Reduction of the number of spectral bands in LANDSAT images with projection methods: Pertinence of the resulting information

Journaux L., Foucherot I. & Gouton P.
Laboratoire LE2I UMR-CNRS 5158
Université de Bourgogne - Faculté des Sciences - Aile de l'ingénieur
BP 47870 - 21078 DIJON Cedex FRANCE
ljourn@u-bourgogne.fr

Abstract

This paper describes applications of linear and nonlinear projection methods, in order to obtain a reduction of the number of spectral bands in LANDSAT multispectral images.

We present Curvilinear Component Analysis (CCA, nonlinear method) and an optimisation of it based on the use of Principal Component Analysis (PCA, linear method).

In order to evaluate the pertinence of the information kept by each transformation, we then apply segmentation on the transformed and original images. This processing allows us to show that the structure (the landscape organization) of the image is preserved by each transformation.

This paper tends to show several results : CCA is an improvement for dimensions reduction of multispectral images ; CCA is really a nonlinear extension of PCA ; CCA optimisation through PCA (called CCAinitPCA) allows a reduction of the calculations, providing a result identical to that of CCA.

1. Introduction

To analyse natural heritage, ecologists study the relations between fauna distribution and landscape features [1]. They are particularly interested in the relationship between fauna and flora along the rivers : There, environmental conditions vary gradually, which makes them interesting models for the study of ecological gradients [2].

Satellite multispectral images seem to be nowadays the most encouraging tools for landcover analysis. This kind of data provides the relevant way to describe faunistic habitats in space and time through usable variables in multivariate analysis. For example, a segmentation of such an image allows to detect the limits of the different areas of the landscape.

The satellite sensors used nowadays supply images with a large number of spectral bands. This set of data could be represented in a vectorial space with a number of dimensions corresponding to the number of spectral bands [3, 4].

The main problem of multispectral images is the considerable quantity of data contained in them. This makes the analysis of such images very heavy. So, the reduction of the vectorial space, that is to say, the reduction of the number of spectral bands of the collected data, is a necessary preprocessing step before multispectral image analysis. It would present several advantages such as data compression and simplification, that will allow to maximize automatic processes, and reduction of the run-time during the segmentation, classification or fusion of images.

The problem is to reduce the quantity of data while keeping sufficient information for structural and informational analysis of the image.

Some methods make it possible to reduce the quantity of data while preserving their intrinsic properties.

The best known is Principal Component Analysis (PCA), also called "Hotelling transform" or "Karhunen-Loëve transform" [5, 6]. It consists in a linear transformation of the vectorial space that leads to the reduction of the number of spectral bands. The resulting image is represented by a set of new bands in which the quantity of information decreases from the first band to the last. This well-known method has several advantages, among which data decorrelation and compression.

Other linear transformation methods, such as Independent Component Analysis [7, 8] or the Projection Pursuit introduced by Friedman and Tukey [9, 10], are also frequently used in this context [3]. Unfortunately, these methods only highlight the linear relations between the spectral bands at the expense of the nonlinear relations. Therefore, their exploitation is limited.

Nonlinear projection methods, such as "Sammon's mapping" [11], or "Multidimensional Scaling" [12] require prohibitive run-times and the convergence of their algorithms is not certain.

A new nonlinear projection method, called Curvilinear Component Analysis (CCA) [13], is interesting. Used until now in several domains such as time series prediction [14], this method has recently been used for hyperspectral image processing [3].

In this article we compare various multidimensional projection methods intended to reduce the quantity of data in satellite multispectral images, while keeping the necessary and sufficient information for segmentation or classification.

Section 2 of the paper contains a short presentation of CCA and of an optimization, called CCAinitPCA, that we propose in order to transform LANDSAT images into 3-band images.

In section 3, we show on an example the redundancy of information between the spectral bands of a LANDSAT image and then explain how to define the sufficient number of bands to be retained as output of the nonlinear processing.

In section 4, we present our experiment and comment its results. We apply three projection methods (PCA, CCA and CCAinitPCA) on LANDSAT images and compare the reduced images to the initial one in term of relevance and conservation of information.

