of small, but relevant, number of pixels and then use BPSS for the estimation based on positivity of the source spectra using the spectral mixtures provided by this reduced set of pixels. In our previous work we succesfully applied this combination of ICA and BPSS on a raw hyperspectral data cube of the South Polar Cap of Mars, provided by the Mars Express on board imaging spectrometer OMEGA [9]. In this paper we investigate the results obtained in the case of a data set pre-processed by astrophysicists, where a number of bands has been removed, being corrupted by noise or atmospheric phenomena. This task, which at first glance seems to be easier, is complicated by the fact that in this case pre-processing discard some information which could be useful for the spectra recognition.

III-A. Bayesian positive source separation

In the Bayesian approach, one can ideally incorporate any prior knowledge as long as this knowledge can be stated in statistical terms. The approach is founded on the likelihood $p(\mathbf{x}|A)$ and prior probabilities of the sources and mixing matrix. The Bayes theorem leads to

$$p(\mathbf{A}, s|\mathbf{x}) = p(\mathbf{x}|\mathbf{A}, s) + p(\mathbf{A}, s) - p(\mathbf{x}) \tag{4}$$

where the probability $p(\mathbf{x})$ is a constant, and given the independence between sources and mixing matrix the following equation is reached

$$p(\mathbf{A}, s|\mathbf{x}) = p(\mathbf{x}|\mathbf{A}, s) + p(\mathbf{A}) + p(s) + c \tag{5}$$

where c is a constant. From this posterior law, both the mixing matrix $\mathbf{A}$ and the sources s can be estimated with a joint *maximum a posteriori* criterion. The marginal posterior probabilities $p(\mathbf{A}|\mathbf{x})$ and $p(s|\mathbf{x})$ can also be used to focus the estimation on either the sources or the mixing matrix. To incorporate non-negativity of sources of the mixing matrix, a Gamma distribution is assumed. Each source s is supposed to be independent and identically distributed with a Gamma distribution. To force non-negativity of the mixing matrix, each column of the mixing matrix is assumed to have a Gamma distribution as well, giving a new posterior law. A Markov Chain Monte Carlo is then applied to simulate the resulting joint posterior density. From this, the source signals and the mixing coefficients are estimated by using marginal posterior densities.

III-B. Experimental results

The OMEGA spectrometer, carried by MARS Express spacecraft, has three channels, a visible channel and two near infrared channels. The spatial resolution of the instrument ranges from 300 m to 4 km. In this work we will focus on the near-infrared channels, which allow to better discriminate the behaviour of the major chemical components. The considered data cube is obtained by looking to the South Polar Cap of Mars in the local summer, where CO_2 ice, water ice and dust were previously detected. In the spatial domain it is composed by 604*128 pixels, and after the pre-processing step contains 184 bands from 0.93 μm to 4.15 μm. Since the area covered by the sensor is very large, we divided the

Table I. Training-test samples for the different data sets.

	Training samples	Testing samples
ROSIS	3921	42776
AVIRIS limited	320	10366
AVIRIS complete	2521	10366

whole image into 5 parts. In conformity with our *a priori* knowledge, at least three components are known to be on the soil of the planet. We chose the number of IC after a PCA analysis: with 4 principal components, 99.50% of the variance of the initial image is preserved. We obtained from each window a subset of 120 signifiant pixels (30 for each source), and then we used BPSS to recover the sources. Fig. 1 shows a comparison between the reference spectra and the spectra retrieved with our approach.

IV. ICA FOR CLASSIFICATION

One of the major difficulties in hyperspectral classification is the curse of dimensionality. The high number of features in a hyperspectral image is a major drawback for several reasons, such as the large number of training samples required and the Hughes'phenomenon [1]. PCA is a commonly used technique for dimensionality reduction, for its simplicity and ease of use: Projecting the data with a linear transformation, it maximizes the variance. Unfortunately, PCA only considers second-order statistics, and this can make it not effective in the case of hyperspectral data, since many substances covered by very high spectral resolution sensors cannot be characterized by second-order statistics. ICA considers higher order statistic to force independence of the components, retaining information which won't be preserved by common second-order statistic algorithms, and therefore seems to be an attractive method for dimensionality reduction. Another advantage is that the density of the probability distribution of the observed components can be obtained with a non-parametric density estimator. In this work, these two characteristics of ICA are exploited: in Section IV-A, the use of ICA as a pre-processing step before SVM classification is investigated. In Section IV-B, we use ICA to force the independence of the observed components. This allows to easily compute the conditional densities and to apply a Bayes detector for the class assignment.

