Similarity of Dimensionality Reduction Methods Applied on Artificial Hyperspectral Images

J. Khoder[1], R. Younes[2], and F. Ben Ouezdou[2]
[1]Azm Center for research in biotechnology, Lebanese University, Tripoli, Lebanon
[2]LISV Laboratory, University of Versailles Saint-Quentin en-Yvelines, Paris, France

Abstract - *Dimensionality reduction is a big challenge in many areas. In this research we address the problem of high-dimensional hyperspectral images in which we are aiming to preserve its information quality. This paper introduces a study similarity of the non parametric and unsupervised methods of projection and of bands selection used in dimensionality reduction of different noise levels determined with different numbers of data points. The quality criteria based on the norm and correlation is employed obtaining a good preservation of these hyperspectral image's vectors in the reduced dimensions. The added value of these criteria can be illustrated in the evaluation of the reduction's performance, when considering the similarity of two categories of bands selection methods and projection methods. A classification of the methods of projection and bands selection is ranked of three categories (good, average and bad) presenting the performances of the application in comparison with the similarity criterion observed..*

Keywords: Dimensionality reduction; manifold learning; similarity spectral criteria; hyperspectral data.

1 Introduction

Hyperspectral imaging has become an active research topic in recent years due to its wide-spread applications in areas such as resource management, agriculture, mineral exploration, and environmental monitoring. With the number of channels in the hundreds instead of in the tens, hyperspectral imagery possesses much richer spectral information than multispectral imagery [1]. However, identifying the material reflecting specific spectral signature remains a challenge for realizing the full potential of hyperspectral technology. It is clear that more effective data processing techniques are needed to deal with hyperspectral cubes. Because it is necessary to have a minimum ratio of training pixels to the number of spectral bands [2], dimension reduction has become a significant part of hyperspectral image interpretation. Dimension reduction is the transformation that brings data from a high order dimension to a low one, thus conquering the curse of dimensionality [3]. Similar to a lossy compression method [4], dimension reduction reduces the size of the data, but unlike compression, dimension reduction is application-driven. Mathematically, given n points $x_1,\ldots,x_n$ in a high dimensional subspace of R^D the goal of dimensionality reduction is to find a mapping: F: $R^D \longrightarrow R^d$, y_i= F $(x_i), i = 1, \ldots, n$. Where $y_i \in R^d$, i=1… n and d is the dimensionality of the embedding space. 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. Ideally, the reduced representation should have a dimensionality that corresponds to the intrinsic dimensionality of the data. Which is the minimum number of parameters needed to take into account for the observed properties of the data [5]? As a result, dimensionality reduction facilitates, among others methods, classification, visualization, and compression of high-dimensional data. Several local approaches of dimension reduction methods were used to address this problem. Chang and al. have proposed a robust modification of Locally Linear Embedding (LLE), Robust LLE [6]. They provided an efficient algorithm to detect and remove the large noises, namely, the outliers. However, RLLE would also fail when the data have some small noises. Pan and Ge have generated a multiple weights LLE, NLLE [7]. This method uses the (k–d) linear independent combination weights to represent the local structure. Chen et al. have also proposed an effective preprocessing procedure for current manifold learning algorithms [8]. They analyzed the input data statistically and then detected the noises. Ridder et al. have also solved the robustness problem of LLE [9] by introducing a weighted reformulation in the embedding step. Hou and Zhang have been developed a large number of local approaches, stemming from statistics or geometry [10]. In practice, these local approaches are often in lack of robustness, since in contrast to maximum variance unfolding (MVU), which explicitly unfolds the manifold; they merely characterize local geometry structure. Moreover, the eigenproblems encountered are hard to solve. These methods try to tackle this problem through a unified framework that explicitly unfolds the manifold and reformulate local approaches as semi-definite programs instead of the above-mentioned eigenproblems. Three well known local approaches (LLE-LE-LTSA) are interpreted and improved within this framework. These methods proposed several experiments on both synthetic and real datasets. These results have shown that the dimensionality reduction techniques SLLE-SLTSA-SLE also have some troubles are not stable, so sensitive in the presence of noises levels and more stable the parameter k. Tsai studied current linear and nonlinear dimensionality reduction techniques in the context of data visualization [11]. Experiments were conducted on varying the neighborhood, density and needed noise levels of data taken into account. He used the manifold metric based on the correlation coefficient that computes the pair wise geodesic distance vector between the original manifold and the lower-dimensional embedding results used. The calculation of the metric is similar to the correlation used by Geng et al [12], but the pair wise geodesic distance vector is calculated for the original data instead of the Euclidean distance vector. A previous study carried out by Tsai and Chan [13] showed that this metric was more suitable for representing the visualization results if the original data lies on a manifold.

