Dimensionality Reduction on Hyperspectral Images: A Comparative Review Based on Artificial Datas

Jihan Khodr
Laboratoire Tsi2m UPRES JE 2529
Université de Renne 1 - ENSSAT
Lannion, France

Rafic Younes
Laboratoire R.I.T.C.H.
Faculty of Engineering, Lebanese University
Beirut, Lebanon

***Abstract*— in this research we address the problem of high-dimensional in hyperspectral images, which may contain rare /anomaly vectors introduced in the subspace observation that we wish to preserve. Linear techniques Principal Component Analysis(PCA), and non linear techniques Kernel PCA, Isomap, Multidimensional scaling (MDS), Local Tangent Space Alignment (LTSA), Diffusion maps, Sammon mapping, Symmetric Stochastic Neighbor Embedding (SymSNE), Stochastic Neighbor Embedding (SNE), Locally Linear Embedding(LLE), Locality Preserving Projection(LPP), Neighborhood Preserving embedding (NPE), Linear Local Tangent Space Alignment (LLTSA) was presented. Classical approaches criterion based on the norm l_d, derivative spectral, nearest neighbors and quality criteria are used for obtaining a good preservation of these vectors in the reduction dimension. We have observed from the results obtained that Sammon and Isomap are less sensitive to these rare vectors compared to the other presented methods.**

Keywords-component; Dimensionality reduction; manifold learning; rare vectors; quality criteria;

I. INTRODUCTION

Advances in data collection and storage capabilities during the past decades have led to an information overload in most sciences. Researchers working in domains as diverse as engineering, astronomy, biology, remote sensing, economics, and consumer transactions, face larger and larger observations and simulations on a daily basis. Such datasets, in contrast with smaller, more traditional datasets that have been studied extensively in the past, present new challenges in data analysis. Traditional statistical methods break down partly because of the increase in the number of observations, but mostly because of the increase in the number of variables associated with each observation. The dimension of the data is the number of variables that are measured on each observation. High-dimensional datasets present many mathematical challenges as well as some opportunities, and are bound to give rise to new theoretical developments [33]. Ideally, the reduced representation should have a dimensionality that corresponds to the intrinsic dimensionality of the data. The intrinsic dimensionality of data is the minimum number of parameters needed to account for the observed properties of the data [25]. As a result, dimensionality reduction facilitates, among others, classification, visualization, and compression of high-dimensional data. Traditionally, dimensionality reduction was performed using linear techniques such as Principal Components Analysis (PCA) [2]. In the last decade, a large number of nonlinear techniques for dimensionality reduction have been proposed See for an overview, e.g., [35, 36, 37]. In contrast to the traditional linear techniques, the nonlinear techniques have the ability to deal with complex nonlinear data; previous studies have shown that nonlinear techniques perform their linear counterparts on complex artificial tasks. The investigation is performed by both a theoretical and an empirical evaluation of the dimensionality reduction techniques. The identification is performed by a careful analysis of the empirical results on specifically designed artificial datasets and on a selection of real world datasets.

Next to PCA, the paper investigates the following non linear techniques: (1) Kernel PCA, (2) Isomap, (3) MDS, (4) LTSA, (5) Diffusion maps, (6) Sammon mapping, (7) SymSNE, (8) SNE, (9) LLE, (10) LPP, (11) Neighborhood Preserving embedding (NPE), (12) LLTSA Although our comparative review includes the most important nonlinear techniques for dimensionality reduction. The outline of the remainder of this paper is as follows. In Section 2, we give a formal definition of dimensionality reduction and subdivide the 13 dimensionality reduction techniques into two linear technique and eleven non-linear techniques. Section 3 present and discuss the linear technique. Subsequently, Section 4 describes and discusses the eleven non-linear techniques for dimensionality reduction. Then, in Section 5, we present a definition of the quality criteria and present the results of the experiments in section 6; Section 7 provides our discussions and we finalize by the conclusions.

II. DIMENSIONALITY REDUCTION

The problem of nonlinear dimensionality reduction can defined as follows. Assume we have dataset represented in n × D matrix X consisting of n data vectors $\boldsymbol{x_i}$ for i = {1, 2, . .,n} with dimensionality D. Assume further that this dataset has intrinsic dimensionality d where d < D.

Here, in mathematical terms, intrinsic dimensionality means that the points in dataset X are lying on or near a manifold with dimensionality d that is embedded in the D-dimensional space. Dimensionality reduction techniques transform dataset X with dimensionality D into a new dataset Y with dimensionality d, while retaining the geometry of the data as much as possible. Throughout the paper, we denote a high-dimensional data point by, where $\boldsymbol{x_i}$ is the ith row of the D-dimensional data matrix X. The low dimensional counterpart of is denoted by $\boldsymbol{y_i}$, where $\boldsymbol{y_i}$ is ith the row of the d dimensional data matrix Y. We subdivide techniques for dimensionality reduction into linear and non-linear techniques. Linear techniques assume that the data lie on or near a linear subspace of the high-dimensional space. Nonlinear techniques for dimensionality reduction do not rely on the linearity assumption as a result of which more complex embeddings of the data in the high-dimensional space can be identified.

III. LINEAR TECHNIQUES DIMENSIONALITY REDUCTION

Is a linear technique [1, 2] for dimensionality reduction, which means that it performs dimensionality reduction by embedding the data into a linear subspace of lower dimensionality, PCA is by far the most popular unsupervised linear technique. Principal Components Analysis (PCA) [3] constructs a low-dimensional representation of the data that describes as much of the variance in the data as possible. This is done by finding a linear basis of reduced dimensionality for the data, in which the amount of variance in the data is maximal. In mathematical terms, PCA attempts to find a linear mapping M that maximizes cov(X) M, where cov(X) is the covariance matrix of the data X. It can be shown that this linear mapping is formed by the d principal eigenvectors (i.e., principal components) of the covariance matrix of the zero-mean data PCA maximizes cov(X)M with respect to M, under the constraint that $|M| = 1$.

Linear Local Tangent Space Alignement (LLTSA) uses the tangent space in the local neighborhood of each sample point to represent the local geometry, and the labels of the samples are not taken into account, which might greatly degrade the performance for a recognition problem. Therefore, it is necessary to develop a supervised method to enhance the performance of LLTSA. In the following section, we will propose a novel method to investigate the discriminant completely using the tangent space to represent the local geometry.