IV-A. ICA and SVM

The support vector machines (SVM) is surely one of the most used kernel learning algorithm. It performs robust non-linear classification of samples using the kernel trick. The idea is to find a separating hyperplane in some feature space induced by the kernel function while all the computations are done in the original space. However if SVMs have shown remarkable abilities to deal with high-dimensional data, irrelevant or correlated features could deteriorate the generalization performance of SVM due to the "curse of dimensionality", especially when limited training sets are available. Thus, it seems useful to perform feature extraction in SVM [10]. In this paper, we use a SVM classifier where the parameters are optimized with a gradient descent method, as shown in [11], and we investigate the effect of dimensionality reduction with PCA and ICA on the classification accuracy.

IV-B. Independent Component Discriminant Analysis

The second method we propose is a generalization of the quadratic discriminant analysis, where the ability of ICA to retrieve components as independent as possible is exploited to estimate the marginal densities of the transformed components. The risk incurred when performing a classification of a measured vector **x** into one of K possible classes is given by:

$$R(\hat{k}|\mathbf{x}) = \frac{\sum_{k=1}^{K} L(k,\hat{k}) f_k(\mathbf{x})\pi_k}{\sum_{k=1}^{K} f_k(\mathbf{x})\pi_k} \quad (6)$$

where π_k is the *a priori* probability that **x** belongs to the class k, f_k is the class-conditional *a priori* density of k. By choosing $\hat{k}$ such that minimizing the numerator of the equation 6, this leads to the so-called Bayes decision rule. In the case of hard classification, the Bayes rule reduces to the following rule: allocate **x** to population $\hat{k}$ such that

$$\hat{k} = d(\mathbf{x}) = argmax\{f_k(\mathbf{x})\pi_k\} \quad k = 1,..,K. \quad (7)$$

The most often applied classification rules are derived by assuming that the class-conditional densities are p-variate normal with mean vectors μ_k and variance-covariance matrices assumed to be equal. This approach works well when good estimates can be obtained from the population parameters, but it is uneffective for hyperspectral data. In [12], Amato *et al.* used ICA to enforce independence to the components of the analyzed data. In the new transformed space the components are as independent as possible: This allows to use as a multivariate density estimator the product of univariate densities. The results obtained will be then substituted in the Bayes rule for the class assignment. Later on, we will refer to this approach as Independent Component Discriminant Analysis (ICDA).

IV-C. Experimental results

Two hyperspectral data sets were considered in this work. The first one is an airborne data from the ROSIS-03 with 103 bands, ranging from 0.43 to 0.86 μm, with nine classes of interest. The second data set is a segment of an AVIRIS data set over the agricultural area of Indiana, composed of 220 spectral channels acquired in the 0.4-2.5 μm region. Sixteen groundtruth classes were considered. Different training sets were randomly constructed from the reference data; each experiment was repeated five times and the average results

Table II. Classification accuracies: XCA_N indicates that N components were retained for the classification

Pre-processing	Nothing	PCA_10	PCA_20	ICA_10	ICA_20	ICDA_10	ICDA_20
	AVIRIS Full training set						
OA	76.60 $\pm$ 0.48%	78.17 $\pm$ 0.60%	79.94 $\pm$ 0.53%	80.80 $\pm$ 0.53%	**85.59** $\pm$ 0.45%	73.2 $\pm$ 0.61%	75.04 $\pm$ 0.72%
κ	73.18 $\pm$ 0.55%	75.05 $\pm$ 0.71%	76.48 $\pm$ 0.61%	77.99 $\pm$ 0.63%	**83.48** $\pm$ 0.52%	70.0 $\pm$ 0.41%	71.83 $\pm$ 0.68%
	AVIRIS Reduced training set						
OA	56.61 $\pm$ 1.56%	56.84 $\pm$ 1.17%	57.82 $\pm$ 2.18%	58.51 $\pm$ 1.82%	64.33 $\pm$ 1.93%	64.42 $\pm$ 1.48%	**66.27** $\pm$ 0.69%
κ	51.46 $\pm$ 1.46%	51.89 $\pm$ 1.04%	52.95 $\pm$ 2.18%	53.95 $\pm$ 1.97%	60.22 $\pm$ 2.07%	60.15 $\pm$ 1.52%	**62.22** $\pm$ 0.82%
	ROSIS data set						
OA	77.89%	72.44%	78.06%	79.87%	74.48%	**80.74**	80.33
κ	72.34%	65.94%	72.56%	75.77%	70.10%	**75.25**	75.20

reported. Informations about the training and test size of each data set are presented in Table I. Table II shows the results obtained with the different algorithms. For both ROSIS data set and AVIRIS with reduced training set, the best results are achieved with ICDA. In the case of AVIRIS with complete training set, the SVM with ICA pre-processing outperforms the other techniques.

