Independent Component Analysis as a tool for the dimensionality reduction and the representation of hyperspectral images

M. Lennon[1], G. Mercier[1], M.C. Mouchot[1], L. Hubert-Moy[2]

[1]Ecole Nationale Supérieure des Télécommunications de Bretagne - Département ITI
Technopôle de Brest Iroise - BP 832 - 29285 Brest Cédex - France
[2]Laboratoire Costel - Université de Rennes 2 - 6, avenue Gaston Berger - 35043 Rennes Cédex- France
Tél : (33) (0)229001069 ; Fax : (33) (0)229001098 ; Email : marc.lennon@enst-bretagne.fr

Abstract – **Independent Component Analysis (ICA) is a multivariate data analysis process largely sudied these last years in the signal processing community for blind source separation. This paper proposes to show the interest of ICA as a tool for unsupervised analysis of hyperspectral images. The commonly used Principal Component Analysis (PCA) is the mean square optimal projection for gaussian data leading to uncorrelated components by using second order statistics. ICA rather uses higher order statistics and leads to independent components, a stronger statistical assumption revealing interesting features in the usually non gaussian hyperspectral data sets.**

I. INTRODUCTION

Current hyperspectral imaging sensors provide images including a huge number of spectral bands, typically from a few to several hundred ones. These data can be represented in a vectorial space whose dimension is equal to this number of spectral bands. However, the intrinsic dimensionality of data is usually largely less important than the dimension of the support vectorial space. In order to analyse such data, it is also necessary to project the data in a lower dimension space according to the intrinsic dimensionality of the data. Original data should be compressed while preserving the maximum amount of information for processing or visualisation. The main reason is that pattern recognition algorithms have poor capabilities in large dimension spaces. For example, the estimation of the statistical properties of classes in a supervised classification process needs the number of training samples to exponentially increase when the number of dimensions of the data increase. Multidimensional segmentation process also needs a “reasonable” number of dimensions to be achievable, usually simply for execution time considerations. Moreover, the visualisation of such data asks for the creation of new channels allowing the data to be represented the most synthetic way while preserving the maximum amount of visual information. PCA is one of the techniques commonly used for the purpose of this dimensionality reduction step in the process of data analysis. PCA leads to optimal projections in the case of a single source of information corrupted with gaussian noise. From a geometrical point of vue, PCA is optimal when the multidimensional scatterplot of data reveals to be hyperelliptic. The distributions of hyperspectral data sets are usually not gaussian and interesting information about “pure endmembers” is concentrated on vertices in the scatterplot of data. Such information can then become indistinguishable with principal components projections. In addition, small size endmembers which do not contribute very much to the total amount of the variance of information may be lost with a principal component analysis. Independent Component Analysis [1] enables the data to be represented by statistically independent components unlike PCA which leads to uncorrelated components. The statistical independance takes into consideration higher order moments and is so a stronger statistical property than decorrelation (the second order statistic used in PCA). ICA has been largely studied these last years by reasearchers from the signal processing community in the field of blind source separation [2]. The technique is based on the optimisation of an independance criterion between the observed signals and leads to a linear non orthogonal projection of the data. In the case of blind source separation, the observed signals are unknown but assumed to be linear combinations of unknown non gaussian sources. With the assumption that the sources are statistically independent, ICA recovers sources from the observed signals. It has been demonstrated [3] that in the case of non gaussian sources, the optimisation of an independence criterion is equivalent to the maximisation of a non gaussianity criterion due to the central limit theorem. Consequently, ICA is able to compute projections which leads to the least gaussian projected data. In other words, probability density functions of projected data are the most skewed or multimodal. This feature is quite interesting for future classification of the data set.

II. INDEPENDENT COMPONENT ANALYSIS

The model of ICA if given by :

$$\boldsymbol{x} = A\,\boldsymbol{s} \qquad (1)$$

with :
- $\boldsymbol{x}$: Vector of observed signals
- A : Scalar matrix of mixing coefficients
- $\boldsymbol{s}$: Vector of source signals

ICA requires 2 hypothesis :

(H1) : The components s_i of $\boldsymbol{s}$ are statistically independent.

(H2) : The components s_i of $\boldsymbol{s}$ have non gaussian distribution (It can be shown that no more than one component can be gaussian).

Model (1) implies two non indentifiabilities :
(NI1) : Impossibility of retrieving the variance of the sources (one term a_{ij} of A can be multiplied by a constant and the component s_j of $\boldsymbol{s}$ by the inverse of this constant without changing the model).
(NI2) : Impossibility of retrieving the order of the sources (the order of the components s_i of $\boldsymbol{s}$ and of the lines of the matrix A can be modified without changing the model).
The goal of ICA is to find the matrix $W = A^{-1}$ so that the sources $\boldsymbol{s}_{est} = W\boldsymbol{x}$ can be estimated from the vector $\boldsymbol{x}$ of the observed signals by optimizing a statistical independence criterion.