IV. NONLINEAR TECHNIQUES FOR DIMENSIONALITY REDUCTION

In part 3 we discussed the main linear technique for dimensionality reduction, which is established and well studied. In contrast, most non linear techniques for dimensionality reduction have been proposed and are therefore less well studied. In this part, we discuss eleven non linear techniques for dimensionality reduction.

Some methods such as LLE and Isomap rely on applying linear techniques on a set of local neighborhoods, which are assumed to be locally linear in nature. As such, they fall into the category of local linear dimensionality reduction techniques. Isometric Feature Mapping (Isomap) [10] is a nonlinear dimensionality reduction technique that uses MDS techniques with geodesic interpoint distances instead of Euclidean distances. Geodesic distances represent the shortest paths along the curved surface of the manifold. Isomap is a very useful no iterative, polynomial-time algorithm for nonlinear dimensionality reduction if the data is severely nonlinear. Isomap is able to compute a globally optimal solution and for a certain class of data manifolds, is guaranteed to converge asymptotically to the true structure However, Isomap may not easily handle more complex domains such as non-trivial curvature or topology and it has the same (high) complexities as MDS.

Locally Linear Embedding (LLE) [16] is a nonlinear dimensionality reduction technique that computes low-dimensional, neighborhood preserving embeddings of high-dimensional inputs. Unlike Isomap, LLE eliminates the need to estimate pairwise distances between widely separated data points and recovers global nonlinear structure from locally linear fits [16] LLE assumes that the manifold is linear when viewed locally. Compared to Isomap, LLE is more efficient. However, LLE finds an embedding that only preserves the local structure, is not guaranteed to asymptotically converge and may introduce unpredictable distortions. LLE should be able to perform well for open planar manifolds, but only if the surface is a smooth curve and the neighborhood chosen is small enough.

Local Tangent Space Alignment (LTSA) [15] is a method for nonlinear dimensionality reduction that constructs approximations of tangent spaces in order to represent local geometry of the manifold and the global alignment of the tangent spaces to obtain the global coordinate system. Based on a set of unorganized data points sampled with noise from a parameterized manifold, the local geometry of the manifold is learned by constructing an approximation for the tangent space at each data point and those tangent spaces are then aligned to give the global coordinates of the data points with respect to the underlying manifold [7] In general, LTSA is less sensitive to the choice of k neighborhoods, as compared to LLE. Although LTSA as well as the other nonlinear dimensionality reduction algorithms are able to handle nonlinearities in data, they are generally not as robust as the linear dimensionality reduction techniques and not able to handle certain types of nonlinear manifolds.

Kernel principal component analysis (Kernel PCA) [11, 13] is an extension of Principal Component Analysis (PCA) using techniques of kernel methods. Using a kernel, the originally linear operations of PCA are done in a reproducing kernel Hilbert space with a nonlinear mapping. Since kernel PCA is a kernel-based method, the mapping performed by Kernel PCA relies on the choice of the kernel function K, the polynomial kernel and the Gaussian kernel [12].

Diffusion Maps [17, 18] are based on defining a Markov random walk on the graph of the data. The technique defines a system of coordinates with an explicit metric that reflects the connectivity of a given data set and that is robust to noise. This diffusion metric is a transition probability of a Markov chain that evolves forward in time and is very robust to noise, unlike the geodesic or Euclidean distance.

Sammon's Mapping (SM) [19] provides a mapping from a high-dimensional vector space onto a 2-dimensional output space. The basic idea is to arrange all the data points on a 2 dimensional plane in such a way, that the distances between the data points in this output plane resemble the distances in vector space as defined by some metric as faithfully as possible. More formally, given a set of data points x_i in

$\mathcal{R}^n$ with d(x_i, x_j) being the distance between two data points according to some metric, we obtain a distance matrix D with elements d_{ij} in input space.

Locality Preserving Projections (LPP) in contrast to traditional linear techniques such as PCA and local nonlinear techniques for dimensionality reduction are capable of the successfully identifying the complex data manifolds such as the Swiss roll. This capability is due to the fact that cost functions are minimized by local nonlinear dimensionality reduction techniques, which aim at preserving local properties of the data manifold. However, in many learning settings, the use of a linear technique for dimensionality reduction is desired, e.g., when an accurate and fast out-of sample extension is necessary, when data has to be transformed back into its original space, or when one wants to visualize the transformation that is constructed by the dimensionality reduction technique. Linearity Preserving Projection (LPP) is a technique that aims at combining the benefits of linear techniques and local nonlinear techniques for dimensionality reduction by finding a linear mapping that minimizes the cost function of Laplacian Eigenmaps [20].

Neighborhood Preserving Embedding (NPE) similar to LPP, [21] minimizes the cost function of a local nonlinear technique for dimensionality reduction under the constraint that the mapping from the high dimensional to the low-dimensional data representation is linear. NPE is the linear approximation to LLE. NPE [22] defines a neighborhood graph on the dataset X, and subsequently computes the reconstruction weights $\boldsymbol{W_i}$ as in LLE.

Sochastic Neighbor Embedding (SNE) constitutes an unsupervised projection which follows a probability based approach. A Gaussian function is centered at every data point $\boldsymbol{x_i}$ which induces a conditional probability of a point $\boldsymbol{x_j}$ given $\boldsymbol{x_i}$.The goal of SNE is to find a low-dimensional data representation that minimizes the mismatch between $\boldsymbol{p_{j|i}}$ and $\boldsymbol{q_{j|i}}$. This is done by the minimization of the sum of the Kullback–Leibler divergences. The minimization of this objective function is difficult and may stick in local minima.

Symmetric Stochastic Neighbor Embedding(SSNE) [39] suppose the pair wise similarities of a set of D-dimensional data points X=$\{\boldsymbol{x_i}\}_{i=1}^{n}$ are encoded in a symmetric matrix P$\in \boldsymbol{R}_+{}^{n\times n}$, where $\boldsymbol{P_{ii}=0}$ and $\sum_{ij}\boldsymbol{P_{ij}}$=1, seeks d-dimensional (d$\ll \boldsymbol{D}$) representations of X, denoted by Y= =$\{\boldsymbol{y_i}\}_{i=1}^{n}$. Compared with an earlier method Stochastic Neighbor Embedding (SNE) [40], SSNE uses a symmetries cost function with simpler gradients most mapped points in the SSNE visualizations are often compressed near the center of the visualizing map without clear gaps that separate clusters of the data. Completely using the tangent space to represent the local geometry.