V. CONCLUSIONS

In this paper, we have investigated the use of ICA for hyperspectral image analysis. In a first part, source separation was addressed. Since the independence of sources is not verified in hyperspectral real data images, ICA, if used alone, is not a suitable tool to unmix sources. We propose the use of ICA as a pre-processing step for a Bayesian Positive Source Separation in order to reduce the computational load without penalizing the performances of this method. The spectral accuracy allows to easily identify the retrieved spectra, when compared with the reference ones. In a second part, the use of ICA is studied in the frame of hyperspectral data classification. The experiments show that the pre-processing step allows to improve the classification accuracy obtained with a SVM. In particular, ICA outperforms Principal component analysis when used for dimensionality reduction. The proposed method ICDA, which uses ICA before a Bayesian classification, shows interesting results and gives the best results for the ROSIS and AVIRIS data set with reduced training set. As a conclusion, ICA has indeed a strong potential for the analysis of hyperspectral data, but it should be used cautiously, in an appropriate way and in complement with other algorithms.

ACKNOWLEDGMENTS

This work has been supported by the European Community's Marie Curie Research Training Networks Programme under contract MRTN-CT-2006-035927, Hyperspectral Imaging Network (HYPER-I-NET).

VI. REFERENCES

[1] C. Lee and D. A. Landgrebe, "Analyzing high dimensional multispectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 4, no. 31, pp. 792-800, Jul. 1993.

[2] P. Comon, "Independent component analysis, a new concept?," *Sign. Proc.*, vol. 36, pp. 287-314, 1994.

[3] T.W. Lee, M. Girolami, A.J. Bell and T.J. Sejnowski, "An unifying information-theoretic framework for Independent Component Analysis," *Comp. Math. with Appl.*, vol. 31, no. 11, pp. 1-21, Mar. 2000.

[4] A. Hyvarinen, J. Karhunen and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.

[5] L. Parra, K.R. Mueller, C. Spence, A. Ziehe and P. Sajda, "Unmixing hyperspectral data," *Adv. Neur. Inf. Proc. Sys.*, vol. 12, pp. 942-948, 2000.

[6] J.M.P. Nascimento and J.M. Bioucas Dias, "Does Independent Component Analysis play a role in unmixing hyperspectral data?," *IEEE Trans. on Geosci. Remote Sens.*, vol. 43, no. 1. pp. 175-187, Jan. 2005.

[7] J. Wang and C.I. Chang, "Independent Component Analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosc. Remote Sens.*, vol. 44, no. 6, pp. 1586-1600, Jun. 2006.

[8] M. Lennon, G. Mercier, M.C. Mouchot and L. Hubert-Moy, "Independent component analysis as a tool for the dimensionality reduction and the representation of hyperspectral images", in *IEEE International Geoscience and Remote Sensing Symposium*, IGARSS '01. Proceedings, 2001

[9] S. Moussaoui, H. Hauksdottir, F. Schmidt, C. Jutten, J. Chanussot, D. Brie, S. Doute and J.A. Benediktsson, "On the decomposition of Mars hyperspectral data by ICA and Bayesian positive source separation", *Neurocomputing*, Volume 71, Issues 10-12, Pages 2194-2208, 2008.

[10] F.E.H. Tay and L.J. Cao, "A Comparative Study of Saliency Analysis and Genetic Algorithm for Feature Selection in Support Vector Machines", *Int. Data Analysis*, 5 (3), 2001.

[11] A. Villa, M. Fauvel, J. Chanussot, P. Gamba and J.A. Benediktsson, "Gradient Optimization for multiple kernel's parameters in support vector machines classification," in *IEEE International Geoscience and Remote Sensing Symposium*, IGARSS '08. Proceedings, 2008

[12] U. Amato, A. Antoniadis and G. Gregoire, "Independent Component Discriminant Analysis", *Intern. Math. Journal*, vol. 3, no. 7, pp. 735-753, 2003.