A set of signals s_i are said to be independent if a signal s_j gives no information on the other signals of the set. The joint probability density function is then equal to the product of marginal probability density functions :

$$p(s_1, \ldots, s_i) = \prod p_{si}(u_i) \tag{2}$$

It can be trivially shown that independence implies decorrelation but the inverse is not true. This observation enables the estimation procedure of the independent signals to be restricted to the space of uncorrelated signals. The first step in computing the ICA is then a decorrelation of the observed signals, which can be achieved with a usual PCA. This step also allows the original data set to be reduced according to the magnitude of covariance matrix eigenvalues. There is a strong relation between independence and non gaussianity. The Central Limit Theorem tells that the sum of N independent random variables tends in law towards a gaussian distribution when N tends to the infinity. The corollary of this theorem tells that the sum of two independent random variables has a more gaussian distribution than anyone of the initial random variables. The estimation of the independent components consists then in finding the weights of the matrix W so that the vector $\boldsymbol{s}_{est} = W\boldsymbol{x}$ has the least gaussian components. A measure of non gaussianity is then needed. The Kurtosis is a usual measure which has the properties of being null for a gaussian distribution and respectively negative and positive for super and subgaussian distributions. Unfortunately, kurtosis is very sensitive to outliers. A more robust measure is given the negentropy. The entropy of a random vector $\boldsymbol{x}$ of density $p_x(\boldsymbol{u})$ is defined by :

$$H(p_x) = -\int p_x(\boldsymbol{u}) \log p_x(\boldsymbol{u})\, d\boldsymbol{u} \tag{3}$$

$H(p_x)$ is maximal for a gaussian random vector $\boldsymbol{x}$. The negentropy is then defined by the deficit of entropy between $\boldsymbol{x}$ and a gaussian random vector $\boldsymbol{x}_g$ of the same covariance matrix as $\boldsymbol{x}$:

$$J(\boldsymbol{x}) = H(p_{xg}) - H(p_x) \tag{4}$$

The Kullback-Leibler divergence allows the distance between two distributions p_v and p_w to be measured :

$$k(p_v, p_w) = -\int p_v(\boldsymbol{u}) \log[\, p_v(\boldsymbol{u}) / p_w(\boldsymbol{u})]\, d\boldsymbol{u} \tag{5}$$

It can be shown that negentropy $J(\boldsymbol{x})$ is equal to the Kullback-Leibler divergence between the vector $\boldsymbol{x}$ and the gaussian vector $\boldsymbol{x}_g$:

$$J(\boldsymbol{x}) = k(p_x, p_{xg}) \tag{6}$$

This expression needs an estimation of the densities and is thus difficult to compute. Hyvärinen [3] has then introduced the following estimation of negentropy :

$$J(\boldsymbol{x}) = \{ E[G(\boldsymbol{x})] - E[G(\boldsymbol{x}_g)] \}^2 \tag{7}$$

with : G : non quadratic function. More details on independence and demonstrations of the preceding equations can be found in [3].

The maximisation of the measure of non gaussianity (7) needs an optimisation algorithm. Fast-ICA algorithm [4] has been used. It is based on a fixed-point iteration scheme and is given for one component :

1. choose an initial (e.g random) weight vector $\boldsymbol{w}$.
2. Let $\boldsymbol{w}^+ = E\{\boldsymbol{x}\, g(\boldsymbol{w}^T\boldsymbol{x})\} - E\{g'(\boldsymbol{w}^T\boldsymbol{x})\}\, \boldsymbol{w}$
   with : $g(\boldsymbol{u}) = \tanh(\boldsymbol{u})$
3. Let $\boldsymbol{w} = \boldsymbol{w}^+ / \|\boldsymbol{w}^+\|$

After every iteration, the vectors $\boldsymbol{wx}$ are decorrelated using a symmetric decorrelation of the matrix W :

$$W = (W W^T)^{-1/2}\, W \tag{8}$$

with W : matrix $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n)^T$ of vectors $\boldsymbol{w}_i$ and $(WW^T)^{-1/2}$ is obtained by the eigenvalue decomposition of W. This step avoid a direction to be estimated several times and do not privilegiate a vector among others.

In our case of the dimensionality reduction of hyperspectral data, the physical sources do not exist and then model (1) does not hold and assumption (H1) is not important. ICA is only used to find the projection where all the projected components are "the most independent" in the sense of negentropy, what is made possible according to (8). From a practical point of vue, the estimation of esperances may be computed on a sample of the original data set in order to limit the computational time if the size of the data set is too large. To avoid problems of convergence distinct samples for each iteration should be chosen. It is also interesting to perform a prefiltering of the original data in order to reduce outliers and then obtainig a better estimation of the directions of projection. The estimated matrix of projection can then be applied on the original data set. It can be trivially shown than linear filtering the original data does not affect the model (1). This consideration is however not important in our case because the model (1) does not hold. A simple mean filter leads to significative improvements.