2. Nonlinear methods

The principle of CCA is to reproduce the topology of the initial set of data from a space $\mathbb{R}^n$ into a subspace $\mathbb{R}^p$ $(p < n)$, without constraining statically the configuration of the topology. It's an automatic auto-adaptation to the real shape of the scatter of points in the vectorial space. In an image, the topology is defined by the euclidian distances between all pixels (all pairs of vectors of the original data). Thus, CCA tries to find all new vectors in the new subspace $\mathbb{R}^p$ reproducing the topology of the vectors of $\mathbb{R}^n$ into $\mathbb{R}^p$. The mathematical aim of CCA is to minimize the following error function which characterizes the topology between initial and final projection space:

$$E_{CCA} \quad \frac{1}{2}\sum_{i}\sum_{j,(i\neq j)}(d_{ij}^{n}-d_{ij}^{p})^{2}F(d_{ij}^{p})$$

with:

d_{ij}^n : Euclidian distances between the vectors x_i and x_j in the original space $\mathbb{R}^n$

d_{ij}^p : Euclidian distances between the vectors x_i and x_j in the projection subspace $\mathbb{R}^p$

$F:\mathbb{R}^+ \rightarrow [0,1]$, Decreasing function of d_{ij}^p, allows the local topology to be favoured with respect to the global topology

We propose here an optimization of CCA, adapted to multispectral images, combines PCA and CCA. This method is called CCAinitPCA [15].

The principle of PCA is to determine a vectorial subspace (whose number of dimensions is lower than in the original space) in which the distribution of observations is preserved at best. The resulting set of points is then an optimized linear projection of the initial one in a reduced subspace. It is obvious that the result of PCA is close to the result of CCA.

In practice, the projection of the points in the subspace $\mathbb{R}^p$ resulting from CCA is randomly initiated. In CCAinitPCA, the random initialized matrix is replaced by the matrix of the main components resulting from PCA.

3. The satellite images

3.1. Spectral bands

The images used in this study have been provided by the French Institute of the Environment (I.F.E.N). They are multispectral images shot by the LANDSAT 7 satellite in 2001. The spatial resolution of the images is 30 meters. It includes 7 spectral bands and a panchromatic band. More details could be found in [15].

From satellite images, we extracted 40 64-by-64-pixel-images. We have chosen this size in order to respect the work scale of the avian population. Moreover, this choice allows us to obtain tolerable computing times during the processing.

3.2. Information redundancy

A calculation of correlation coefficients between each spectral band on LANDSAT multispectral images shows that they are highly redundant. To reduce the number of bands, we need an "a priori" knowledge of the "whole dimension" of the initial projection space. In other words, we have to evaluate the real informative dimension of the data in order to preserve the maximum of useful information during the projection into the new subspace.

This dimension is called the intrinsic dimension of the data, and can be defined as the number of minimum freedom degrees of the initial space projection [13]. The estimation of this intrinsic dimension is thus very important if we don't want to lose too much information

3.3. Estimation of the intrinsic dimension

Several methods can be used to estimate the intrinsic dimension of our multispectral images like fractal dimensions[16]. Here we use local eigenvalues [17] .

The intrinsic dimension is generally lower than the dimension of the initial set of data, corresponding here to the number of spectral bands. We estimated the intrinsic dimension of the Landsat 7 images by applying Local Linear Transform (LLT) [17] also called method of local eigenvalues.

The Local Linear Transform considers the local linear relations between the spectral bands. The LLT allows to study the local variability of the data (via a mobile mask) and to extract local intrinsic dimension (p_i) from each zone via the extraction of the local eigenvalues of the various covariance matrixes. The intrinsic dimension (p) of the initial projection can be evaluated by calculating the average of local dimensions (p_i) according to the following formula:

$$\bar{p} = \frac{1}{N}\sum_{i=1}^{N} p_i$$

According to this method, the intrinsic dimension of the data resulting from the multispectral images was estimated at $\bar{p} = 3$. So CCA will project the image into a 3-dimensional subspace.

4. Experiment results

We have applied PCA, CCA and CCAinitPCA to LANDSAT images and so obtained 3-dimensional new images. Figure 1 shows the 3 bands resulting from the same image after each transformation. On the new bands we can see regions that correspond to areas of the original image. However, these new components have no signification compared with the original spectral definition.