V. Performance And Quality Criteria

Before beginning the study of the reduction of dimensionality images (DR), it is necessary to define several normalized quality criteria for the reduction. These non supervised criteria will allow comparing these 13 methods of reduction in the analysis of image and in particular to measure the different types of degradations (loss of information, etc.) caused by the various methods DR. The interest of these criteria is to have evaluated the performances of reduction and the stability of these methods. An approach is then proposed to appreciate the appropriateness of these criteria, to applications of an artificial image of size 100x100x21.

A. Criteria derived from the norms $\boldsymbol{l_p}$

In this section, these criteria are directly derived from classical statistical measures. Every value individually is considered according to the spatial and dimensions spectral. The artificial image is represented as a three-dimensional matrix$\boldsymbol{I(x,y,\lambda)}$, with x is the position of the pixel in the line, y there is the number of the line and $\boldsymbol{\lambda}$ the spectral considered band. $\boldsymbol{n_x, n_y, n_\lambda}$ are respectively the number of pixels by line, the number of lines and the number of spectral bands. Note also equally $\sum_{x=1}^{n_x}\sum_{y=1}^{n_y}\sum_{\lambda=1}^{n_\lambda}\boldsymbol{I(x,y,\lambda)}$ by $\sum_{x,y,\lambda}\boldsymbol{I(x,y,\lambda)}$. The measures used are derived from the norms $\boldsymbol{l_p}$(I) =$\|\boldsymbol{I}\|_p$ = $(\sum_{x,y,\lambda}|\boldsymbol{I(x,y,\lambda)}|^p)^{1/p}$, the corresponding distance is defined by: $\boldsymbol{d_p(I,\bar{I}) = l_p(I-\bar{I})}$ =$(\sum_{x,y,\lambda}|\boldsymbol{I(x,y,\lambda)-\bar{I}(x,y,\lambda)}|^p)^{1/p}$.

1) Mean Absolute Error (MAE) or mean of $\boldsymbol{l_1}$: we mean here by the number of pixels to render the measurement independent of the size of the image.

$$\text{MAE }(\mathbf{I},\bar{\mathbf{I}}) = \frac{1}{n_x n_y n_\lambda}\sum_{x,y,\lambda}|\mathbf{I}(x,y,\lambda)-\bar{\mathbf{I}}(x,y,\lambda)|$$

2) Mean Square Error (MSE) or mean of $\boldsymbol{l_2^2}$: Here ,we also mean the number of pixels :

$$\text{MSE}(\mathbf{I},\bar{\mathbf{I}}) = \frac{1}{n_x n_y n_\lambda}\sum_{x,y,\lambda}\left(\mathbf{I}(x,y,\lambda)-\bar{\mathbf{I}}(x,y,\lambda)\right)^2$$

3) Maximum Absolute Distortion or mean of $\mathbf{l}_\infty$: (used by Motta [30])

$$\text{MAD} = \boldsymbol{l_\infty(I-\bar{I}) = max_{x,y,\lambda}\{|I(x,y,\lambda)-\bar{I}(x,y,\lambda)|\}}$$

4) First Spectral Derivative: The bandwidth, or wavelength range, of each band is a variable in a hyperspectral sensor [31]. This method explores the bandwidth variable as a function of added information. It is apparent that if two adjacent bands do not differ greatly then the underlying geo-spatial property can be characterized with only one band. The mathematical description is shown below, where I represent the hyperspectral value, x is a spatial location and λ is the central wavelength. Thus, if $\mathbf{D_1}$ is equal to zero then one of the bands is redundant. In general, the adjacent bands that differ significantly should be retained, while similar adjacent bands can be reduced.

$$\mathbf{D_1(\lambda_i)} = \sum_x\|\mathbf{I}(x,\lambda_i) - \mathbf{I}(x,\lambda_{i+1})\|$$

B. Criteria derived from the theory of information:

1) Entropy criterion:

The Shannon entropy [32] corresponds to the quantity of contained information or delivered by an information source $\boldsymbol{Q_i = -\log_2 p_i}$Where $\boldsymbol{p_i}$ is the probability of appearance of the level i. Since the probability $\boldsymbol{p_i}$ between 0 and 1, it follows that the quantity of information $\boldsymbol{Q_i}$ varies of 0 to the infinite . One will note that more the level is rare ($\boldsymbol{p_i}$ weak), more the quantity of information is big; however, A level brings few information. H= $\sum_{i=0}^{N-1}\boldsymbol{H_i}$ = $\sum_{i=0}^{N-1}\boldsymbol{P_i\,Q_i}$

2) Variance Criterion:

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{I_1 I_2} \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} (\mathbf{r}_{i_1 i_2} - \widehat{\boldsymbol{\mu}})^2$$

Avec $\widehat{\boldsymbol{\mu}} = \frac{1}{I_1 I_2} \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \mathbf{r}_{i_1 i_2}$

3) Similarity Criterion (SS):

Appears in [28,29] , it tries to measure the resemblance between two vectors, seen as vectors $\mathbf{n}_\lambda$ dimensional, v and $\mathbf{v}'$ defined by :

$$\mathrm{SS}\,(\boldsymbol{v}', \boldsymbol{v}) = \sqrt{\boldsymbol{RMSE}(\boldsymbol{v}, \boldsymbol{v}')^2 + (\mathbf{1} - \boldsymbol{corr}(\boldsymbol{v}, \boldsymbol{v}')^2)^2}$$

with $\mathrm{RMSE} = \sqrt{\frac{\sum_{\lambda=1}^{n_\lambda} (v(\lambda) - v'(\lambda))^2}{n_\lambda}}$ and

$$\mathrm{Corr}\,(\boldsymbol{v}, \boldsymbol{v}') = \frac{\frac{1}{n_\lambda - 1} \sum_{\lambda=1}^{n_\lambda} (v(\lambda) - \mu_v)(v'(\lambda) - \mu_{v'})}{\sigma_v \sigma_{v'}}$$

$\mathbf{v_k}$ **is different** $\Leftrightarrow \mathbf{SS}(\mathbf{v_k}, \mathbf{v_i}) > \mathbf{1}\ \forall \mathbf{i} = \mathbf{1}, \mathbf{2}, .., \mathbf{k} - \mathbf{1}$. In the non normalized case or of other term $\mathbf{v_k} \notin \mathrm{B}\ (\mathbf{v_i}, \mathbf{1})$ (B the closed ball of centers vi and radius equal to 1).