III. RELATION TO PROJECTION POURSUIT

Independent Component Analysis is closely related to Projection Poursuit (PP) [5], a statistical data analysis tool designed for reducing the dimensionality of multivariate data sets by finding “interesting projections” according to a projection index. PP has recently been proposed for the analysis of hyperspectral images both for supervised classification [6] and for unsupervised data analysis [7]. The main difference is that ICA simultaneously looks for all the components and find the directions where all the projected components are “the most independent” in the sense of a measure to independence. ICA does not privilegiate any direction among others. In opposition, standard PP, as used in [7], sequentially looks for one projection. Once a direction is found, data are projected onto the subspace orthogonal to the latter and so on. The choice between both methods should be guided by the application. Moreover, a discrete version of differential entropy is used for the index projection in [7]. We rather used the neguentropy, a best and faster-to-compute estimation of independance. It has also been demonstrated that the maximization of neguentropy is equivalent to the minimization of mutual information, the natural information theory measure of independence.

IV. RESULTS

ICA has been applied to a 2 meters ground reolution image from the CASI Sensor including 14 spectral bands from 450 to 950 nm. The image has been corrected to reflectance by the mean of the empirical line method. Fig. 1 and Fig. 2 shows the first three components for PCA and ICA.

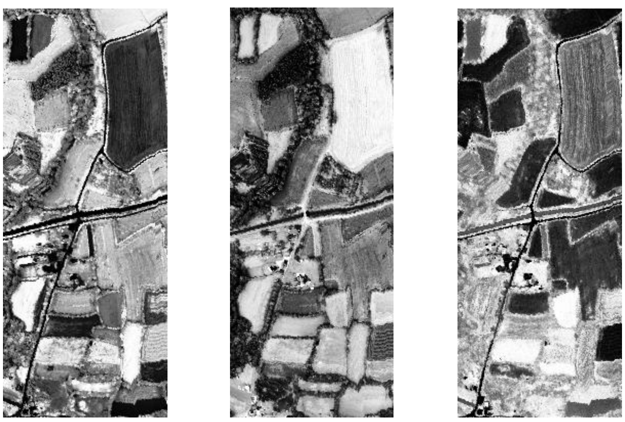

Fig. 1. Principal Component Analysis

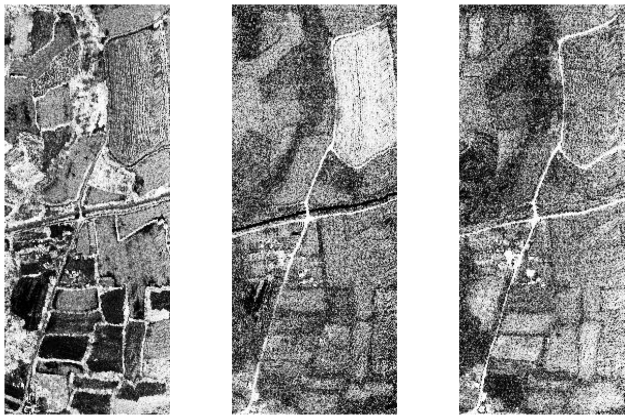

Fig. 2. Independent Component Analysis

It can be noticed that,unlike PCA, ICA extracted separately the main different materials present in the scene (wood, different types of soil occupation, roads). Even if it is more noisy, a false color composition of the components reveals the superiority of ICA for vizualisation. Moreover, no one of histograms of ICA components is gaussian.

V. CONCLUSION

An multivariate data analysis tool refered as Independent Component Analysis has been presented. It reveals to be quite interesting for the visualisation of large hyperspectral data sets and for dimensionality reduction, a necessary preprocessing step for segmentation, classification or other image analysis procedure. Unlike Principal Component Analysis, it is able to show interesting features even in the case of non gaussian data.

REFERENCES

[1] P. Comon, “Independent component analysis, A new concept ?”, *Signal processing*, vol. 36, pp. 287-314, 1994.

[2] C. Jutten, J. Hérault, “Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture”, *Signal processing*, vol. 24, pp. 1-10, 1991.

[3] A. Hyvärinen, E. Oja , “Independent component analysis: algorithms and applications”, *Neural Networks*, vol. 13, pp. 411-430, 2000.

[4] A. Hyvärinen, ”Fast and Robust Fixed-Point Algorithms for Independent Component Analysis”, *IEEE Tr. on Neural Networks*, vol. 10(3), pp. 626-634, 1999.

[5] M.C. Jones, R. Sibson, “What is projection poursuit ?”, *J. of R. Statist. Soc. A*, vol. 150 (1), pp. 1-36, 1987.

[6] L.O Jimenez, D.A. Landgrebe, “Hyperspectral data analysis and supervised feature reduction via projection poursuit”, *IEEE Tr. on Geoscience and Remote Sensing*, vol. 37 (6), pp. 2653-2667, 1999.

[7] A. Ifarraguerri, C.-I. Chang, “Unsupervised hyperspectral image analysis with projection poursuit”, *IEEE Tr. on Geoscience and Remote Sensing*, vol. 38(6), pp. 2529-2538, 2000.