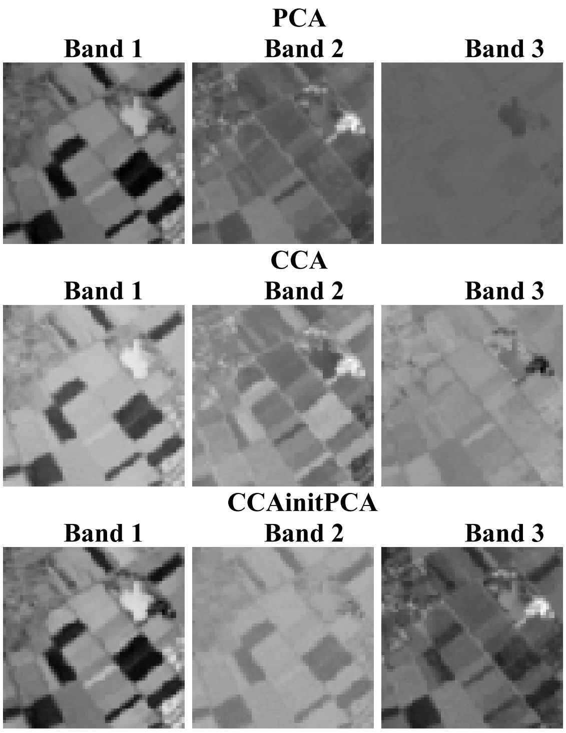


Figure 1: Resulting bands in images after reducing the number of dimensions

The images obtained with the three methods are different. We shall now compare them from different angles.

- We shall first compare the quality of the projection between CCA and CCAinitPCA.
- We shall then compare the images obtained from the various transformations, and thus evaluate the quality of the transformations. We have chosen two approaches : The first approach consists in a statistical analysis of the resulting images. The second approach is a segmentation of the images by unsupervised classification. Thus we shall show that the structure of landscape occupation is preserved by each transformation. We shall then carry out a comparison of the computing times of each algorithm.
- Finally, we shall compare convergences of CCA and CCAinitPCA algorithms.

4.1. Evaluation of the projection quality

The projection quality of CCA and CCAinitPCA can be evaluated by plotting the "dy-dx" representation proposed by Demartines and Hérault [13].

"dy-dx" projection represents the joint distribution of d_{ij}^{n} and d_{ij}^{p} for each pair of vectors. In such a representation the points close to the origin correspond to local topology while the distant points correspond to global topology. If the topology is respected, all the points are concentrated along the y=x line. While there are points outside of the y=x line, CCA needs an other iteration. When all the points are along the y=x line, the initial topology of the vectors is found again and CCA is successful.

The "dy-dx" representation is also interesting to estimate the number of necessary iterations for the convergence of CCA. Figure 2 shows the "dy-dx" representations at different stages of CCA and CCAinitPCA.

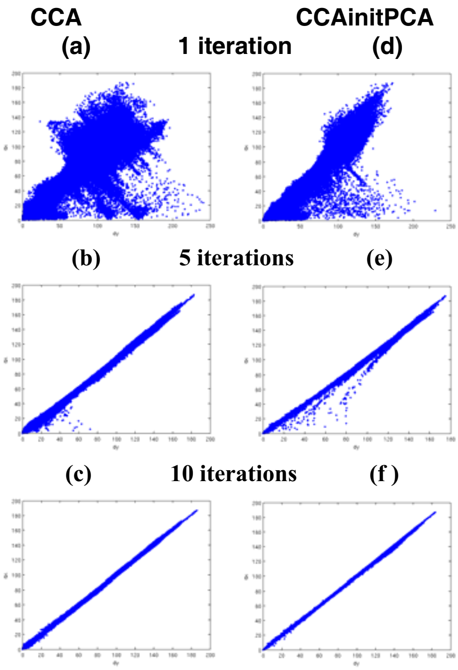


Figure 2:"dy-dx" representation of CCA (a, b, c) and of CCAinitPCA (d, e, f). One iteration corresponds to an iteration for each point that is to say $(64)^2 = 4096$ iterations.

We can see on this figure (figure 2 c, f) that CCA and CCAinitPCA converge towards the same results after 10 iterations. However, we can note that, after the first iteration, the projection of the points from CCAinitPCA is more concentrated along the y=x axis than the projection of the points from CCA. This tends to confirm that the use of PCA matrix for the initialization of ACC corresponds to an improvement of CCA.

4.2. Statistical analysis of images

In order to evaluate the differences between resulting images, we have calculated a correlation coefficient between the resulting bands (all bands, all methods) (table 1).

•=The analysis of the inter-band correlation coefficients shows three main facts:High correlation between first band of PCA, first band of CCA and first band of CCAinitPCA.

•=Low correlation between second band of PCA and second band of CCA or CCAinitPCA. We can observe the same phenomenon with the third bands.

These two points show that, for the first bands, the result of PCA is close to the result of CCA and CCAinitPCA and CCA is really a nonlinear extension of PCA.

High correlation between the second band of CCA and the second band of CCAinitPCA (same observation with the third bands of these methods).This allows us to prove that the initialization of CCA by the PCA matrix leads to a result equivalent to the result obtained by CCA initialized at random. Moreover, this method offers two advantages: a faster convergence (graphic 1) and a reduction of calculations and manipulations of data, so a reduction of run-time.

Method		**PCA**			**CCA**			**CCA init PCA**		
	Bands	*1*	*2*	*3*	*1*	*2*	*3*	*1*	*2*	*3*
	1	1	0,00004	0,00110	0,98187	0,81070	0,01990	0,98701	0,94614	0,48880
	2	0,00004	1	0,00228	0,18376	0,56869	0,95644	0,15501	0,07143	0,87074
PCA	*3*	0,00110	0,00228	1	0,02357	0,08374	0,19446	0,03626	0,24229	0,04712
	1	0,98187	0,18376	0,02357	1	0,68552	0,18310	0,94011	0,94582	0,63937
	2	0,81070	0,56869	0,08374	0,68552	1	0,48841	0,89022	0,71659	0,09668
CCA	*3*	0,01990	0,95644	0,19446	0,18310	0,48841	1	0,12424	0,06409	0,84870
	1	0,98701	0,15501	0,03626	0,94011	0,89022	0,12424	1	0,91031	0,34946
	2	0,94614	0,07143	0,24229	0,94582	0,71659	0,06409	0,91031	1	0,51060
CCAinitPCA	*3*	0,48880	0,87074	0,04712	0,63937	0,09668	0,84870	0,34946	0,51060	1

Table 1: coefficients of correlation between every new spectral band according to three methods (with a grey background, the coefficients with a correlation higher than 50 %)

4.3. Analysis by segmentation

To estimate the quality of the different transformations, we have applied a segmentation on the initial and reduced images. We have chosen an unsupervised method of classification usually applied in remote sensing: The method of aggregation around mobile centres called K-means method [18, 19].We applied the k-means method to the resulting images from each reduction method and to the original Landsat image (8 bands). The goal was to compare the segmented images and so to see if the transformations preserve the space structure of the image. In other words, we wanted to know if our transformations enable us to preserve the space configuration of the different zones of the landscape in spite of the reduction of dimensions. We arbitrarily fixed the number of classes at 9 for each image (9 different kinds of landscape). Figure 3 shows the classification results.

The observation of the segmented images obtained by the k-means method shows several results.

The first one is the conservation of the space organization of the image. Indeed, for each transformation method and more particularly for the nonlinear methods (d and e), the space organization and the geometry of the different zones of the original image are respected. This tends to show that the transformations by CCA (d) and CCAinitPCA (e) preserve the space organization of the various elements constituting the landscape in the original image (a).

Moreover, the comparison of the segmented images resulting from CCA (d) and CCAinitPCA (e) shows that the processing leads to the same results. This comparison proves that the initialization of CCA by PCA does converge towards the same result.

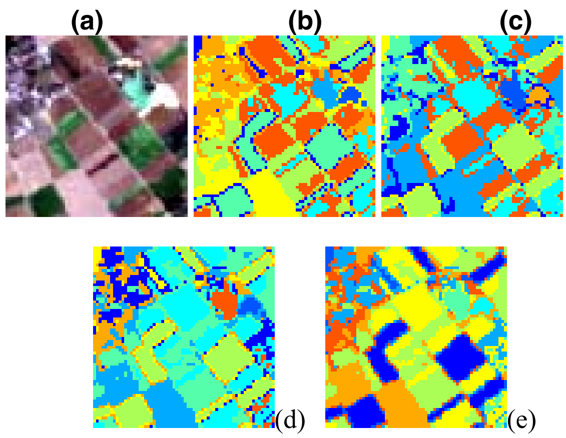


Figure 3: Segmented images obtained by unsupervised classification on (a) Original image (8 bands), (c) PCA image, (d) CCA image, (e) CCAinitPCA image. Image (b) is the representation of the initial image in the RGB space.

The second result constitutes the main interest of the reduction of dimensions in multispectral images.

Indeed, the segmentation of the initial 8-band image leads to an over-segmentation, in particular in textured zones. This could be explained by the information redundancy between the various bands of the Landsat image. On the reduced images, over-segmentation is lower. This tends to show that the reduction of dimensions allows to keep only the essential information of the image

The original color image (figure 4a) enables us to compare the quality of the segmentation on the reduced images. On the image resulting from PCA (figure 4b), several different parcels are merged in a same region by the segmentation. On the images resulting from the nonlinear methods (figure 4c) the different parcels are better differentiated.

Moreover, the image resulting from PCA leads to an over-segmentation in the textured zones (figure 4 b) while the images resulting from the nonlinear methods smooth the noise giving the segmentation a better quality (figure 4 c).

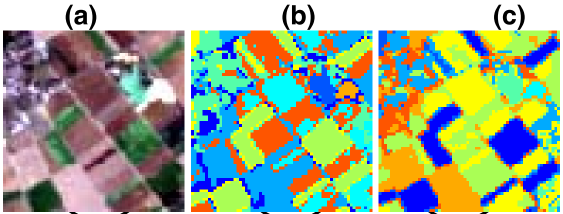


Figure 4: (a) color image, (b) segmented PCA image, (c) segmented CCAinitPCA image.

The segmentation of our images allowed us to observe unexpected and interesting results.

Indeed, we can see that the same labels (here, the same color of regions) are associated to the same kinds of landscape (Figure 5). This was not obvious at the beginning, especially with the nonlinear transformation methods.

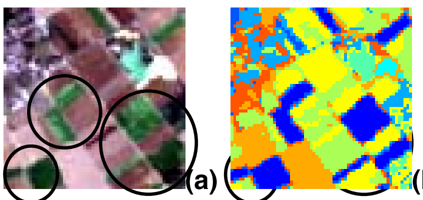


Figure 5: (a) color image, (b) segmented image resulting from CCAinitPCA.

Finally, we can observe areas with pond labels in the segmented images resulting from the nonlinear methods (Figure 6). These areas do not appear on the segmented image resulting from the PCA image. Verification at the site has revealed the presence of ponds under vegetable cover. This enables us to consolidate the idea that CCA and CCAinitPCA make it possible to reduce considerably the quantity of data with a maximum preservation of information sometimes lost by the linear methods.

From the analysis by segmentation, we draw the following conclusions:

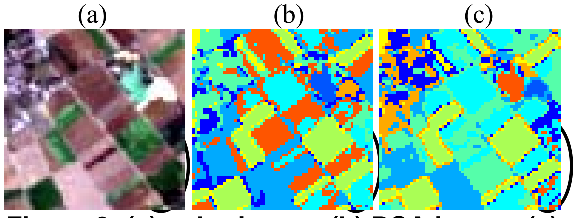


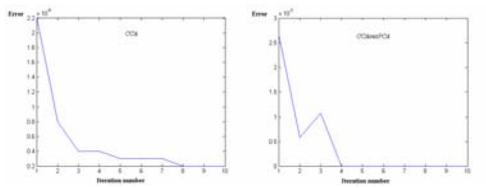
Figure 6: (a) color image, (b) PCA image, (c) CCA image.

All methods preserve the spatial organization of the landscape.

CCA and CCAinitPCA preserve sufficient information (more than PCA) to recognize the different kinds of landscape in reduced images

4.4. Algorithmic convergence

The preliminary results are encouraging, but there are important improvements to realize. The main difficulties concern the speed of the convergence and its robustness. Actually, during CCA, the minimization of the E_{CCA} function is realized with a method of stochastic gradient descent. This method is not optimal because it is a quasi-Newton method of the first order [20].



Graphic 1: Convergences obtained according to CCA and CCAinitPCA; the convergence is faster for CCAinitPCA method.

To validate our transformation method, we have used the stockastic gradient descent described by Demartines [13], supposed to avoid the algorithmic problem of being blocked in a local minimum. At this point in our work, about 65 % of the gradient descent succeeds in converging towards a global minimum, verified by error estimation.

To solve the problem of convergence, we envisage to use methods of second order gradient resolution, such as BFGS and the Levenberg-Marquart method [20]. These methods allow to tend towards a global minimum of the function while preserving the convergence speed. They ensure an algorithmic robustness by the Hessian control even if on the other hand, the computing time may increase significantly.

5. Conclusion

The reduction of the number of dimensions is a necessary preprocessing for analyzing multispectral images. PCA, CCA and CCAinitPCA are transformation methods which allow to reduce the quantity of data while maintaining the best part of the objects properties contained in the image. Applying such a transformation on a multispectral image thus permits to facilitate its analysis.

We have noticed that the use of the matrix originating from PCA to initiate CCA allows to improve the quality of the resulting images, because of bypassing vectorial quantization, and to reduce the time cost of the transformation. The images obtained by CCAinitPCA are very clearly close to the CCA images while reproducing the real topology of the data.

A segmentation of the reduced images enabled us to consider the relevance of the information kept by each transformation. We have observed that the nonlinear methods make it possible to decrease the quantity of data while keeping the capacity to reveal the main part of the properties of the objects contained in the image. They make it possible to extract the structure of the landscape, and thus to facilitate the classification or segmentation. Other classification methods should be tested to support these results. And the results obtained should be compared to the reality in order to evaluate the rate of "good" classification of the pixels.

A convergence algorithm problem remains. To overcome it, we envisage to use the resolution gradient method of the second order in the error function. Even if its time cost is high, it should ensure convergence of the E_{CCA} function towards a global minimum and thus an optimal projection of the data in the reduced subspace. According to this results and in order to validate best our approach, we envisage to realize a program of compression/decompression for colors images processing.

6. References

[1] S. A. Cushman and K. McGarigal, "Patterns in the species-environment relationship depend on both scale and choice of response variables," *Oikos*, vol. 105, pp. 117-124, 2004.

[2] B. Frochot, M.-C. Eybert, L. Journaux, J. Roché, and B. Faivre, "Nesting birds assemblages along the River Loire: result from a 12 years-study," *Alauda*, vol. 71, pp. 179-190, 2003.

[3] M. Lennon, G. Mercier, M. C. Mouchot, and L. Hubert-Moy, "Curvilinear Component Analysis for nonlinear dimensionality reduction of hyperspectral images.," presented at Image and Signal Processing for Remote Sensing VII, SPIE's International Symposium on Remote Sensing 2001, Toulouse, France, 2001.

[4] B. Tso and P. M. Mather, *Classification Methods for Remotely Sensed Data*: Taylor and Francis Ltd (London), 2001.

[5] J. C. Devaux, P. Gouton, and F. Truchetet, "The Karhunen-Loeve Transform Applied to Région-Based Segmentation of Color Aerial Images," *Optical Engineering*, vol. 40, pp. 1302-1308, 2001.

[6] H. Hotelling, "Analysis of a complex statistical variable into principal component," *J. Edu. Psy.*, vol. 24, pp. 417-441 & 498-520, 1933.

[7] P. Comon, "Independent Component Analysis, a new concept ?," *Signal Processing*, vol. 36, pp. 287-314, 1994.

[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," presented at IGARSS 2001, Sydney,Australia, 2001.

[9] A. Ifarraguerri and C.-I. Chang, "Unsupervised Hyperspectral Image Analysis with Projection Pursuit," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, pp. 2529-2538, 2000.

[10] J. H. Friedman and J. W. Tukey, "A projection pursuit algorithm for exploratory data analysis," *IEEE Transactions on computers*, vol. C23, pp. 881-890, 1974.

[11] J. W. Sammon, "A nonlinear mapping for data analysis," *IEEE Transactions on Computers*, vol. C-18, pp. 401-409, 1969.

[12] B. D. Ripley, *Pattern recognition and neural networks*: Cambridge University Press, 1996.

[13] P. Demartines and J. Hérault, "Curvilinear Component Analysis : A self-organizing neural network for nonlinear mapping of data sets," *IEEE Transactions on neural networks*, vol. 8, pp. 148-154, 1997.

[14] A. Lendasse, J. Lee, E. De Bodt, V. Wertz, and M. Verleysen, "Input data reduction for the prediction of financial time series," presented at ESANN'2001, 2001.

[15] L. Journaux, I. Foucherot, and P. Gouton, "Nonlinear reduction of multispectral images by Curvilinear Component Analysis: Application and optimization," presented at CSIMTA'04 International Conference, Cherbourg, France, 2004.

[16] F. Camastra and A. Vinciarelli, "Estimating the Intrinsic Dimension of Data with a Fractal-Based Method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1404-1407, 2002.

[17] J. Bruske and E. Merényi, "Estimating the Intrinsic Dimensionality of Hyperspectral Images," presented at ESANN'1999, Bruges (Belgium), 1999.

[18] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*, 2001.

[19] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*. Massachusetts, 1974.

[20] R. Fletcher, *Practical Methods of Optimization,*, 2nd Edition ed: John Wiley & Sons, 2000.