4) Structural Content Criterion (SC):

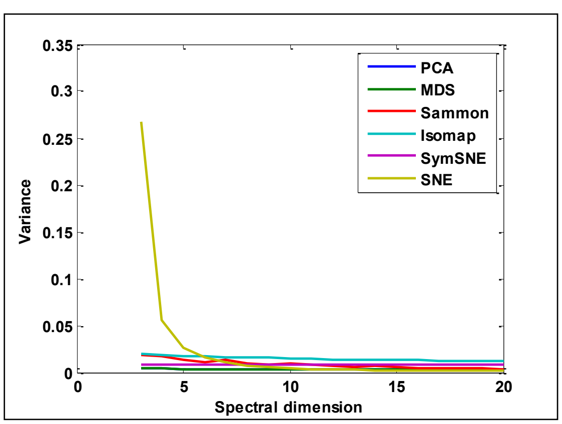
Structural Content (SC) is mentioned in [26, 27], it is defined as:

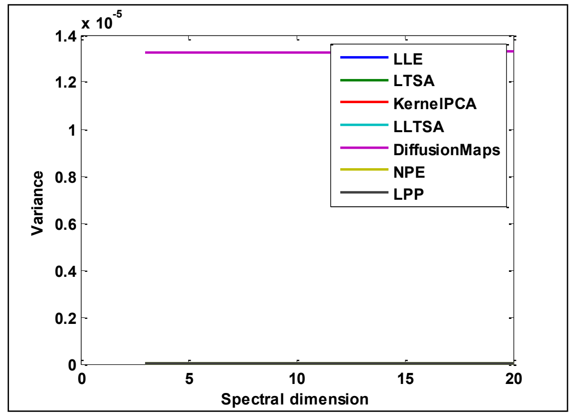
$$\mathrm{SC} = \frac{\sigma_{\tilde{I}}^2 + \mu_{\tilde{I}}^2}{\sigma_I^2 + \mu_I^2}$$

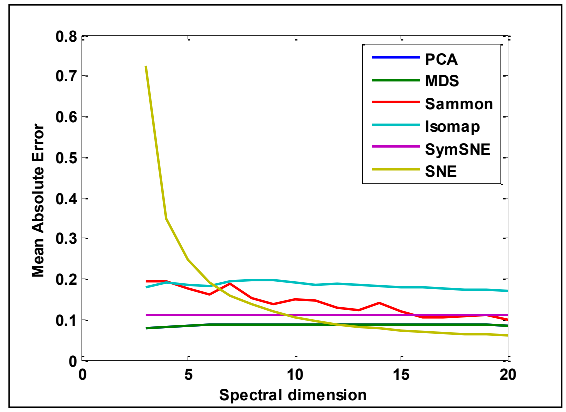
And for the images hyperspectral: $\mathrm{SC} = \frac{\sum_{x,y,\lambda} [\tilde{I}(x,y,\lambda)]^2}{\sum_{x,y,\lambda} [I(x,y,\lambda)]^2}$

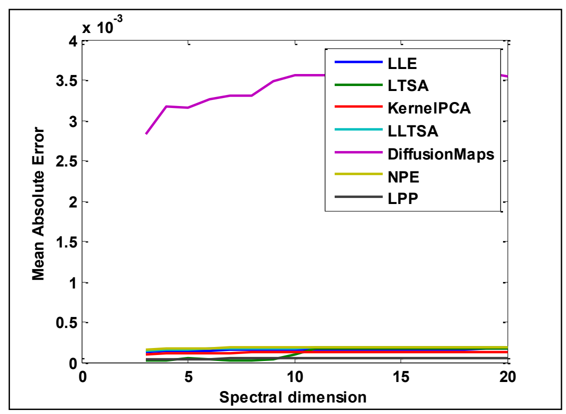
VI. Experiments

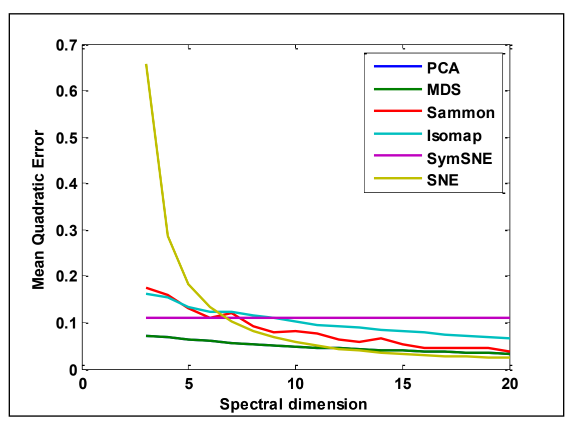
We can classify all the mathematical techniques of dimensions reduction in two categories. The curves are traced by nine quality criteria of intrinsic dimension equals to 20, for different levels. These results are compared to selected performances. Every curve is traced while representing the value of the criterion quality as the ordinate and in abscissa the value of the rate of reduction (spectral dimension). The minima and maxima are selected for every evolution (the selection is done in comparison with the results on an artificial image).