A new approach we used to measure the similarity of these linear and nonlinear techniques to reduce the size of a disturbance in the data set is based on the noise variance at different scales. Similarity criteria, as defined in section 3 are used. These criteria take into account the intrinsic structure of the original hyperspectral and disturbed images. This comparative study focuses on the influence of noise variance in the data set with respect to the spectral dimension.

The outline of the remainder of this paper is as follows. Section 2 describes the methodology of the approaches used. We give a formal definition of linear techniques for dimensionality reduction and subdivide the dimensionality reduction techniques into three linear and fifteen non-linear techniques. Moreover, in Section 3, we define the quality criteria and present the results of the experiments in Section 4. Section 5, discusses the influence of the rate disturbance in the phenomenon of reduction allowing us to conclude about the main contributions of this paper.

2 Methodology approach

Regards to methodology when considering the stability of two categories of selection bands BandClust and projection methods (linear and non linear). There are mainly two kinds of dimensionality reduction approaches [14]: linear and non linear. Linear methods may mainly include Principal Component Analysis (PCA) [15], LLTSA and non linear techniques: Kernel PCA [16], Isomap [6], MDS [17], LTSA [18], Diffusion maps [19], Sammon mapping [20], SymSNE [21], SNE[23], LLE [22], LPP [5], Neighborhood Preserving embedding (NPE) [24], HLLE [25], Laplacian Eigenmaps [26], Landmark Isomap [32], Kernel LDA [31]. An unsupervised approach to band reduction in hyperspectral remote sensing imagery [27]. The band selection involves selecting a minimal subset of M bands $S=(S_1,S_2,\ldots,S_M)$ from the original set $F = (F_1,F_2,\ldots,F_N)$, where ($M<<N$) and ($S\subseteq F$). Hyperspectral imaging offers high richness of information which is often necessary to achieve good classification performance at the pixel level. Hyperspectral images generally show a high amount of correlation between adjacent spectral bands. Therefore, removing this redundancy would reduce the amount of data that are relevant to further classification and interpretation stages. This selects an appropriate subset of image bands to ful fill the same applications the full image can, to some extent The BandClust method using mutual information between two bands as selection criterion provides a rather increased stability is presented in [28].

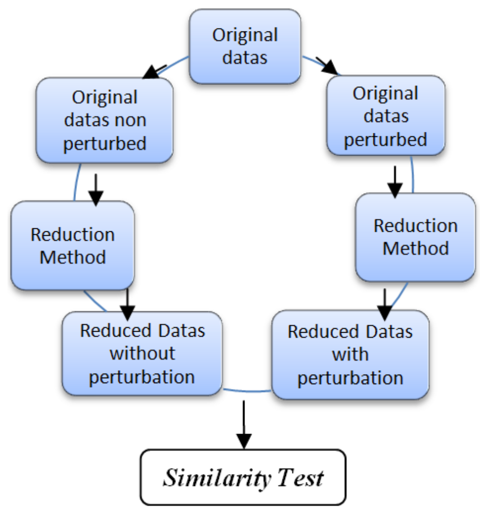


Fig. 1. Proposed Method for Similarity Study applied on Hyperspectral Images.

Although, our comparative review study for stability includes the most important nonlinear techniques for dimensionality reduction, our aim is to identify an demonstrate an efficient algorithm to detect and remove the large noises. The approaches of this framework that we propose deal with reduction for high-dimensional noisy signals. The problem of nonlinear dimensionality reduction can be defined as follows. Assume we have dataset represented in $n \times D$ matrix X consisting of n data vectors x_i for $i = \{1, 2... n\}$ with dimensionality; assume further that this dataset has intrinsic dimensionality d where $d < D$. In practice, the signal-subspace perturbation ξ_i from observed vectors has to satisfy the following general model: $\xi_i = x_i + z_i$, $i=1\ldots n$. Where $x_i\in R^D$ is the observed random vector $z_i \in R^D$ is the data-acquisition or/and model noise; finally, the stability test will be used on a set data and data perturbed. The approached is illustrated in the **figure 1**.

3 Performances And quality criteria

Before beginning the study of the Dimensionality Reduction of images (DR), it is necessary to define several normalized quality criteria derived from classical statistical measures for the reduction. These unsupervised stability criteria will allow comparing and evaluated the performances of reduction and the stability of these 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. An approach is then proposed to appreciate the appropriateness of these criteria, to applications of a Hyperspectral images. Every value individually is considered according to the spatial and dimensions spectral. The artificial image is represented as a three-dimensional matrix I(x,y,λ), with x is the position of the pixel in the line, y there is the number of the line and λ the spectral considered band. n_x ,n_y, n_λ are respectively the number of pixels by line, the number of lines and the number of spectral bands. Note also equally $\sum_{x=1}^{n_x}\sum_{y=1}^{n_y}\sum_{\lambda=1}^{n_\lambda} I(x,y,\lambda)$ by $\sum_{x,y,\lambda} I(x,y,\lambda)$.

3.1 Similarity Criteria (SS)

The similarity criteria was appeared in [29] and tries to measure the resemblance between two vectors, seen as vectors L dimensional, v and L defined by :

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

With RMSE = $\sqrt{\frac{\sum_{\lambda=1}^{n_\lambda}(v(\lambda)-v'(\lambda))^2}{n_\lambda}}$,

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

3.2 Fidelity (F)

This criterion was developed by Eskhicioglu [29].We define fidelity by a ratio of the spectral density (DSP) of the error image and reference. U and V two sets of values (initially pixels of two images) of which we wants to evaluate the difference. Defines by:

F (U, V) = $1-\frac{L_2^2(U-V)}{\sigma_u^2+\mu_u^2}$ (2)

Either for hyperspectral images, spatial fidelity define by :

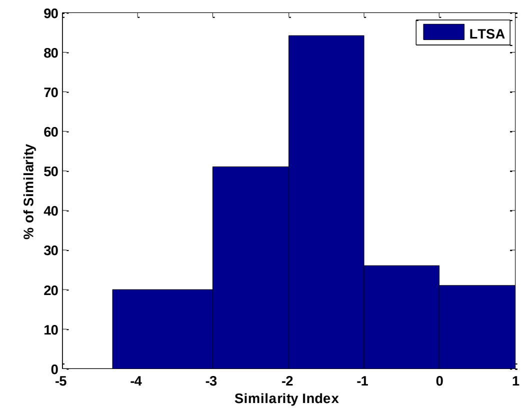
F $(I,\tilde{I})$= $1-\frac{\sum_{x,y,}[I(x,y,\lambda)-\tilde{I}(x,y,\lambda)]^2}{\sum_{x,y,}[I(x,y,\lambda)]^2}$ (3)

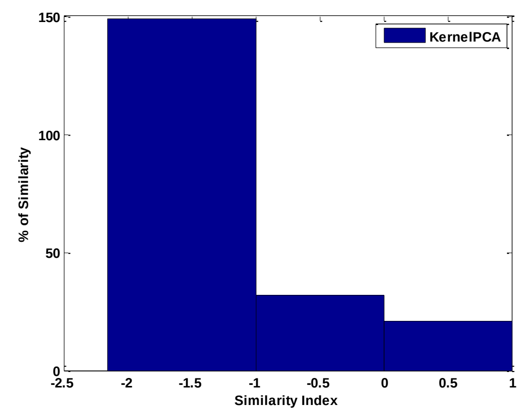
The fidelity is equal to one, when the output image is equal to the input image. Due to the quadratic action, the small errors will be minimized.

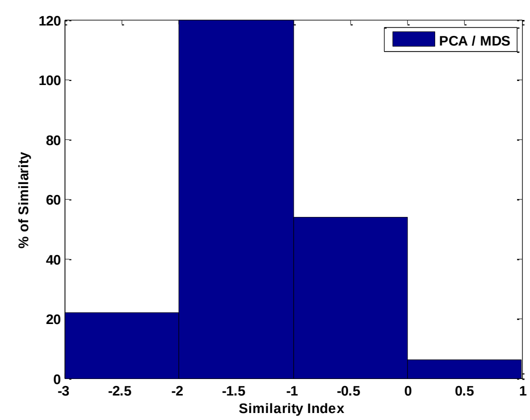
4 Experiments and results

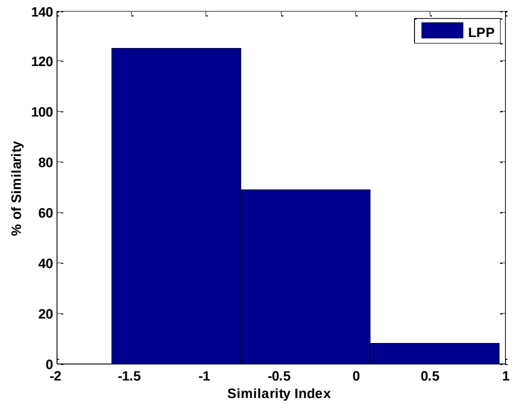
A selection of the curves obtained for the tested criteria of the Similarity Criteria and the fidelity are presented above. In the following, we explain the influence rate of disturbance of a pixel introduced to the data observation in the reduction phenomenon and the results obtained from simulations.

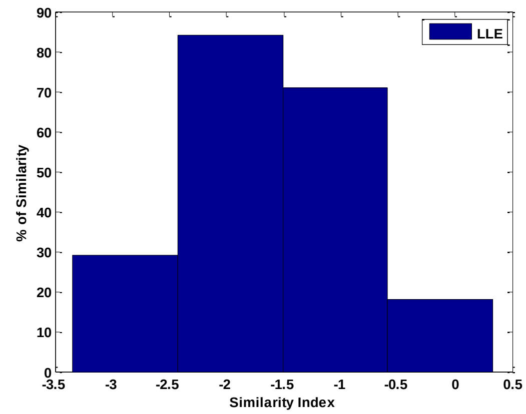
Fig.2. Similarity percentage obtained for the different reduction methods

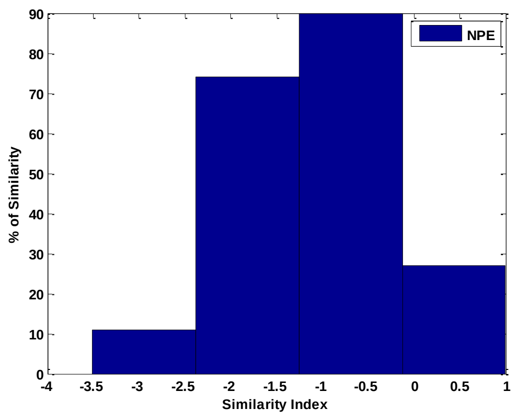


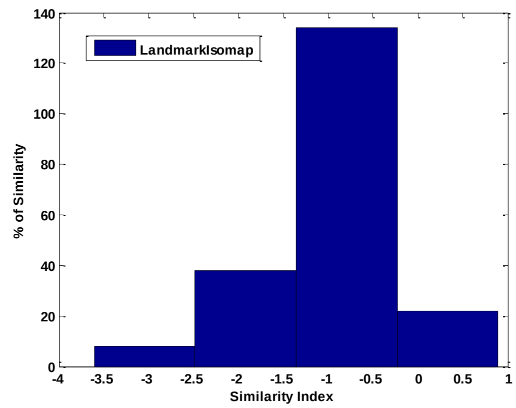


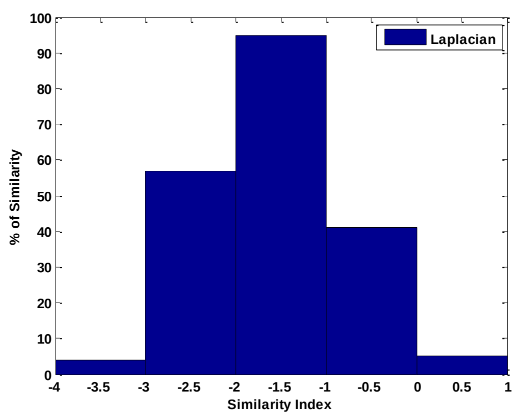


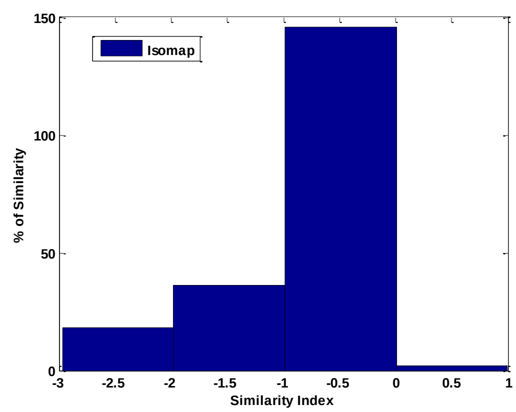

LPP
% of Similarity
Similarity Index

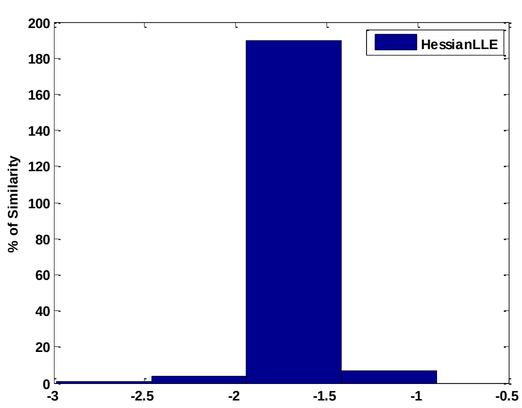

LLE
% of Similarity
Similarity Index

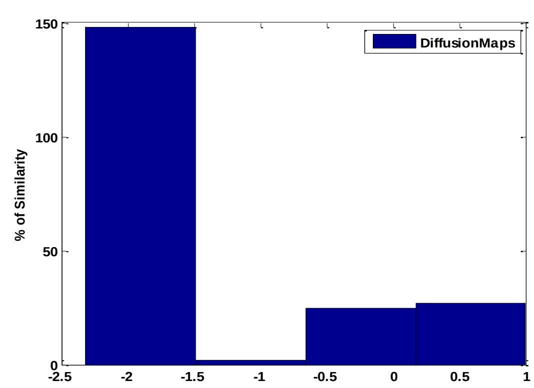

NPE
% of Similarity
Similarity Index

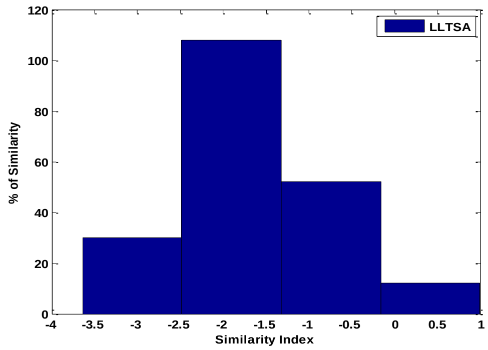

LandmarkIsomap
% of Similarity
Similarity Index

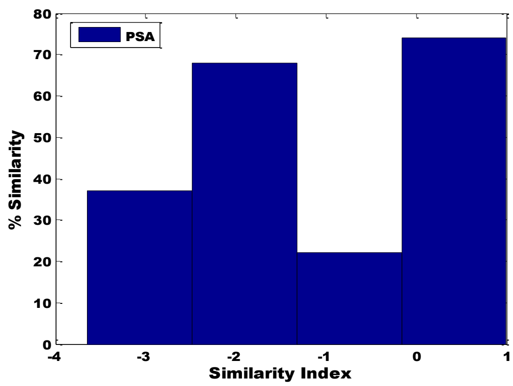

Laplacian
% of Similarity
Similarity Index

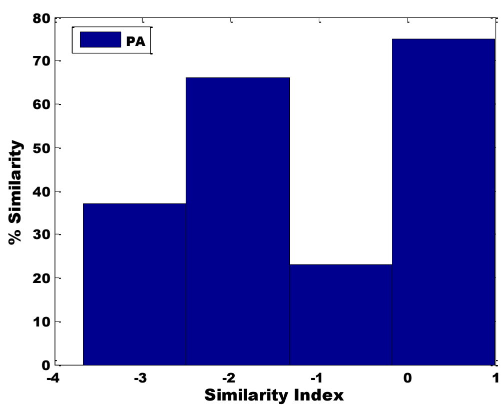

Isomap
% of Similarity
Similarity Index

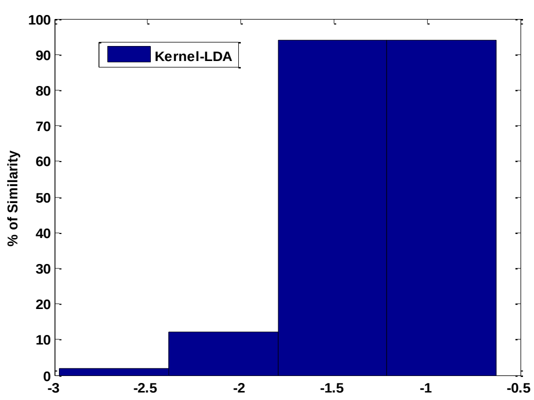

HessianLLE
% of Similarity
Similarity Index


DiffusionMaps
% of Similarity
Similarity Index









4.1 Projection methods Structural Similarity Criteria (SS)

Similarity Criteria numerically values the stability of the methods, the more this value is close to the unit, and the more the method is stable. We disturb numerically the original image, and then we calculate the value of this criterion proportionally to the noise degradation on a scale [0%; 10%]. These NCC curves tested on 15 reduction methods show the following facts: We notice many similar performances for the methods: PCA, MDS, NPE LTSA, Laplacian, KernelPCA, LPP and Sammon. Also, we notice many similar performances for the other categories of nonlinear techniques like: Diffusion Maps, LLTSA, LLE, and KernelLDA. The similarity criterion tested on NPE, MDS and PCA reflects a good performance compared to other methods in the same category. In reality, the SS value that varies around the unit, offers a well increased stability under a noise scale [0%, 10%] followed by a small deformation of the original image. The quantity of information stored in diffusion Maps, LLE, DM and LLTSA is almost equivalent. But not constant according to degradation of noise, bound to a geometrical deformation of increasing SS value all around the unit proportionally to the variation of noise. This is illustrated by the deformation of information bound by a strong fall of degradations compared to the other methods of reduction on a scale of noise [2%; 10%], and SS value rather near to the unit. In conclusion, the SS criterion shows a better partial performance on the PCA, MDS followed by NPE method. The variation rate of noise equal to 40 seems to be a best alternative.

4.2 Band selection methods: *Structural Similarity Criteria (SS)*

This category of methods seems giving different advantages compared to the first category of linear or non-linear projection. Indeed, the two indexes evolving near the unit which allows concluding a stable behavior for the three compared methods. The BandClust method using mutual information between two bands as selection criterion provides a rather increased similarity, reflects a good performance compared to other methods bands selection in the same category PSA and PA. On the other hand, the influence of this noise parameter on the bands selection methods performance probably imposes the use of the natural images of large dimension. Note that a noise percentage of 20% was not sufficient to disturb the performances of these methods.

In the **Table 1**, a classification of the methods of projection and of bands selection is ranked of three categories (good, average and bad) presenting the performances of the application in comparison with the similarity criterion observed.

Classification derived by reduction methods		
Good	Average	Bad
PCA – MDS – NPE – LTSA - KernelPCA - LPP - Laplacian-Sammon - BandClust	Isomap – PSA - PA Landmark isomap – LLE - Kernel LDA	HessianLLE

Table 1. The various reduction methods used, in order to categorize, with similarity criteria, on artificial data with different noise variances.

5 Conclusion

The paper presents a review and comparative study of techniques for dimensionality reduction**.** In this comparative study on the similarity of non-parametric algorithms, unsupervised by dimension reduction of large images (by projection and by selection of spectral bands) and taking into account the similarity evolution criteria presented in section 3. It is revealed that the techniques unsupervised by projection are either limited by their linear character (ACP, MDS), or difficult to use because of their algorithm complexity when working on high dimensional data. Moreover, the majority of them are sensitive to the different undergoing variations as the noise degradation and the information loss. This is the reason making these techniques do not fully meet our two main points of interest: stability and preservation of the rare event. Despite the PCA-MDS algorithms followed by NPE that are quite sensitive to noise degradation in comparison with other reduction techniques of projection, to preserve the geometrical structure local / global of reduced data**.** It seems that one should focus on selection techniques by projection, which at the moment can stay in sensitive observations space with some little variation. The selection of bands in unsupervised learning is not explored sufficiently and the existing techniques strongly rely on similarity measures between attributes (spectral bands) or on variance measures. We can conclude from this study on the robustness of the selection method BandClust and MDS projection method, which found very encouraging results on the similarity of artificial data compared to the influence of noise during the reduction .Several extension of this work can be considered such as the development of new hybrid techniques by band selection with projection methods for dimensionality reduction, which does not rely on the local properties of data.

6 References

[1] H. Grahn and P. Geladi, Eds, Techniques and Applications of Hyperspectral Image Analysis. Chichester, U.K.: Wiley, 2007.

[2] J.A. Richards. Remote Sensing Digital Image Analysis: An Introduction. 2nd ed. New York: Springer-Verlag, 1993.

[3] D.W. Scott, The curse of dimensionality and dimension reduction, in Multivariate Density Estimation: Theory, Practice and Visualization. New York: Wiley, 1992, ch. 7, p. 195–217.

[4] Gao, Qinfeng Shi, Tibério S. Caetano “Dimensionality reduction via compressive sensing”. Pattern Recognition Letters, Volume 33, Issue9, 1 July 2012, 1163-1170.

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

[6] Andrés Álvarez-Meza, Juliana Valencia-Aguirre, GenaroDaza-Santa coloma,Germán Castellanos Domínguez. "Global and local choice of the number of nearest neighbors in locally linear embedding".Pattern Recognition Letters, Volume 32, Issue 16, 1 December 2011, Pages 2171-2177.

[7] Yaozhang Pan, Shuzhi Sam Ge Abdullah Al Mamun. "Weighted locally linear embedding for dimension reduction". Pattern Recognition, Volume 42, Issue 5, May 2009, Pages 798-811.

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

[9] B. Scholkopf, A. Smola and K.R. Muller. Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation, 10, 1299–1319. (1998).

[10] Y. Wu. Jiao, C. Hou and C. Zhang. Stable local dimensionality reduction approaches. Pattern Recognition, vol. 42, Issue 9, p. 2054-2066. 2006.

[11] F.S. Tsai. Comparative Study of Dimensionality Reduction Techniques for Data Visualization. Journal of Artificial Intelligence, vol. 3, Issue 3, p. 119-134. 2010.

[12] X. Geng, X., D.C. Zhan and Z.H. Zhou. Supervised nonlinear dimensionality reduction for visualization and classification. IEEE Trans. Syst. Man Cybernetics Part B, 35: 1098-1107. 2005.

[13] F.S. Tsai and K.L. Chan. A manifold visualization metric for dimensionality reduction. Nanyang Technological University Technical Report. 2009.

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

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

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

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

[18] 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.
[19] B. Nadler, S. Lafon, R.R. Coifman and I.G. Kevrekidis. