

Spectral transformation based on nonlinear principal component analysis for dimensionality reduction of hyperspectral images

Giorgio Licciardi^{a,b} and Jocelyn Chanussot^{a,c}

^aSigmaphy, GIPSA-Lab, Grenoble, France; ^bHypatia Research Consortium, Rome, Italy; ^cFaculty of Electrical and Computer Engineering, University of Iceland, Reykjavik, Iceland

ABSTRACT

Managing transmission and storage of hyperspectral (HS) images can be extremely difficult. Thus, the dimensionality reduction of HS data becomes necessary. Among several dimensionality reduction techniques, transform-based have found to be effective for HS data. While spatial transformation techniques provide good compression rates, the choice of the spectral decorrelation approaches can have strong impact on the quality of the compressed image. Since HS images are highly correlated within each spectral band and in particular across neighboring bands, the choice of a decorrelation method allowing to retain as much information content as possible is desirable. From this point of view, several methods based on PCA and Wavelet have been presented in the literature. In this paper, we propose the use of NLPCA transform as a method to reduce the spectral dimensionality of HS data. NLPCA represents in a lower dimensional space the same information content with less features than PCA. In these terms, aim of this research is focused on the analysis of the results obtained through the spectral decorrelation phase rather than taking advantage of both spectral and spatial compression. Experimental results assessing the advantage of NLPCA with respect to standard PCA are presented on four different HS datasets.

ARTICLE HISTORY

Received 19 June 2017
Revised 9 February 2018
Accepted 13 February 2018

KEYWORDS

Hyperspectral image processing; dimensionality reduction; noise suppression; NLPCA

Introduction

Hyperspectral (HS) sensors collect information on a very high number of wavelengths, corresponding to tens or hundreds of bands. These kinds of data become increasingly popular and are extremely useful in several fields of application. Usually, in Earth observation, HS images are acquired by sensors mounted on airborne or satellite-borne carriers. While airborne-based sensors can cover only limited regions of the Earth surface, satellite based sensors have the ability to collect information over the entire globe. However, due to the size of a typical HS dataset, not all the acquired data can be downlinked to a ground station. For this reason the dimensionality reduction of HS data becomes necessary in order to match the available transmission bandwidth. From this point of view it is possible to take advantage of the high degree of spectral and spatial correlation of HS data. In general, image compression approaches can be grouped as lossless or lossy compressions. Lossless approaches are usually based on the suppression of the statistical redundancy of the data (Jing & Guizhong, 2006; Wei et al., 2010). On the other hand, a lossy algorithm aims to minimize the data volume by discarding non-relevant part of information. Lossy compression is usually used when higher ratios are required. Moreover, as long as the

radiometric resolution of the image increases, expressed as number of bit per pixel, lossy approaches obtain better results than lossless techniques, in terms of quality of reconstructed images. In the literature, several lossy approaches have been proposed for the compression of HS images (Abousleman et al., 1995; Conoscenti, Coppola, & Magli, 2016; Fowler et al., 2007; Karami, Heylen, & Scheunders, 2015; Kulkarni et al., 2006). Many of these techniques are based on decorrelation transforms, in order to exploit both spatial and spectral correlations, followed by a quantization stage and an entropy coder. In particular these approaches involve the combination of a 1-D spectral decorrelator such as the principal component analysis transform (PCA), the Discrete Wavelet Transform (DWT), or the Discrete Cosine Transform (DCT), followed by a spatial decorrelator (Abrardo, Barni, & Magli, 2010; Christophe, Mailhes, & Duhamel, 2008; Kaarna et al., 2000). It is not difficult to understand that the spectral decorrelation phase plays a critical role for an effective HS compression. Wavelet-based techniques include the 3D extensions of JPEG2000, SPIHT, and SPECK (Kim, Xiong, & Pearlman, 2000; Penna et al., IEEE GRSL, 2006a; Tang et al. 2005). These approaches can be seen as direct 3D extensions of approaches designed for 2D imagery, where a 1D

DWT perform the spectral decorrelation, while a 2D DWT works spatially. Even if these approaches have been widely used, offering good performances (Fowler & Rucker, 2007), other approaches for spectral decorrelation have been proposed. In the literature, there exist many methods for the decorrelation of HS images in order to represent the inherent information content in lower dimensionality domain (Serpico et al., 2003), however, from a coding gain point of view, the PCA is considered to be the optimal transform for gaussian sources. In Du et al. (2007) it has been demonstrated that combining PCA with JPEG2000 yields superior rate-distortion performance than using DWT for spectral decorrelation. However, even if this approach is compliant with the Part 2 of the JPEG2000 Standard,¹ its practical application has been limited due to its high computational complexity. Possible solutions have been proposed in Du et al. (2008) and Penna et al., IGARSS (2006b), where different approaches for a low-complexity PCA have been proposed. The main advantage of using the PCA is based on the assumption that in a HS image the number of spectrally distinct signal sources is limited. The PCA transformation reproject the data in the direction of the highest variance. Consequently it is possible to assume that, after the PCA transformation, the relevant information is retained in the first few principal components (PCs) having the highest variance, while the remaining one contain essentially noise only. In this way the 1D spectral compression is obtained by considering the statistically more relevant PCs. However, since the PCA is a linear transformation, it is not able to decorrelate data presenting nonlinear correlations between variables. This will result in part of the relevant information to be retained by the last PCs and consequently to a loss of a relevant part of information by discarding them. It has been demonstrated that this loss of information can have a negative influence on the following processing stages (García-Vílchez et al., 2011). From this point of view, to be really useful, a lossy compression algorithm should be able to deal with both linear and nonlinear correlations in order to retain as much as possible of the original information in less components as possible. In other terms, a nonlinear generalization of the PCA is advisable. In the literature several approaches have been proposed to perform the nonlinear version of PCA, among which, the most effective ones are the Nonlinear Principal Component Analysis (NLPCA) (Kramer, 1991) and the Kernel Principal Component Analysis KPCA (Scholkopf, 1998). While both techniques perform a dimensionality reduction by projecting the original data into a lower dimensional feature space, only

NLPCA provides a demapping function to reproject the data into the original space. Indeed, the KPCA use the so-called kernel trick to project the data into the feature space implicit and unknown. Thus, it is not possible to reproject data back to the original space but is possible to reconstruct the data by means of minimization approaches. For this reason, the reconstructed images obtained with kernel methods present strong spectral distortions and consequently, the use of KPCA for compression purposes is discouraged. In this paper, in order to reduce the loss of information derived from the dimensionality reduction, we propose the use of the NLPCA to project the original data into a reduced dimensionality subspace (or feature space) by extracting meaningful components while still retaining the structure of the raw data. The proposed method will be evaluated in terms of compression and distortion performances. The remainder is organized as follows. In section 2, we will present the NLPCA technique, while in section 3 experimental results will be presented. Finally, we make some concluding comments in Section 4.

Nonlinear principal component analysis

NLPCA, commonly referred to a nonlinear generalization of the standard PCA, is based on multi-layer perceptrons (MLP) and is performed by Autoassociative Neural Networks (AANN) or as auto-encoder networks (Bishop, 1995; Kramer, 1991). An AANN is a conventional feedforward NN having sigmoidal activation functions in each node,

$$\sigma(x) = 1/(1 + e^{(-x)}) \quad (1)$$

with $\sigma(x) \rightarrow 1$ as $x \rightarrow +\infty$ and $\sigma(x) \rightarrow 0$ as $x \rightarrow -\infty$. The NN is trained by the Standard Conjugated Gradient (SCG) (Moller, 1993) in order to minimize the sum-of-squares error of the form:

$$E = \frac{1}{2} \sum_{n=1}^N \sum_{k=1}^d \{y_k(x^n) - x_k^n\}^2 \quad (2)$$

where y_k ($k = 1, \dots, d$) is the output vector. Differently from the standard NN topology, the nonlinear AANN sees the use of three hidden layers, including an internal bottleneck layer of smaller dimension than either input or output. The network is trained to perform identity mapping, where the input Y has to be equal to the output Y' . This means that if the training phase finds an acceptable solution, that is, a solution that gives an error E below a predefined threshold, a good compressed representation of the input must exist in the bottleneck layer. Since there are fewer nodes in the bottleneck layer

¹Jpeg2000 part 2—extensions, document ISO/IEC15444-2, www.iso.org.

than in the input/output, the bottleneck nodes must represent or encode the information obtained from the inputs for the subsequent layers to reconstruct the input. In other words, data compression caused by the network bottleneck may force hidden units to represent significant features in the data.

Nonlinear AANN topology

The ability of an AANN to fit arbitrarily nonlinear functions depends on the presence of a hidden layer with nonlinear nodes. Without the hidden layer the network is only capable of producing linear combinations of the inputs, given linear nodes in the output layer. This can be explained by considering the AANN as a combination of two successive functional mappings. The first part represents the encoding or extraction function:

$$T = G(Y) \quad (3)$$

that projects the original d -dimensional data Y onto a lower dimensional subspace defined by the activations of the units in the central hidden layer (bottleneck). In a similar way, the second half of the network defines an arbitrary functional mapping:

$$Y' = H(T) \quad (4)$$

that project from the lower dimensional space back into the original d -dimensional space (Figure 1). In general, the output of each node is defined as the compositions and superposition of a single, simple linear, or nonlinear activation function f (Cybenko, 1989):

$$y_k(x) = f\left(\sum_{j=1}^N w_{kj}x_j + w_0\right) \quad (5)$$

If we consider an AANN having only one hidden layer, the coding and decoding subnets correspond to a simple input-output network. In this way, having linear activation functions, each node in the output layer is a linear combination of the input nodes:

$$y_{out}(x) = f\left(\sum_{j=1}^N w_jx_j + w_0\right) \quad (6)$$

that would correspond exactly to the linear PCA (Sanger, 1989). A subnetwork lacking a hidden layer but including sigmoidal activation functions is only capable of generating multivariable sigmoidal functions, that is, linear functions compressed into the range (0,1) by the sigmoid:

$$y_{out}(x) = \sigma\left(\sum_{j=1}^N w_jx_j + w_0\right) \quad (7)$$

To be nonlinear, each individual function requires one hidden layer with nonlinear activation functions. In this way each node of the output layer depends on the previous hidden layer:

$$y_{out}(x) = \frac{1}{1 + e^{-\sum_{i=1}^M w_i h_i(x) + w_0}} \quad (8)$$

where:

$$h_i(x) = \frac{1}{1 + e^{-\sum_{j=1}^N w_{ij}x_j + w_{0i}}} \quad (9)$$

Substituting h_i in Equation (8) we obtain:

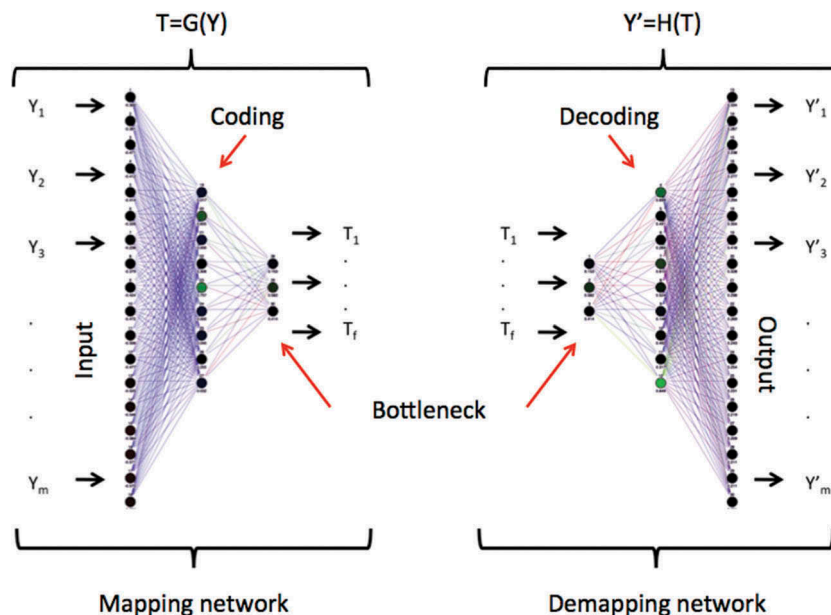


Figure 1. Networks implementing mapping (left) and demapping functions (right).

$$y_{out}(x) = \sigma \left(\sum_{j=1}^N w_j \phi_j(x) + w_0 \right) \quad (10)$$

where $\phi_j(x)$ are nonlinear basis functions that are able to approximate any continuous function. This leads to the conclusion that three hidden layers are necessary in order to obtain an optimal nonlinear feature extraction.

Nonlinear AANN training

If we consider the two subnetworks separately, for the subnetwork representing G , we know the input Y while the desired output T is unknown. On the other hand, for the network representing H , the desired output Y' is known, while the input T is not. In order to map the input space into a nonlinear feature space using a supervised training of neural networks, it is necessary to have a complete knowledge of the relations between the two spaces, and this is not always possible. This means that it is not possible to train the subnets separately in order to define a mapping function between input space and feature space. However, considering that T is at the same time output of G and input of H .

$$Y' = H(G(Y)) \quad (11)$$

is possible to combine the two networks so that the NN performing G feed directly the NN performing H . In this case there is no more need to know the relations between the input and feature spaces since both input Y and output Y' of the combined NN are known and a supervised training to learn identity mapping is now possible (Cottrell, Munro, & Zipser, 1986). With this configuration the AANN, and consequently the two subnets composing it, can be trained to minimize Equation (2). In the same way, E measures the loss of information in the same sense as PCA.

One of the main difficulties in designing the AANN relies in the selection of the correct number of nodes that minimizes the loss of information produced in the three hidden layers, and in particular in the bottleneck layer. Being the AANN designed in order to minimize the reconstruction error, the best NN topology can be retrieved by using a simple grid search algorithm that varies recursively the number of nodes of the hidden layers and evaluated the respective error. Then the topology presenting the lowest error is selected. However, without a starting point, this approach can be extremely time consuming and a different solution should be found. Analyzing the structure of the AANN, it can be found that the number of adjustable parameters in this kind of networks is:

$$(M_1 + M_2)(d + k + 1) + d + k \quad (12)$$

where M_1 and M_2 are the numbers of nodes in the mapping and demapping layers, respectively. The term d represents the number of nodes of the input/output layers while the term k refers to the number of nodes in the bottleneck layer. Equation (12) implies the following inequality:

$$M_1 + M_2 \ll d \frac{n - k}{d + k + 1} \quad (13)$$

where n is the number of training samples. Equation (13) greatly reduces the number of different AANN configurations to be tested. Aim of a dimensionality reduction method is to reduce the original spectral dimension into a lower dimensional space. This can be translated into the AANN structure as:

$$k \ll d, n \quad (14)$$

Once selected the number of nodes in bottleneck layer, that corresponds to the number of nonlinear principal components (NLPCs), then Equation (13) becomes:

$$M_1 + M_2 \ll n \quad (15)$$

Assuming a balanced structure of the AANN, M_1 and M_2 should have the same dimensions ($M_1 = M_2 = M$), and Equation (15) can be simplified as:

$$2M \ll n \quad (16)$$

Equation (16) is effective only if the number of mapping/demapping nodes M is greater than the number of nodes in the bottleneck layer k , on the contrary there will not be enough data to effectively extract k nonlinear components (Kramer, 1991).

NLPCA applied to hyperspectral images

The NLPCA can be used to project HS images into a lower dimensional feature space. The training of the AANN can be performed by using the pixels of the image, where each band corresponds to one input of the network. It has to be noted that, as the output has to simply replicate the input, no independent target data are provided, and there is no need to have an a priori knowledge for the implementation of the learning phase. This implies that the AANN training can be performed in a fully automatic way and that all pixels in the image can be considered for this task, which has actually been the technique adopted in this paper.

Once trained the AANN to perform the identity mapping, the NLPCs can be obtained directly from the bottleneck layer. In the same way, the obtained NLPCs can be subsequently used as input to the decoding layer in order to obtain the reconstructed original data. This means that if the training phase of the AANN finds an acceptable solution, the obtained

NLPCs present the same information of the input data but in a lower dimension. Thus the NLPCA permits to compress the spectra of the HS images in few components. Moreover, since the goal of the supervised training of the AANN is identity map, it is possible to suppose that the training error E is associated to the noise. This suggests that the use of the NLPCA permits to suppress or completely remove noise and artifacts present in the image.

Noise suppression and spectral dimensionality reduction can be obtained also with linear PCA. However, compared to linear decorrelation techniques, NLPCA has many advantages. First of all, with the PCA or similar approaches the information content is firstly reprojected onto an orthogonal space and then the obtained components are ordered in terms of variance. Then the image dimensionality reduction through PCA can be obtained by discarding the less relevant components in terms of variance. Since this kind of approaches detect only linear correlations among spectral bands, a relevant part of the original information can be retained by the last components and consequently lost during the compression phase (Licciardi et al. 2011). From this point of view, the NLPCA approach has the advantage to directly compress the original information into an already defined number of components, allowing an almost perfect reconstruction of the original data (Licciardi, Khan, et al., 2012). This characteristic also has great impact on the distribution of information among components. In fact, differently from other approaches, in NLPCA the nonlinear components are not ranked in term of variance. This means that the compressed information tends to be distributed among the components (Licciardi, Del Frate, & Duca, 2009).

As stated before, the PCA can be obtained using an AANN without the coding and decoding layers. In this way it is possible to avoid discarding less relevant information and project all the information in few components. However, the main difference between PCA and NLPCA obtained using AANN is that the latter is able to map both linear and nonlinear relations between variables, while PCA is only able to deal with linear ones. This means that, if nonlinear correlations exist between variables, NLPCA has the relevant advantage to describe the data with greater accuracy in fewer components than PCA.

Many methods have been proposed to extract component in a nonlinear manner, for example, locally linear embedding (LLE) and Isomap (Saul et al. 2004; Tenenbaum, Silva, & Langford, 2000) visualize high-dimensional data by projecting (embedding) them into a two or three-dimensional space. Principal curves and self-organizing maps (SOM) (Kohonen, 2001) describe data by nonlinear curves and nonlinear planes up to two dimensions. The main limitation of these methods is

related to obtaining low number of features, that may be not sufficient to describe the inherent information of the data. An alternative solution to NLPCA can be offered by the Kernel Principal Components Analysis (KPCA) (Scholkopf, Smola, & Muller, 1998). In KPCA the original data are firstly mapped into a higher dimensional feature space F , and then PCA is performed in F to extract nonlinear PCs of the input data. Due to the high computational complexity, the mapping into a higher feature space can be exploited by applying the “kernel trick” method. The “kernel trick” in machine learning is a way to easily adapt linear algorithms to nonlinear situations. In the case of KPCA, the kernel trick permits to project the input data into a higher dimensional implicit feature space F without having to compute the mapping explicitly. Similarly to PCA, the dimensionality reduction is performed by discarding the less relevant components. Both KPCA and NLPCA methods could be considered as a nonlinear generalization of the standard PCA and tend to produce similar results in terms of feature space. However, being the feature space F implicit and unknown, is not always possible to find the exact demapping function from F to the original data space (Mika et al., 1999). The reconstructed data can be obtained by minimizing the reconstruction error in F with gradient descent method. Thus, the results obtained with this approach are quite far from the optimal solution, presenting high amount of spectral distortion and will not be considered in this paper.

In the literature, NPCA has been proposed as an effective instrument for dimensionality reduction and decorrelation of different types of RS data. In Licciardi & Del Frate (2011) and Licciardi, Marpu, et al. (2012) NLPCA has been used to reduce the dimensionality of different HS images, while in Licciardi & Del Frate (2011), it has been demonstrated the effectiveness of NLPCA on other feature extraction approaches. Finally, in Licciardi et al. (2015), the NLPCA has been used to effectively remove noise from HS data.

Experiments

In this work, we will use the NLPCA approach described above to process different HS images related to the main fields of application of spectroscopy. The selected images comprise Earth observation images, biological analysis images as well as X-ray-microscope images. These datasets present different characteristics in terms of spectral range, spatial/spectral resolution, acquisition mode, and type of noise. In particular we selected:

- a satellite-borne image, featuring different kind of noises, mainly produced by the sensor and the atmosphere;

- an airborne image, where the effect of the atmosphere can be considered not present, thus presenting only noise from the sensor;
- a laboratory image that did not present relevant amount of any kind of noise;
- an image acquired with a scanning electron microscope (SEM), featuring no correlation between bands.

Statistical information of the abovementioned data are reported in Table 1.

It is important to highlight that, if properly trained, the same AANN can be used for any images acquired from the same sensor. This means that is not necessary to train a new AANN for each image to be compressed.

For each experiment, in order to satisfy Equation (16), we trained an AANN using about 60% of the pixels available (randomly selected) for each image. Once trained we used the coding network to extract the NLPCs from the bottleneck layer and we evaluated the compression ratio obtained. The performance of the dimensionality reduction method is determined in terms of rate and distortion. Rate essentially measures the percentage or the amount of compression that can be achieved. In this case the compression ratio is directly related to the number of nonlinear principal components and is expressed in terms of bits per pixel per band.

The nonlinear components are then used as input to the decoding network reconstructing the original data. The quality of the reconstructed image is evaluated in terms of distortion that can be defined as the fidelity of the reconstructed data to the original data. In this study we evaluated the distortion in terms of SNR, as defined in (Fowler, 2007):

$$SNR = 10 \log_{10} \frac{\sigma^2}{MSE} \quad (17)$$

where σ^2 is the variance of the original image while MSE is the means square error (MSE) between the original and the reconstructed image. However, in the case of real images, noise-free references may not be available. Thus, the SNR can be derived as the ratio between the mean value of the pixels in the image, and the standard deviation of the pixels of a uniform area in the image:

$$SNR = 10 \log_{10} \frac{\mu_{signal}}{\sigma_{noise}} \quad (18)$$

Table 1. Max, min, mean, and standard deviation average values of the datasets used in the experimental section.

Dataset	Min	Max	Mean	Std. Dev.
Rosis	0	8000	1385	720
Hyperion	0	0.549	0.1607	0.0987
Hypex	0	0.6593	0.2025	0.1153
SEM	0	255	82	76

Ideally, a perfect reconstruction will not remove noise from the image, thus would not change the SNR. However, since the NLPCA tends to remove noise and artifacts from the image, it is expected to have an improvement of the SNR of the reconstructed image. A further analysis has been exploited by measuring the spectral distortion introduced by the compression. This has been obtained by means of the Spectral Angle Mapper (SAM) algorithm, that measures the spectral distance between the reconstructed image and the original one:

$$SAM(X, X') = \arccos \left[\frac{X \cdot X'}{\|X\| \|X'\|} \right] \quad (19)$$

SAM will produce positive values with an ideal value of 0. However, due to noise suppression, values that are lower than 3 are referred to a good reconstructed image.

ROSIS

A first experiment has been performed using an airborne data set acquired by the ROSIS sensor over the University of Pavia, Italy (Figure 2). Covers the 0.43 μ to 0.86 μ spectrum with 103 bands having a radiometric resolution of 14 bit. The image represents a measure of the radiance and being acquired from an aircraft, the atmospheric contributions noise can be considered not relevant. The image is mainly affected by additive noise in the first bands of the detected spectrum, mainly due to the sensitivity of the detectors. Moreover, the image is also affected by the so-called smile effect, that is a common artifact to pushbroom-type sensors to which ROSIS belongs. This effect is caused by optical distortions onto the



Figure 2. RGB image of the ROSIS dataset acquired over the University of Pavia.

spatial/spectral detector array which make the instrument spectral response nonuniform for the cross-track dimension. The consequence of this effect is that the central wavelength of a band varies with spatial position across the width of the image in a smoothly curving pattern. The smile effect producing an attenuation of the detected signal can be considered as nonlinear noise (Mouroulis, Green, & Chrien, 2000). The effect of the smile is usually strong on the bands around atmospheric absorption (760 nm), however it can't be detected simply by inspecting the spectral bands. The most popular technique to detect if an image is affected by the smile is to analyse the first components of the Minimum Noise Fraction (MNF) (Dadon, Ben-Dor, & Karnieli, 2010).

The image has 340×610 pixels, and 120,000 samples have been used to train the AANN. In this experiment, we evaluated several configurations of the bottleneck layer in order to detect the best trade-off between SNR and compression. However, in order to concentrate our attention on the compression side, we analyzed only the performances obtained varying the number of nodes in the

bottleneck. For this reason we choose in accordance with Equation (16) the same number of nodes ($M = 30$) in the coding and decoding layers, for all the different configurations.

The training phase of the different AANNs has been performed using 60% of the 120,000 pixels in the image. For each AANN, the training has been considered complete when the sum of square error expressed in Equation (2) is minimized. The training performance has been evaluated in terms of MSE computed on the complete dataset.

Figures 3 and 4 report the SAM and SNR values for the PCA and NLPCA approaches as the number of component varies. Analyzing the distortion performances of PCA and NLPCA reported in Table 2 it can be noted that NLPCA reaches the best trade-off between distortion and compression with 4 components, while PCA saturates the SNR and SAM with just three components, but with a quality of the reconstructed image comparable to the one obtained using NLPCA. This can be explained by the non completely linear behavior of the investigated image. In particular, a part of information relevant for the

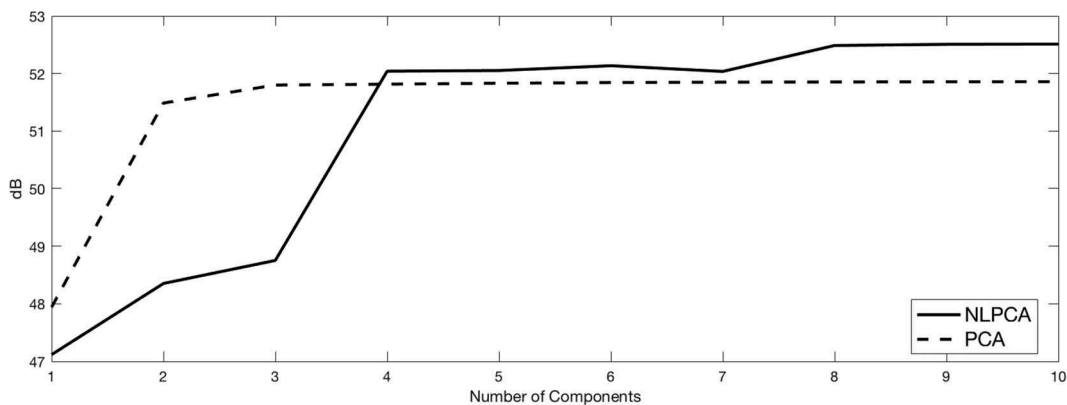


Figure 3. ROSIS dataset: Distortion rate obtained with NLPCA and PCA, measured as SNR (dB) expressed in terms of bits per pixel per band (bpppb).

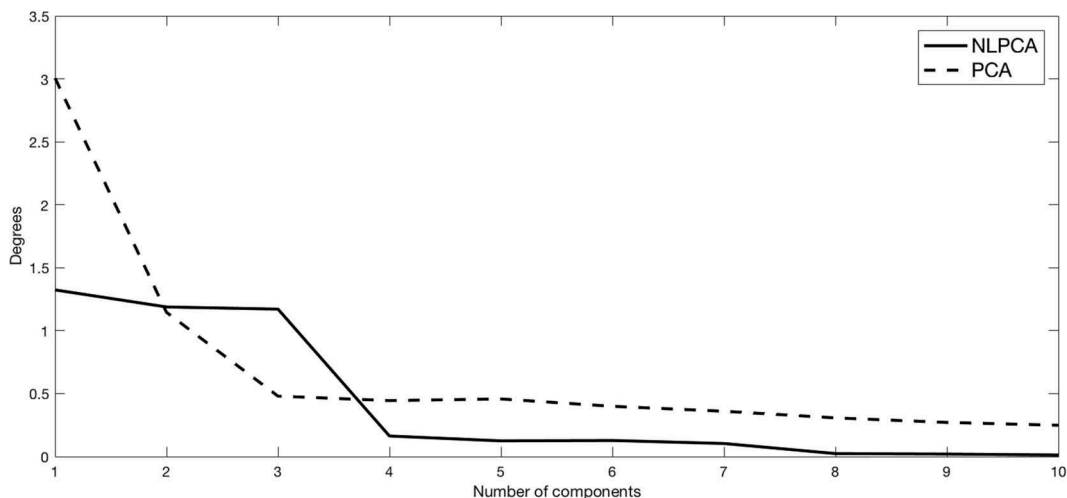


Figure 4. ROSIS dataset: SAM values for the NLPCA and PCA methods measuring the spectral distortion expressed in terms of bits per pixel per band (bpppb).

Table 2. Target bit-rate, expressed in bits per pixel per band (bpppb), SNR and SAM angle as the number of nodes in the bottleneck layer changes for the ROSIS dataset.

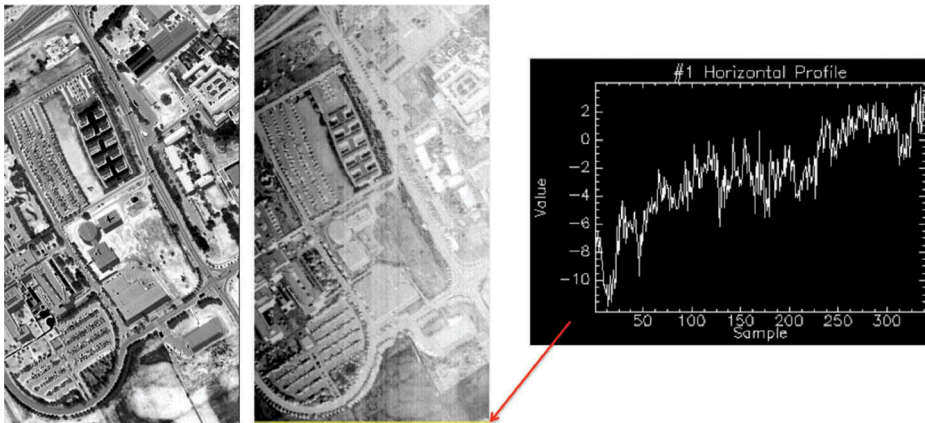
N. of components		1	2	3	4	5	6	7	8	9	10
Target bit rate (bpppb)		0.6214	1.2427	1.8641	2.4854	3.1068	3.7282	4.3495	4.9709	5.5922	6.2136
NL PCA	SNR (dB)	47.1166	48.3528	48.7528	52.0402	52.0517	52.1352	52.036	52.4867	52.5066	52.5111
	SAM (°)	1.3237	1.1883	1.1711	0.1634	0.1248	0.1278	0.1039	0.0238	0.01993	0.01275
PCA	SNR (dB)	47.9403	51.4869	51.7996	51.8144	51.8311	51.8431	51.849	51.8528	51.8556	51.8578
	SAM (°)	3.006	1.1455	0.48	0.4447	0.4578	0.3993	0.36	0.3073	0.2713	0.2486

reconstruction is retained in the components presenting the lowest variance. This means that in order to obtain the same quality of the image reconstructed by the NLPCA method, the PCA approach requires more components.

From a qualitative point of view the two reconstructed images are very similar to the original one. On a further analysis it can be noted that, similarly to the PCA (Shettigara, 1992), also the NLPCA approach introduces an improvement in the image. This is clearly evident in Figure 5 where the band 1 of the original image and the reconstructed ones are depicted. As can be seen, the original data are

strongly affected by additive noise. On the other hand, the reconstructed image seems to not suffer from this kind of noise and the image reconstructed with the NLPCA approach seems to be sharper than the image reconstructed with the PCA approach.

The best performance of the NLPCA over the PCA in terms of distortion of the reconstructed image can be explained not only with the presence of nonlinearities in the original space, but part of the improvement introduced in terms of SNR by the NLPCA technique can be addressed to the suppression of the smile effect. Analyzing the MNF components derived from the reconstructed images (Figure 6) it

**Figure 5.** ROSIS dataset: Detail of band 1 of the original (left) and the best reconstructed images obtained with NLPCA approach (four components, center) and PCA approach (four components, right), respectively.**Figure 6.** ROSIS dataset: Seventh MNF components obtained from the images reconstructed using 10 nonlinear (left) and linear (right) principal components.

is possible to note that the smile effect for the image reconstructed with the PCA is visible in the seventh component (in the original the smile was evident in the fourth component), corresponding to slight attenuation of the artifact. On the other hand, there is no evidence of smile in any MNF components obtained from the image reconstructed with the NLPCA. Thus the NLPCA tends to suppress also the smile effect contribution and results in higher performance in terms of noise filtering.

Hyperion

In a second experiment, we applied the proposed method to a Hyperion image acquired in 2008 over the Campi Flegrei area, North-West of Naples, Italy. Hyperion features 242 bands from 0.4 μm to 2.5 μm with a radiometric resolution of 12 bit. Differently from the previous experiment, in this case the image has been acquired from a satellite. This means that the atmospheric contribution has a relevant role in terms of noise of the image. For this reason atmospheric correction has been applied and the data converted into reflectance at ground level. Similar to the ROSIS instrument, Hyperion is a pushbroom type sensor, meaning that it could be affected by the smile effect. Moreover, the Hyperion instrument is characterized by poorly calibrated detectors, that is the result of small variations in the gain of each column of detectors. These detectors cause high frequency errors in the VNIR or SWIR regions, which can be identified as vertical strips in the image bands. These stripping errors can affect the mean and standard deviation of the data values for particular Hyperion band. Before compressing the Hyperion image a pre-processing step to remove the most noisy bands not

containing relevant information has been performed on the original dataset, resulting in 155 spectrally unique and good quality bands (Datt, Vicar, Niel, Jupp, & Pearlman, 2003).

The considered image consists of 100×100 pixels, and also this time we trained the AANNs using 60% of all the pixels from the image. Also in this experiment we evaluated the performances of the NLPCA dimensionality reduction method by varying only the number of nodes in the bottleneck. In particular, the number of nodes in the coding/decoding layers has been chosen to be 50 for all the configurations. Similarly to the previous experiment, the rate-distortion of the NLPCA dimensionality reduction method has been evaluated in terms of SNR and spectral distortion and compared with the performance of the standard PCA. Figures 7 and 8 report the SNR and SAM values as the number of components varies for PCA and NLPCA, respectively. Analyzing these curves it can be noted that the best tradeoff between spectral compression and quality of the reconstructed image is obtained with nine components. Analyzing the values reported in Table 3 it is also evident that with the PCA approach it is possible to obtain a similar quality with 30 components.

In Figure 9 the reconstructed images, obtained using only nine components for PCA and NLPCA, are depicted. As expected, both PCA and NLPCA performed as noise filters for the input image. However, a further analysis can be carried out by inspecting the bands where both linear and nonlinear noise is present. More in particular, while in band 1 only additive noise is present, band 88 is affected also by the striping. Thus, due to its linear nature, PCA is not able to filter all the noises present in band 88, as it is possible with NLPCA.

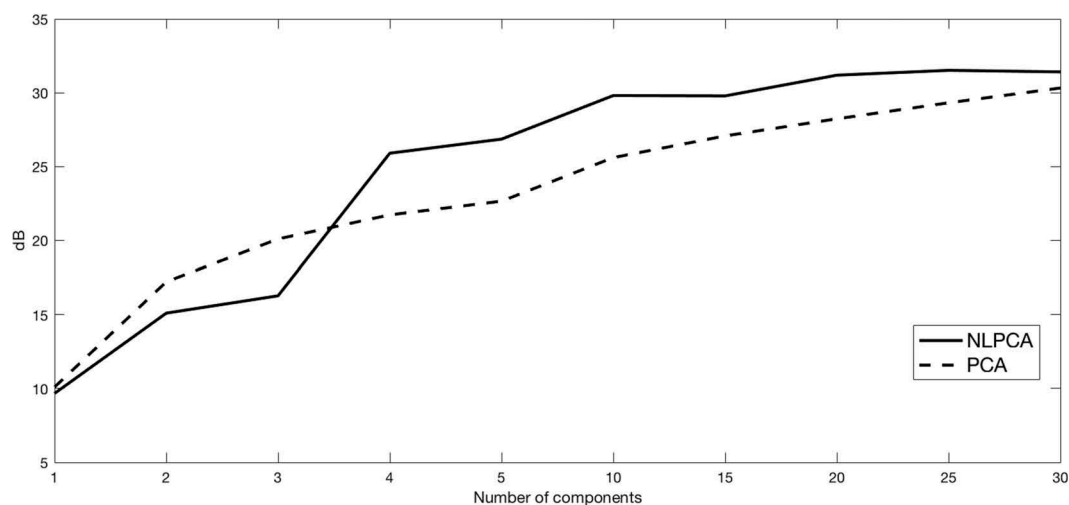


Figure 7. Hyperion dataset: Distortion rate obtained with NLPCA and PCA, measured as SNR (dB) expressed in terms of bits per pixel per band (bpppb).

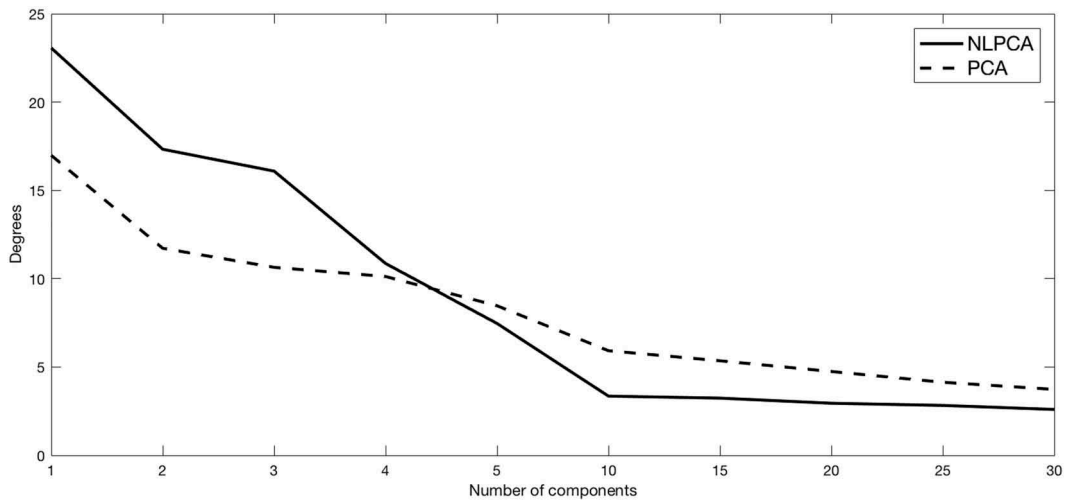


Figure 8. Hyperion dataset: SAM values for the NLPCA and PCA methods measuring the spectral distortion obtained as the number of components changes.

Table 3. Target bit-rate, expressed in bits per pixel per band (bpppb), SNR and SAM angle as the number of nodes in the bottleneck layer changes for the Hyperion dataset.

N. of components	1	2	3	4	5	10	15	20	25	30	
Target bit rate (bpppb)	0.4129	0.8258	1.2387	1.6516	2.0645	4.1290	6.1935	8.2581	10.3226	12.3871	
NL PCA	SNR (dB)	9.6663	15.1024	16.2748	25.9095	26.8618	29.8141	29.788	31.1918	31.5197	31.4154
	SAM (°)	23.062	17.328	16.0905	10.8611	7.4664	3.3512	3.2381	2.9487	2.8255	2.5979
PCA	SNR (dB)	10.0958	17.2185	20.105	21.734	22.6621	25.6123	27.079	28.2384	29.3227	30.3227
	SAM (°)	16.982	11.7237	10.642	10.1242	8.4702	5.9231	5.3539	4.748	4.1366	3.7322

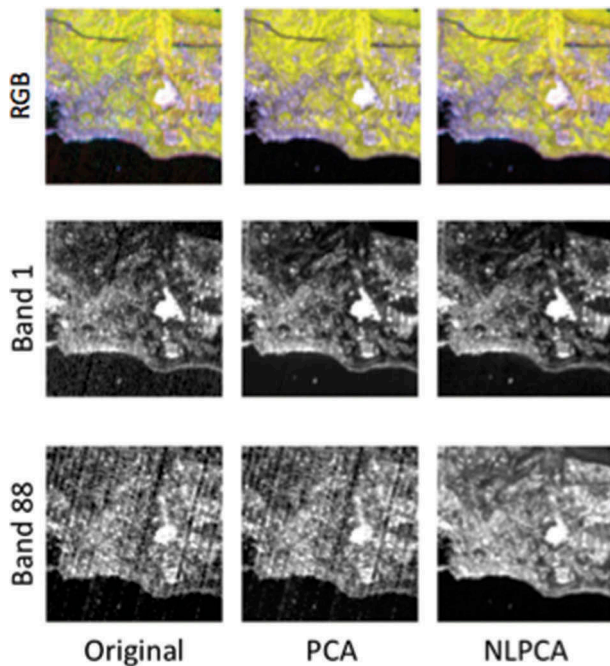


Figure 9. Hyperion dataset: Detail of bands 1 and 88 of the original Hyperion dataset (left), the PCA-reconstructed (center) and the NLPCA-reconstructed (right) images, respectively.

Hypex

A third experiment has been conducted using a field HS image. In particular we analyzed the performance of the NLPCA dimensionality reduction on a Hypex image acquired on the field for the biochemical

analysis of the vegetation. The considered Hypex is a pushbroom sensor with a spectral coverage within 0.4–1.0 μm , for a total of 160 bands having radiometric resolution of 12 bit. Before being processed the image has been converted to reflectance. In this case, the sensor has been mounted on a tripod very close to the target vegetation (1m). The static nature of the support reduced to the minimum the amount of noise present in the image. Moreover, the proximity of the target resulted in the almost total absence of any nonlinear atmospheric contribution. For these reasons, this dataset is expected to be almost completely linear.

For the evaluation we considered a part of the image, consisting of 364×365 pixels. In this case the AANNs have been trained using only a small set of randomly selected pixels (9395). A grid search approach has been exploited in order to find the best number of nodes in the coding/decoding layers. 100 nodes have been found to be sufficient to extract all the different components. Similarly to the previous experiments, the analysis of the training performance as the number of nodes in the bottleneck layer changes, expressed in terms of MSE, indicates that two NLPCs are sufficient for retaining all the relevant information.

Figures 10 and 11 report the SNR and SAM values, respectively, as the number of components changes for NLPCA and PCA, respectively. From the analysis of the values reported in Table 4 it can be noted that

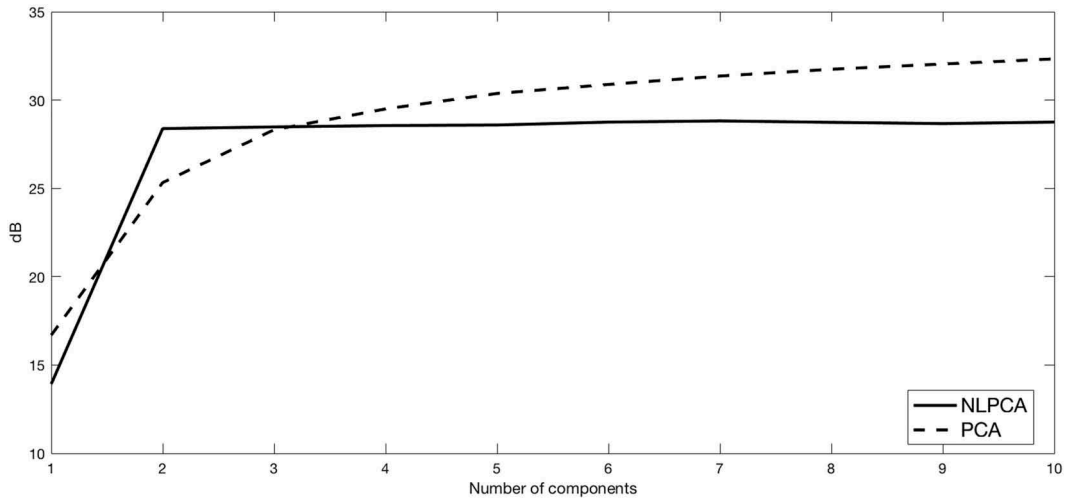


Figure 10. Hypspx dataset: Distortion rate obtained with NLPCA and PCA, expressed in terms of SNR (dB), as the number of components changes.

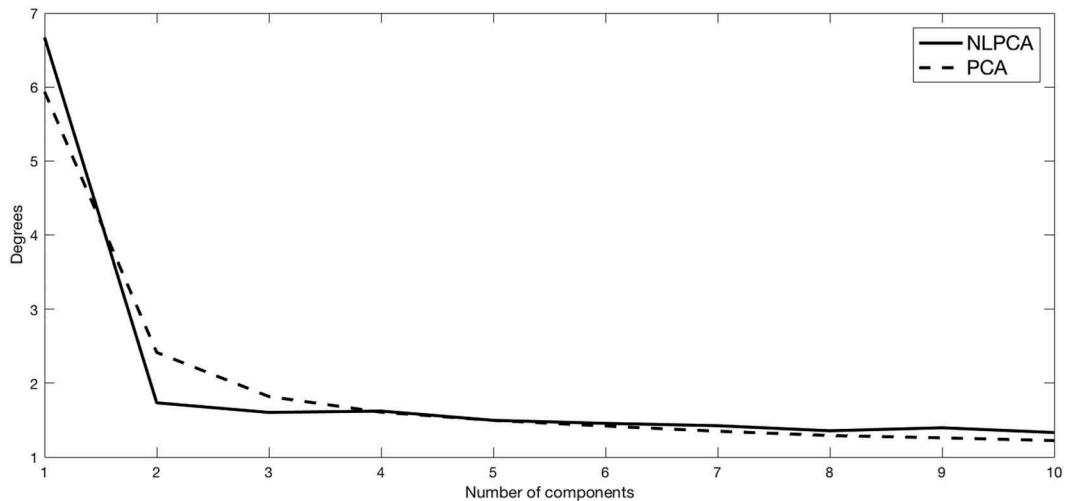


Figure 11. Hypspx dataset: SAM values for the NLPCA and PCA methods measuring the spectral distortion obtained as the number of components changes.

Table 4. Target bit-rate, expressed in bits per pixel per band (bpppb), SNR and SAM angle as the number of nodes in the bottleneck layer changes for the HYSPEX dataset.

N. of components	1	2	3	4	5	6	7	8	9	10
Target bit rate (bpppb)	0.4	0.8	1.2	1.6	2.0	2.4	2.8	3.2	3.6	4.0
NL PCA										
SNR (dB)	13.9248	28.3841	28.4801	28.5573	28.5878	28.7536	28.820	28.7402	28.6677	28.7554
SAM (°)	6.6698	1.7341	1.6043	1.6228	1.4978	1.4567	1.4251	1.3564	1.3971	1.3327
PCA										
SNR (dB)	16.6928	25.3289	28.3198	29.5001	30.3739	30.8837	31.3594	31.7443	32.0421	32.3321
SAM (°)	5.9352	2.4155	1.8192	1.6079	1.4965	1.4203	1.3504	1.292	1.2611	1.2239

for both methods the best tradeoff between quality of the reconstructed image and spectral compression is obtained with 2 components. Aside from slight differences, both methods offer similar performances in terms of compression and reconstruction, confirming the initial hypothesis of an almost completely linear image. This is also confirmed by comparing the two principal components obtained with the two methods. Figure 12 reports the scatter plots obtained comparing PC1 with NLPC1 and PC2 with NLPC2. In both cases the components present high degrees of

correlation. Figure 13 reports the RGB representations of the original image, and the reconstructed images obtained using six components for the PCA and NLPCA methods, respectively. On a qualitative analysis there is no relevant difference between the original image and the reconstructed ones.

This experiment shows that in case of images that are mainly characterized by linear correlations, both PCA and NLPCA obtain similar results. This because, referring to Equation (10), linear basis functions can be considered as a subset of $\phi_j(x)$, demonstrating that

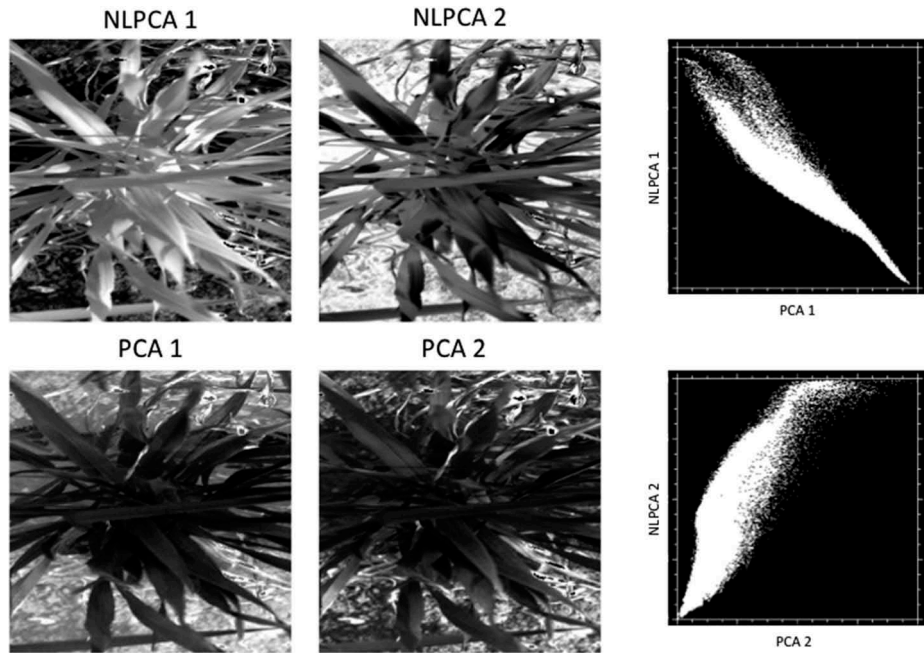


Figure 12. Hypspec dataset: The first two components obtained with the PCA and NLPCA transformation. The components present correlation as reported on the scatter plots on the right part of the picture.

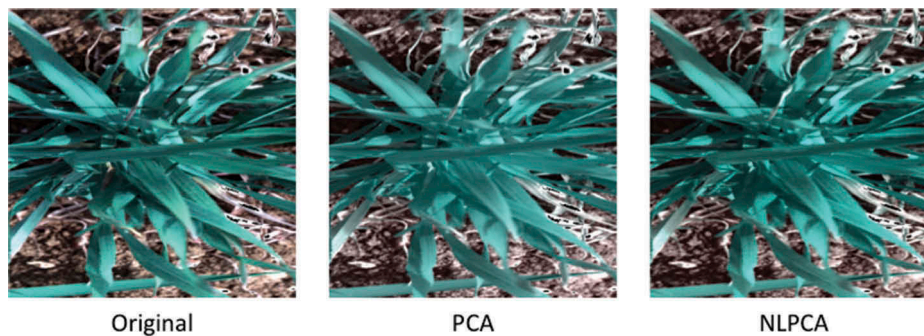


Figure 13. Hypspec dataset: RGB images obtained combining bands 60 (R), 40 (G) and 30 (B) of the original dataset (left), of the reconstructed ones obtained using 6 components for PCA (center) and NLPCA (right) approaches, respectively.

NLPCA is able to manage both linear and nonlinear correlations between variables

Scanning electron microscope

To further appreciate the effectiveness of the proposed method, a last experiment is conducted on a different kind of image. In this experiment we applied the NLPCA technique to an image acquired by a scanning electron microscope. This kind of microscope is based on energy-dispersive X-ray spectroscopy (EDX) and investigate the interaction of X-ray and target samples by measuring it with a energy-dispersive spectrometer (Goldstein et al., 2003). The energy-dispersive detector permits to separate the characteristic x-rays of different elements into an energy spectrum. It is possible to compose a dataset of images, each of them related to the energy of a specific element. For this experiment we considered an image representing a seed in

water. The dataset was composed by 15 bands digitalized at 16 bit, the first 14 representing the energy values of the following minerals: Al, Ca, Cb, Cl, Fe, K, Mg, Mn, Na, O, P, Si, So, Ti, and a fifteenth band obtained measuring the backscatter electron values of the surface of the sample. Representing the electromagnetic response of different elements, these 15 bands result to be extremely uncorrelated. Moreover, most of them are also noisy, as depicted in Figure 14. Aim of this experiment is to demonstrate the ability of NLPCA to compress images where the bands are poorly correlated with the others.

In this case we choose to train the different AANNs using 20 nodes in the coding/decoding layers, while from the analysis of the training performance we selected 14 nodes in the bottleneck layer.

On a quantitative analysis, Figures 15 and 16 and Table 5 report the SNR and the SAM values

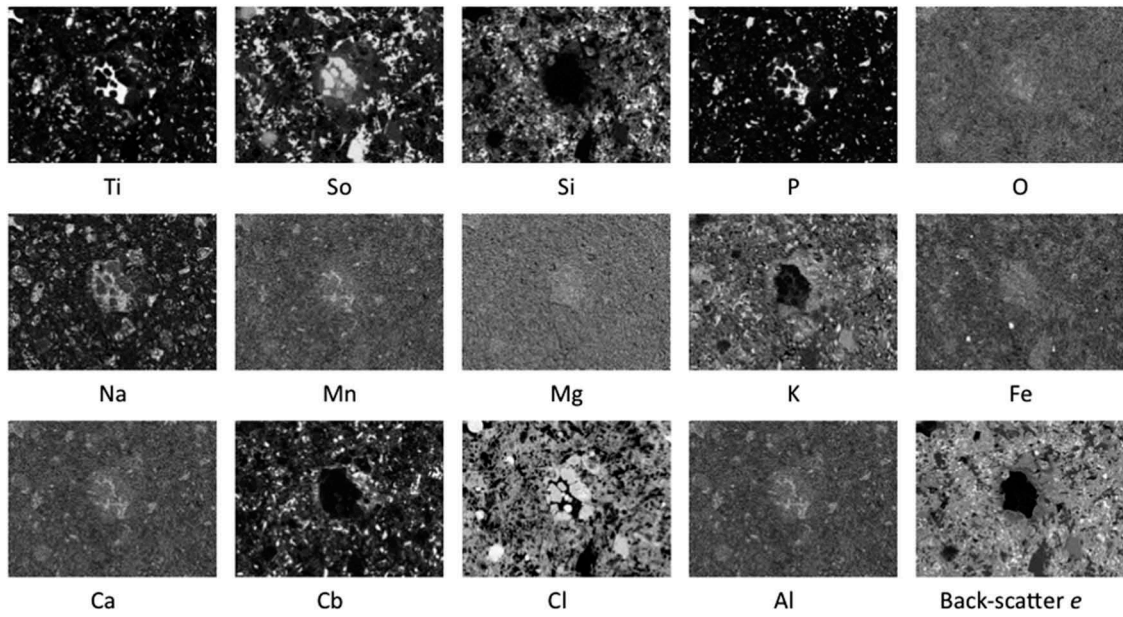


Figure 14. SEM dataset: 15 bands acquired at the energy values related to different elements.

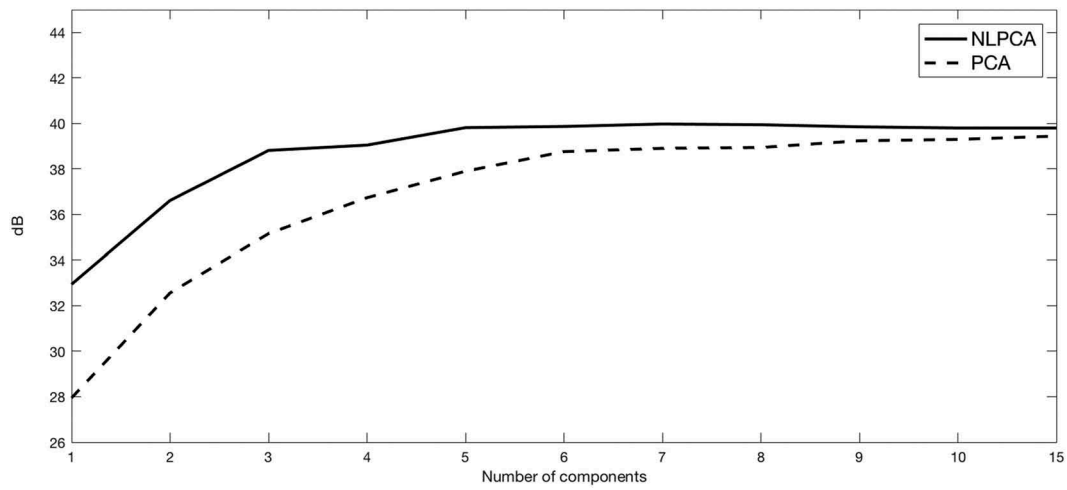


Figure 15. SEM dataset: Distortion rate obtained with NLPCA and PCA, expressed in terms of SNR (dB), as the number of components changes.

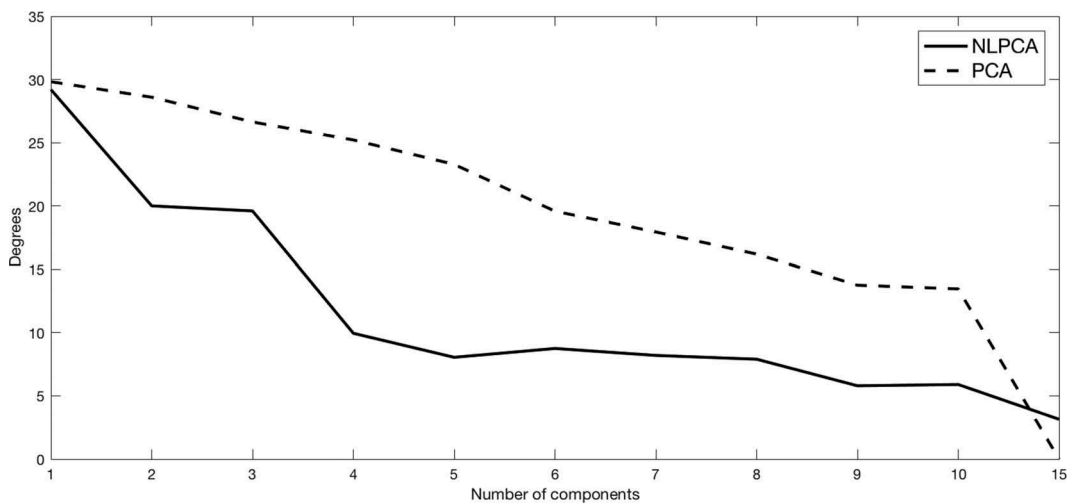


Figure 16. SEM dataset: SAM values for the NLPCA and PCA methods measuring the spectral distortion obtained as the number of components changes.

Table 5. Target bit-rate, expressed in bits per pixel per band (bpppb), SNR and SAM angle as the number of nodes in the bottleneck layer changes for the SEM dataset.

N. of components	1	2	3	4	5	6	7	8	9	10	15
Target bit rate (bpppb)	4.2667	8.5333	12.8	17.0667	21.3333	25.6	29.866	34.1333	38.42	42.6667	64
NL PCA											
SNR (dB)	32.943	36.6146	38.813	39.0478	39.8155	39.8683	39.975	39.9432	39.851	39.8001	39.7994
SAM (°)	29.2	20	19.6	9.95	8.05	8.75	8.2	7.9	5.8	5.9	3.15
PCA											
SNR (dB)	27.9492	32.5649	35.159	36.7354	37.9011	38.7614	38.9065	38.9432	39.2389	39.2981	39.4434
SAM (°)	29.821	28.5923	26.6353	25.2159	23.2841	19.5785	17.939	16.1927	13.7499	13.4586	0

of the proposed method compared to the PCA based one, respectively. As it can be noted, the NLPCA based method is able to obtain a good tradeoff between compression and distortion with four components. On the other hand, even if PCA is able to reach a good SNR with six components, it still requires almost all the principal components to get also a good spectral reconstruction.

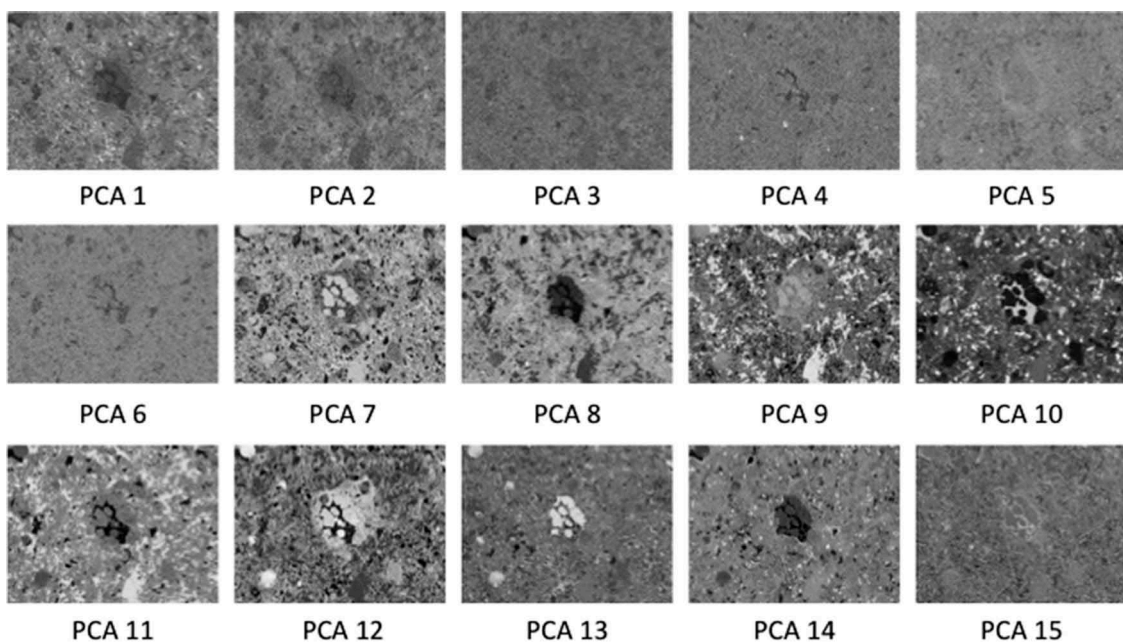
Analyzing the SAM values reported in Table 5 it is possible to note that for 15 components the SAM value for the PCA method is 0, corresponding to a perfect reconstruction of the original image. On the other hand, the SAM value for the NLPCA approach is higher (3.15°), suggesting a moderate distortion between the original image and the reconstructed one. This means that while PCA preserved entirely the original information expressed as a combination of signal and noise, NLPCA was able to remove part of the noise present in the image. This can be quantitatively evaluated by analyzing Figure 17. This also suggests that since the original bands are not linearly correlated, the information is distributed among all the linear principal components, as reported in Figure 17. Thus is no more possible to discard the components with less variance without losing relevant information. This is also evident from

Figure 18, where the original image ($R = Si$; $G = So$; $B = Ti$) and the reconstructed ones, obtained both using 14 components from PCA and NLPCA, are depicted. In particular, even if most of the information is contained in the first 14 principal components, the PCA is not able to correctly reconstruct the original image. On the other hand, 14 nonlinear components are enough to correctly reconstruct the original image.

Conclusions

This paper presented a novel approach for the dimensionality reduction of HS data based on the nonlinear generalization of the standard PCA. Aim of the presented method is to preserve as much of the original spectral information as possible with a compression rate higher than those obtained using PCA-based approaches. The main advantage on using NLPCA relies on the assumption that thanks to the nonlinear functions, it permit to project in a lower dimensional space the same information content of the standard PCA but with less features.

The proposed approach has been tested both qualitatively and quantitatively on several images, featuring different types of information and

**Figure 17.** PCA components obtained from the 15 SEM bands.

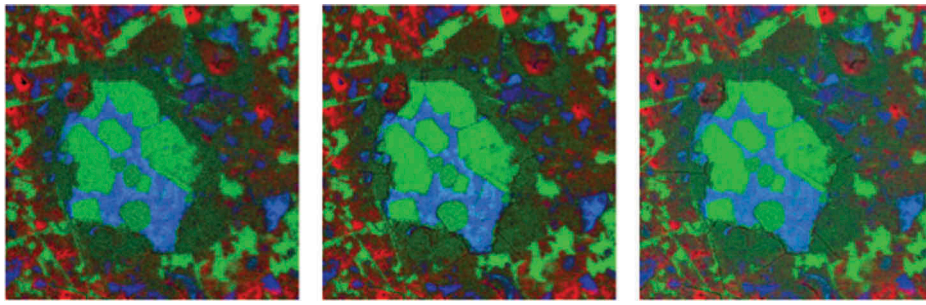


Figure 18. SEM dataset: RGB images, obtained combining Si, So and Ti for R, G and B channels, respectively, of the original SEM dataset (left), the NLPCA-reconstructed (center) and the PCA-reconstructed (right) images, respectively.

affected by different kind of noises. For each image considered, the rate-distortion obtained from the reconstruction of the image was evaluated in terms of SNR and SAM. The experiments demonstrate that the NLPCA method tends to obtain good reconstructions of the original images with less components than PCA. Only in one case, when the investigated image was not affected by any kind of relevant noise, the two methods obtained similar results.

From a computational point of view, a direct comparison between PCA and NLPCA could not be achieved. In particular, it is important to separate the time spent to estimate the projection function from the computational time necessary to project the original data into the feature space. Indeed, NLPCA requires a long time to identify the optimal topology of the AANN while PCA just need to compute the covariance matrix (or the correlation matrix). However, once defined the projection functions, both compression processes are comparable in terms of computational time.

Another important result came from the analysis of the reconstructed images. In particular, while PCA is able to effectively filter only the additive noise present in the original image, thus, enhancing the spectral information of the reconstructed one, NLPCA is able to filter both linear and nonlinear noise. The ability of the NLPCA to deal with both linear and nonlinear noises, results in reconstructed images that have a higher SNR, if compared with those obtained with PCA.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Agence Nationale de la Recherche [project APHYPIS];

References

- Abousleman, G.P. (1995). *Compression of hyperspectral imagery using hybrid and entropy-constrained trellis coded quantization*. Proceedings Data Compression Conference. IEEE Computer Society Press.
- Abrardo, A., Barni, M., & Magli, E. (2010). *Low-complexity lossy compression of hyperspectral images via informed quantization*. In IEEE International Conference on Image Processing (ICIP), Hong-Kong, 26-29 September 2010.
- Bishop, C. (1995). *Neural networks for pattern recognition*. London, UK: Oxford University Press.
- Christophe, E., Mailhes, C., & Duhamel, P. (2008). Hyperspectral image compression: Adapting spilt and ezw to anisotropic 3-d wavelet coding. *IEEE Transactions on Image Processing*, 17(12), 2334–2346. doi:10.1109/TIP.2008.2005824
- Conoscenti, M., Coppola, R., & Magli, E. (2016). Constant snr, rate control, and entropy coding for predictive lossy hyperspectral image compression. *IEEE Transactions on Geoscience and Remote Sensing*, 54(12), 7431–7441.
- Cottrell, G.W., Munro, P., & Zipser, D. (1986). Learning internal representations from gray-scale images: An example of extensional programming. *Proceedings Conference of the Cognitive Sciences SOC, Seattle, WA*, pp 461-473.
- Cybenko, G. (1989). Approximations by superpositions of sigmoidal functions. *Mathematics of Control, Signals, and Systems*, 2(4), 303–314.
- Dadon, A., Ben-Dor, E., & Karnieli, A. (2010). Use of derivative calculations and minimum noise fraction transform for detecting and correcting the spectral curvature effect (smile) in hyperion images. *IEEE Transactions Geoscience Remote Sensing*, 48(6), 2603–2612. doi:10.1109/TGRS.2010.2040391
- Datt, B., Vicar, T.R.M., Niel, T.G.V., Jupp, D.L., & Pearlman, B.J.S. (2003). Preprocessing eo-1 hyperion hyperspectral data to support the application of 30 agricultural indexes. *IEEE Transactions on Geoscience and Remote Sensing*, 41(6), 1246–1259. doi:10.1109/TGRS.2003.813206
- Du, Q., & Fowler, J.E. (2007). Hyperspectral image compression using jpeg2000 and principal component analysis. *IEEE Geoscience Remote Sensing Letters*, 4(2), 201–205. doi:10.1109/LGRS.2006.888109
- Du, Q., & Fowler, J.E. (2008). Low complexity principal component analysis for hyperspectral image compression. *International Journal of High Performance Computing Application*, 22, 438–448. doi:10.1177/1094342007088380

- Fowler, J.E., & Rucker, J.T. (2007). *3d wavelet-based compression of hyperspectral imagery, hyperspectral data exploitation: Theory and applications*. John Wiley and Sons, Hoboken, NJ.
- García-Vílchez, F., Muñoz-Marí, J., Zorteza, M., Blanes, I., González-Ruiz, V., Camps-Valls, G., ... Serra-Sagristà, J. (2011). On the impact of lossy compression on hyperspectral image classification and unmixing. *IEEE Geoscience and Remote Sensing Letters*, 8(2), 253–257. doi:10.1109/LGRS.2010.2062484
- Goldstein, J., Newbury, D.E., Joy, D.C., Lyman, C.E., Echlin, P., Lifshin, E., ... Michael, J. (2003). *Scanning electron microscopy and x-ray microanalysis*. Springer, New York.
- J. E. Fowler and J. T. Rucker, (2007). 3d wavelet-based compression of hyperspectral imagery. *Hyperspectral Data Exploitation: Theory and Applications* - John Wiley and Sons Inc.
- Jing, Z., & Guizhong, L. (2006). *A novel lossless compression for hyperspectral images by adaptive classed arithmetic coding in wavelet domain*. In IEEE International Conference on Image Processing (ICIP), Atlanta, GA.
- Kaarna, A., & Parkkinen, J. (2000). Comparison of compression methods for multi-spectral images, *Proceedings NORSIG Nordic Signalling Processing Symposium Kolmarden Sweden*, 2, 251–254.
- Karami, A., Heylen, R., & Scheunders, P. (2015). *Lossy compression of hyperspectral images optimizing spectral unmixing*. In 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (pp. 5031–5034), IGARSS 2015, 26–31 July, Milan, Italy.
- Kim, B.-J., Xiong, Z., & Pearlman, W.A. (2000). Low bit-rate scalable video coding with 3-d set partitioning in hierarchical trees (3-d spih). *IEEE Transactions on Circuits and Systems for Video Technology*, 10(8), 1374–1387. doi:10.1109/76.889025
- Kohonen, T. (2001). *Self-organizing maps* (3rd ed.). Springer-Verlag Berlin Heidelberg.
- Kramer, M.A. (1991). Nonlinear principal component analysis using autoassociative neural networks. *AICHE Journal. American Institute of Chemical Engineers*, 37, 233–243. doi:10.1002/(ISSN)1547-5905
- Kulkarni, P., Bilgin, A., Marcellin, M.W., Dagher, J.C., Kasner, J.H., Flohr, T.J., & Rountree, J.C. (2006). *Compression of earth science data with jpeg2000, hyperspectral data compression*. New York: Springer-Verlag.
- Licciardi, G., & Chanussot, J. (2015). Nonlinear pca for visible and thermal hyperspectral image quality enhancement. *IEEE Geoscience and Remote Sensing Letters*, 12(6), 1228–1231. doi:10.1109/LGRS.2015.2389269
- Licciardi, G., & Del Frate, F. (2011). *Pixel unmixing in hyperspectral data by means of neural networks*, IEEE Transactions on, Geoscience and Remote Sensing, IEEE Transactions on 49 (pp. 4163–4172).
- Licciardi, G., Del Frate, F., & Duca, R. (2009). Feature reduction of hyperspectral data using autoassociative neural networks algorithms, *Geoscience and Remote Sensing Symposium. 2009 IEEE International IGARSS 2009 1*, 179, 176–595.
- Licciardi, G., Khan, M.M., Chanussot, J., Montanvert, A., Condat, L., & Jutten, C. (2012). Fusion of hyperspectral and panchromatic images using multiresolution analysis and nonlinear pca band reduction. *EURASIP Journal on Advances in Signal Processing*, 4885(8), 347–358.
- Licciardi, G., Marpu, P.R., Chanussot, J., & Benediktsson, J. A. (2012). Linear versus nonlinear pca for the classification of hyperspectral data based on the extended morphological profiles. *IEEE Geoscience and Remote Sensing Letters*, 9(3).
- Mika, S., Scholkopf, B., Smola, A., Muller, K.R., Scholz, M., & Ratsch, G. (1999). Kernel pca and de-noising in feature spaces. *Advances in Neural Information Processing Systems*, 11, 536–542.
- Moller, M.F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6, 525–533. doi:10.1016/S0893-6080(05)80056-5
- Mouroulis, P., Green, R.O., & Chrien, T.G. (2000). Design of pushbroom imaging spectrometers for optimum recovery of spectroscopic and spatial information. *Applications Optical*, 39(13), 2210–2220. doi:10.1364/AO.39.002210
- Penna, B., Tillo, T., Magli, E., & Olmo, G. (2006a). Progressive 3-d coding of hyperspectral images based on jpeg 2000. *IEEE Geoscience and Remote Sensing Letters*, 3(1), 125–129. doi:10.1109/LGRS.2005.859942
- Penna, B., Tillo, T., Magli, E., & Olmo, G. (2006b). *A new low complexity klt for lossy hyperspectral data compression*. In Proceedings of the international, Proceedings of the International Geoscience and Remote Sensing Symposium, Denver. 3525–3528.31 July - 04 August, Denver, CO.
- Sanger, T.D. (1989). *Optimal unsupervised learning in a single-layer linear feed-forward neural network, neural networks*.
- Saul, L.K., & Roweis, S.T. (2004). Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *Journal of Machine Learning Research*, 4(2), 119–155.
- Scholkopf, B., Smola, A., & Muller, K.R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5), 1299–1319. doi:10.1162/089976698300017467
- Scholkopf, B., Smola, A. J., & Muller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10, 1299–1319.
- Serpico, B., D’Inca, M., Melgani, F., & Moser, G. (2003). Comparison of feature reduction techniques for classification of hyperspectral remote sensing data universal image quality index. *Proceedings SPIE Image Signalling Processing Remote Sensing*, 4885.
- Shettigara, V.K. (1992). *A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set*. In Photogrammetric Engineering and Remote Sensing, May 1992, 58 (5), 561–567.
- Tang, X., & Pearlman, W.A. (2005). *Scalable hyperspectral image coding*. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, 19–23 March 2005, Philadelphia, PA (vol. 2), 401–404.
- Tenenbaum, J., Silva, V.D., & Langford, J. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500), 2319–2323. doi:10.1126/science.290.5500.2319
- Wei, Z., Qiwei, L., & Houqiang, L. (2010). *Distributed lossless coding of hyperspectral images*. In IEEE International Conference on Image Processing (ICIP), 9 September 2010, Hong Kong.