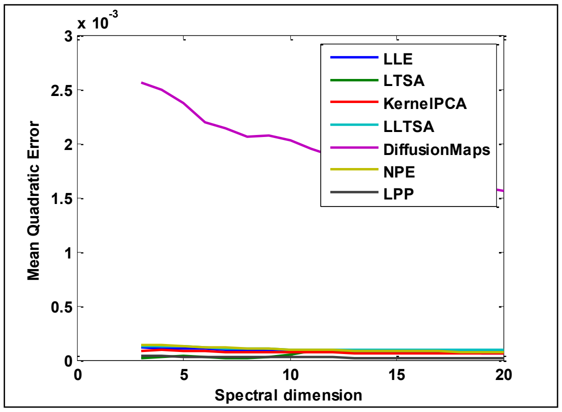


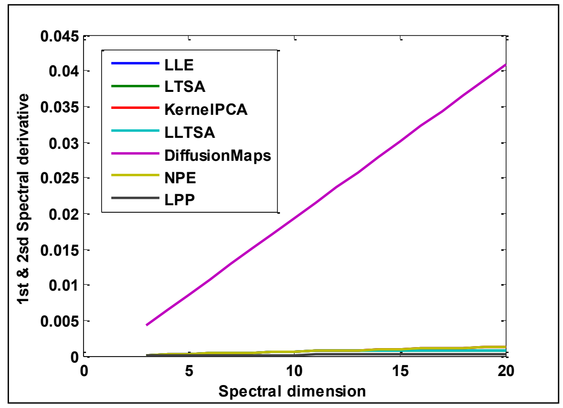


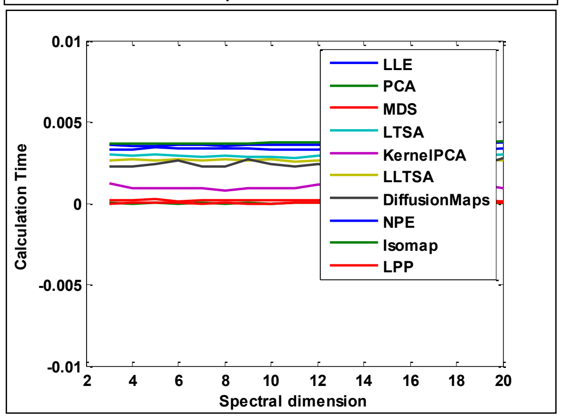


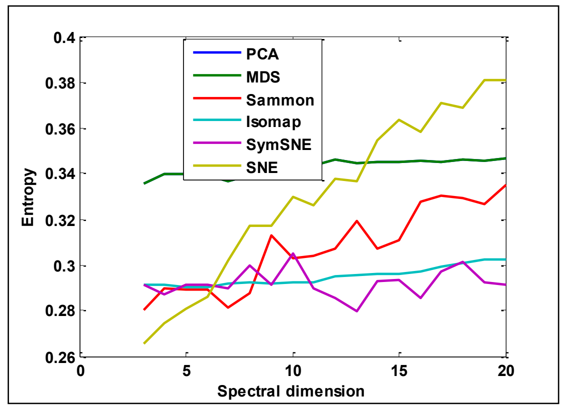


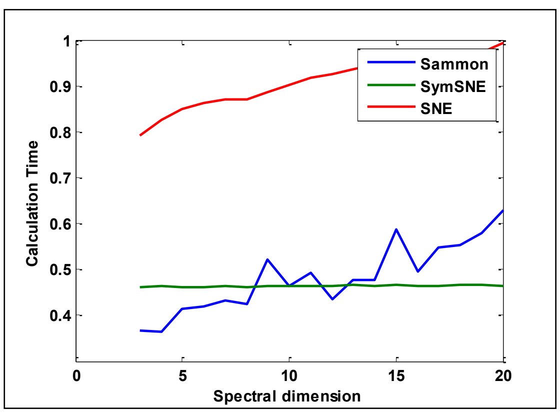


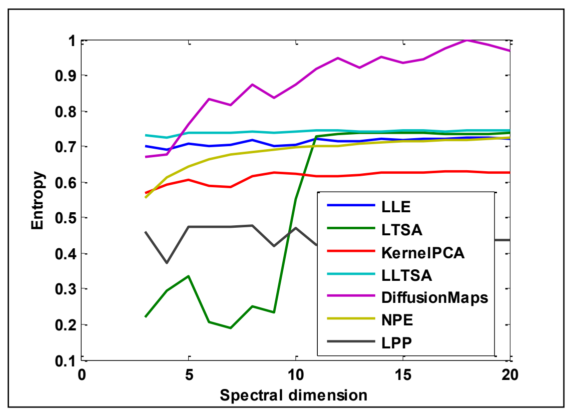

x 10^-3
Mean Quadratic Error
Spectral dimension
LLE
LTSA
KernelPCA
LLTSA
DiffusionMaps
NPE
LPP

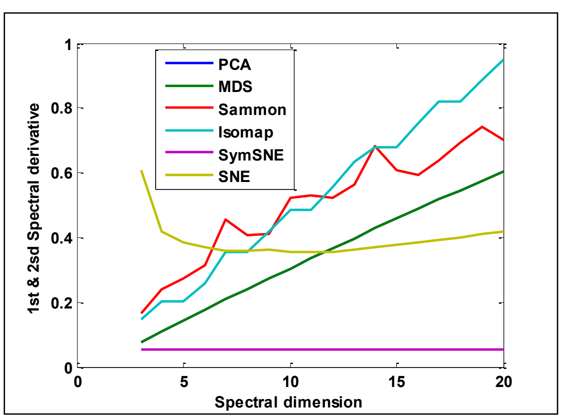

1st & 2sd Spectral derivative
Spectral dimension
LLE
LTSA
KernelPCA
LLTSA
DiffusionMaps
NPE
LPP

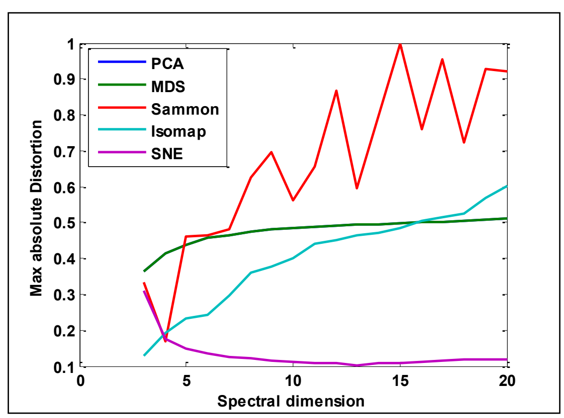

Calculation Time
Spectral dimension
LLE
PCA
MDS
LTSA
KernelPCA
LLTSA
DiffusionMaps
NPE
Isomap
LPP

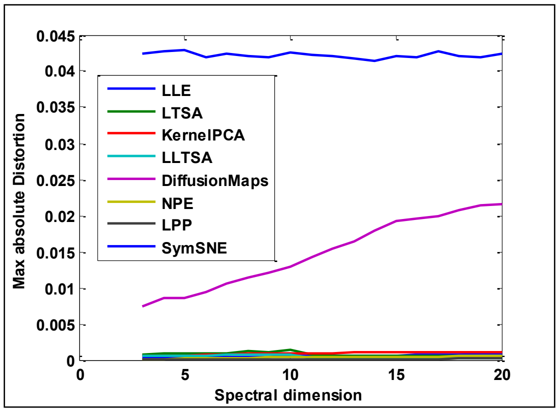

Entropy
Spectral dimension
PCA
MDS
Sammon
Isomap
SymSNE
SNE

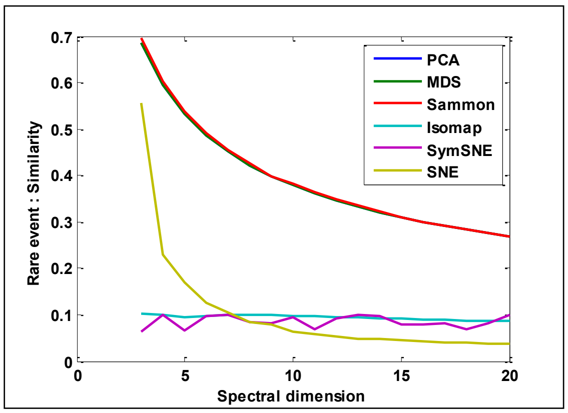

Calculation Time
Spectral dimension
Sammon
SymSNE
SNE

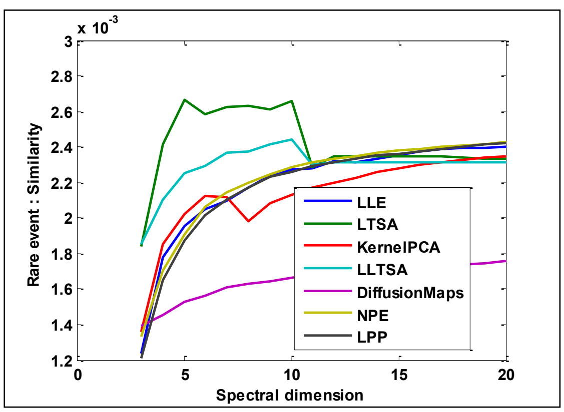

Entropy
Spectral dimension
LLE
LTSA
KernelPCA
LLTSA
DiffusionMaps
NPE
LPP

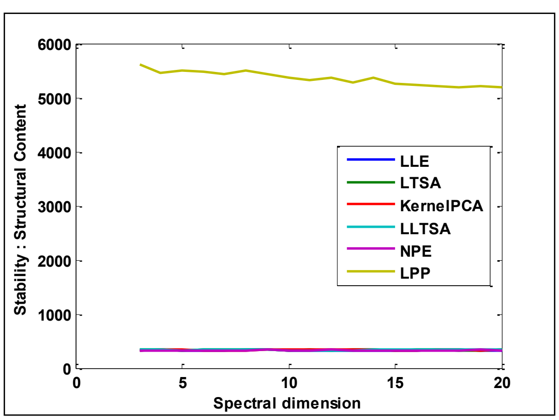

1st & 2sd Spectral derivative
Spectral dimension
PCA
MDS
Sammon
Isomap
SymSNE
SNE

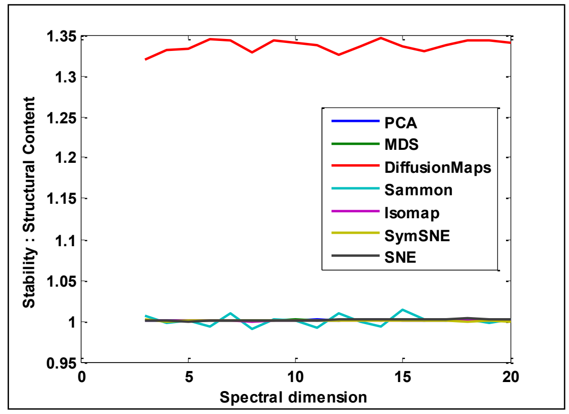

Max absolute Distortion
Spectral dimension
PCA
MDS
Sammon
Isomap
SNE











VII. Results And Discussion

A set of the curves obtained for all the criteria tested is presented above. In the following, we explain the interpretations that one can pull along with the conclusions obtained from simulation results.

A. Entropy :

These curves of experienced entropy on 13 methods of reduction show the following facts: at the time of the phenomenon of the reduction of dimension, the methods LLTSA, LTSA, KACP, LPP, LLE, NPE showed a similar behavior: the entropy is proportional to the reduction rate of the image. The maximum quantity of information kept in NPE is 0, 96 of the maximum value for a reduction rate varying in the interval [2, 20], and also reveals a good performance in comparison with other reduction techniques. The quantity of information kept in LLTSA, LTSA, LLE, LPP, and KACP is almost equivalent, but varying with the spectral dimension. We notice the small perturbations that complicate the choice of an optimal dimension reduction, which reveals the partial performance on the information quantities in comparison with NPE. For the PCA, MDS, DM, Sammon, Isomap, LPP, SymSNE, it is noticed that random degradations in the reduction, which is able to reflect instability of this criterion on the quantities of the data when it is quickly decreasing in the interval [2; 20]. In conclusion, this criterion presents its good illustrated performance on NPE and slightly less on LLTSA and LLE, despite the weak loss of information there is a strong fall in the rate of reduction equals to 2 in comparison with the other methods of reduction.

B. Variance :

The tests on the criterion of the variance show globally the proportional behavior for the 13 methods to the spectral dimension. We can also note a random degradation that seems to be a characteristic of certain methods, such as Sammon, regarding this reduction phenomenon, after preserving a quantity of data very close to SNE. For the PCA, Isomap, MDS, SymSNE, the curve of the tested variance preserves an average quantity of information varying in the interval [0.001; 0.015] almost constantly to the reduction rate. We can also note a quick raise reflecting the stability of this criterion regarding the quantities of data preserved by this set of methods. At last, the study on "Neighborhood Preserving Embedding", Diffusion Maps, LLTSA, LTSA, LLE, KACP and LPP shows bad performances in comparison with the other reduction methods explained above, despite the stability of the small values of variance within the reduction interval [2; 20]. In conclusion, this criterion shows better performance on the Isomap, followed by the method Sammon. The intrinsic dimension 10 seems to be an optimal choice.

C. Mean Absolute Error (MAE) and Mean Quadratic Error (MQE):

This criterion seems to react in a similar way to the criterion of the variance. We notice similar performances of these methods LTSA, LLTSA, NPE, kernel-ACP, LLE, LPP and DM, on one hand and to the six other methods SNE, Sym-SNE, PCA, MDS, Isomap and Sammon on the other hand. Thus, these two criteria (MAE and MQE) tested on Isomap reflect a good performance of the normalized values in comparison with the other methods of the same category. Equally, those two criteria are less effective on SymSNE, MDS, DM and SNE since they begin to lose the information at the time of a fall rather weak of maximum value average. Overall, the second category containing the six methods reflects the sensitiveness of these criteria on the quantities of information. On the other hand, the mean values, in the first category, are too weak within an interval of reduction rate of [2-20], mark the weakness of these seven methods quoted above to have preserved a quantity of favorable information. In conclusion, the criterion of mean absolute error and mean quadratic error show their effectiveness on Isomap and partially on the Sammon better than the other methods. The intrinsic dimension 10 seems to be an optimal choice.

D. Mean Absolute Distortion (MAD):

This criterion reflects the potential of a reduction method to preserve information , of which the distance is the most distant of the value of average information. Thus, Sammon presents a good performance on the MDS (similarly to the PCA) while preserving the maximum quantity of information in all the interval of reduction, despite complex perturbations that reveal a certain instability in this method for this criterion. Then the MAD curve associated to SNE and Isomap preserves a maximum information quantity of normalized value close to 0.2, which is slightly weaker in comparison with what precedes. The associated figures to the methods LLE, LTSA, KACP, DM, NPE, LPP, S-SNE, where some of them have few disturbances that can decrease the overall potential, reveals an unacceptable conservation of this criterion, but with the advantage of a small loss of this information if the reduction rate increases. In conclusion, despite the perturbations of the data that was submitted in the reduction phenomenon, this studied criterion demonstrates its partial performance on the Sammon, and a better performance on MDS while keeping a better maximum quantity than the other reduction methods.

E. First and Second Spectral Derivative:

This criterion can be well explained as a variation rate in the information stocked between two bands. An increase in this value can demonstrate the need to increase the number of bands in order not to lose all the information of a given event. Thus, for a size reduction of an artificial data tree, this criterion increases when the number of the bands is decreased. However, what we should seek for when comparing reduction methods is the maximum for this criterion. This means a maximum of information maximum per band when the number of bands is minimal. Taking this explanation for this test, Sammon will be classified as the most advantageous compared to methods like MDS, S-ENS, ENS, Isomap, PCA which give results not far from the first's. Finally, the methods with least important results are LLTSA, LTSA, NPE, LPP, LLE, KACP and DiffusionMaps.

F. Similarity (SS):

We exploited this similarity criterion (SS) to study the influence of rare events on the data reduction. We use the notion of difference to measure the similarity between two D-dimension vectors to show the effectiveness of this criterion after artificially introducing a new event $\boldsymbol{v_k}$ not similar the remainder of the data. We present in the Table1 below, the corresponding values of the quantity of rare information preserved by the various applications chosen by this similarity criterion. More this values is close to the unity, the more the method is preserving of these rare vectors in the reduction dimension. The results measure the relevance of the 13 methods and reveal the strong execution of techniques that do not use neighborhood graphs (PCA, MDS, Sammon), while preserving rare information at the time of the reduction phenomenon compared to the techniques based on neighborhood graphs (Isomap, LTSA, LLE, LLTSA, SNE and S-SNE). In addition, among the non linear reduction techniques, isomap (lightly less strong) shows its performance in comparison to other techniques of the same category (LLTSA, d-maps, LLE, LPP, S-SNE, SNE, LTSA, KACP, and NPE).

TABLE I. SENSIBILITY CRITERION FOR DIVERSE METHODS

		SS	SC
Linear	***PCA***	0.8	0.099
Nonlinear	***KACP***	1.9 x $\mathbf{10^{-2}}$	32.03
	MDS	0.8	1
	LTSA	2.4 x $\mathbf{10^{-2}}$	36.06
	LLE	1.4 x $\mathbf{10^{-2}}$	33.8
	LPP	1.25 x $\mathbf{10^{-2}}$	558.3
	SNE	0.2	1
	SYM-SNE	0.05	1.007
	DM	0.7 x $\mathbf{10^{-2}}$	1.37
	Sammon	0.8	1.001
	NPE	2.1 x $\mathbf{10^{-2}}$	24.69
	Isomap	0.1	1.001
	LLTSA	2.3 x $\mathbf{10^{-2}}$	36.06

G. Structural Content (SC):

This criterion can serve to value the stability of the reduction methods. The more this value is close to the unity, the more the method is stable. We digitally perturb the base image then we calculate for every spectral dimension the value of this criterion after reduction of the image. This study reveals that although certain non-linear techniques (MDS, SNE, Sammon, Isomap, DiffusionMaps, and SymSNE) and linear techniques (PCA) offer a rather increased stability, the remaining techniques (KPCA, NPE, LLE, LPP, LTSA, LLTSA) are considerably less advantageous at the reduction phenomenon of the data that underwent an initial perturbation.

On the other hand, the LPP seems to be least stable among all methods studied.

H. Time Criterion:

Another classical criterion consists to classify the dimension reduction methods according to the execution time. The nature of the processor influences certainly on this value, but one always can realize a comparative study of the normalized values for this criterion. The first group of the methods (PCA, MDS, LLE, LPP, LTSA, KPCA, DiffusionMaps, LLTSA, Isomap, and NPE) is particularly interesting and equivalent in calculation time . If we take the average of the time of calculation of these methods as reference value, then the methods SNE, Sammon and SymSNE respectively present 15, 30 and 100 times more to realize their reduction performances. These rather high values, can interpret themselves by the usage of the repeated techniques of optimization, by the latter group, to carry out the reduction. However, this inconvenience always can be admitted while observing their good presented performances in certain criteria.

VIII. CONCLUSION

The paper presents a review and comparative study of the techniques for dimensionality reduction. In following, the conclusion is presented followed by the perspectives, obtained from the made simulations.

TABLE II. SELECTION METHODS MORE PERTINENTS

Quality Criteria	Selection Methods
Entropy	NPE, LLTSA, LLE
Variance	Isomap, Sammon
MAE, MQE	Isomap, Sammon
MAD	Sammon, MDS
First and second Spectral Derivative	Sammon, MDS, S-SNE, Isomap, PCA
Time Calculation	PCA, MDS, LLE, LPP, LTSA, DM
SS	PCA, MDS, Sammon
SC	MDS, SNE, Sammon, Isomap, DM, S-SNE, PCA

A selection of the obtained methods is concluded in the Table2, presenting the performances of the application in comparison with the criterion quality observed. Of these obtained results, we can conclude that the non linear techniques for the dimensionality reduction are, despite their high variance, able to surpass the traditional PCA. The Table 2 reveals well the efficiency of Sammon and Isomap on the majority of the criteria through the simulations made during the reduction.

In the future, we plan to develop new techniques for nonlinear dimensionality reduction that does not count on the local property of the diverse data.

Besides, we plan a change towards the developement of techniques of reduction stable and preserving at best the rare events in the hyperspectral images on artificial data.

REFERENCES

[1] H. Hotelling. Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24:417–441, 1933.

[2] Pearson. On lines and planes of closest fit to systems of points in space. Philiosophical Magazine, 2:559–572, 1901.

[3] X. He, D. Cai, S. Yan, and H.-J. Zhang. Neighborhood preserving embedding. In Proceedings of the 10th IEEE International Conference on Computer Vision, pages 1208–1213, 2005.

[4] A. Saxena, A. Gupta, and A. Mukerjee. Non-linear dimensionality reduction by locally linear isomaps. Lecture Notes in Computer Science, 3316:1038–1043, 2004.

[5] X. He and P. Niyogi. Locality preserving projections. In Advances in Neural Information Processing Systems, volume 16, page 37, Cambridge, MA, USA, 2004. The MIT Press.

[6] N. Mekuz and J.K. Tsotsos. Parameterless Isomap with adaptive neighborhood selection. In Proceedings of the 28th DAGM Symposium, pages 364–373, Berlin, Germany, 2006. Springer.

[7] K.V. Mardia, J.T. Kent, and J.M. Bibby. Multivariate Analysis. Probability and Mathematical Statistics. Academic Press, 1995.

[8] T. Cox and M. Cox. Multidimensional scaling. Chapman & Hall, London, UK, 1994.

[9] S. Lafon and A.B. Lee. Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(9):1393–1403, 2006.

[10] Tenenbaum, J., de Silva, V., and Langford, J. (2000). A global geometric framework for nonlinear dimensionality reduction. Science, 290(5500):2319– 2323.

[11] Vapnik, V. N. (1995). The nature of statistical learning theory. New York: Springer Verlag.

[12] Williams, C. K. I. & Barber, D. (1998). Bayesian classification with processes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20:12, 1342–1351.

[13] Wahba, G. (1990). Spline models for observational data. Philadelphia, PA: Society for Industrial and Applied Mathematics. CBMS-NSF Regional Conference series in applied mathematics.

[14] Sch¨olkopf, B., Smola, A., &M¨uller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation, 10, 1299–1319.

[15] Z. Zhang and H. Zha. Principal manifolds and nonlinear dimensionality reduction via local tangent space alignment. SIAM Journal of Scientific Computing, 26(1):313–338, 2004.

[16] S.T. Roweis and L.K. Saul. Nonlinear dimensionality reduction by Locally Linear Embedding. Science, 290(5500):2323–2326, 2000.

[17] S. Lafon and A.B. Lee. Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(9):1393–1403, 2006.

[18] B. Nadler, S. Lafon, R.R. Coifman, and I.G. Kevrekidis. Diffusion maps, spectral clustering and the reaction coordinates of dynamical systems. Applied and Computational Harmonic Analysis: Special Issue on Diffusion Maps and Wavelets, 21:113–127, 2006.

[19] Sammon, J. W., A Non-Linear Mapping for Data Structure Analysis, IEEE Trans. on Computers, C-18(5), 1969.

[20] R. Huber, H. Ramoser, K. Mayer, H. Penz, and M. Rubik Classification of coins using an eigenspace approach. Pattern Recognition Letters, 26(1):61–75, 2005.

[21] University of Chicago, The University of Chicago, In Advances in Neural Information Processing Systems, Volume 16, page 37, Cambridge, MA, USA, 2004. The MIT Press.

[22] Xiaofei He; Deng Cai; Shuicheng Yan; Hong-Jiang Zhang,"Neighborhood preserving embedding", Tenth IEEE International Conference on Computer Vision, Volume 2, Issue , 17-21 Oct. 2005

[23] G. Hinton, S. Roweis, Stochastic neighbor embedding, in: Advances in Neural Information Processing Systems, vol. 15, MIT Press, Cambridge, MA, USA, 2003, pp. 833–840.

[24] C. Spearman. General intelligence objectively determined and measured. American Journal of Psychology, 15:206–221, 1904.

[25] K. Fukunaga. Introduction to Statistical Pattern Recognition. Academic Press Professional, Inc., San Diego, CA, USA, 1990.

[26] M. Eskicioglu et P. S. Fisher. “A survey of quality measures for gray scale image compression”. Dans 9th Computing in Aerospace Conference,93-4514. AIAA, oct. 1993.

[27] B. Lamiscarre et D. Léger. “Nouveaux critères pour l’évaluation globale de la qualitédes images ; applications aux images satellitaires restaurées ou comprimées”. Rap. tech., CERT/DERO, sept. 1995.

[28] S. Rupert, M. Sharp, J. Sweet et E. Cincotta. “Noise constrained hyperspectra data compression”. IEEE International Geoscience and Remote Sensing Symposium, IGARSS’01, vol. 1, p. 94–96. July 2001.

J. N. Sweet, M. Sharp et J. Granahan. “Hyperspectral analysis toolset”. Sensors, systems, and next-generation satellites IV, vol. 4169, p.396–407. SPIE, 2001.

[29] G. Motta, F. Rizzo et J. A. Storer. “Compression of hyperspectral imagery”.Dans Data Compression Conference, DCC, vol. 8, p. 333– 342.IEEE, mars 2003.

[30] Price, J. C., 1994. Band selection procedure for multispectral scanners. Applied Optics 33, 3281-3288.

[31] C. E. Shannon et W. Weaver. The Mathematical Theory of Communication. Illini Books, 1949.

[32] D.L. Donoho. High-dimensional data analysis: The curses and blessings of dimensionality. Lecture delivered at the "Mathematical Challenges of the 21st Century" conference of The American Math. Society, Los Angeles, August 6-11, 2000.

[33] W.S. Torgerson. Multidimensional scaling I: Theory and method. Psychometrika, 17:401–419,1952.

[34] C.J.C. Burges. Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers, chapter Geometric Methods for Feature Selection and Dimensional Reduction: A Guided Tour. Kluwer Academic Publishers, 2005.

[35] J.A. Lee and M. Verleysen. Nonlinear dimensionality reduction. Springer, New York, NY, USA,2007.

[36] J. Venna. Dimensionality reduction for visual exploration of similarity structures. PhD thesis,Helsinki University of Technology, 2007.

[37] Zhang, J. Yang, D. Zhao, X. Ge, Linear local tangent space alignment and application to face recognition,Neurocomputing 70 (2007) 1547–1553.

[38] Zhang, J. Yang, D. Zhao, X. Ge, Linear local tangent Space alignment and application to face recognition,Neurocomputing 70 (2007) 1547–1553.

[39] J. A. Cook, I. Sutskever, A. Mnih, and G. E. Hinton. Visualizing similarity data with a mixture of maps. In Proceedings of the 11th International Conference on Artificial Intelligence and Statistics, volume 2, pages 67–74, 2007.

[40] G. Hinton and S. Roweis. Stochastic neighbor embedding. Advances in Neural Information Processing Systems, 15:833–840, 2003.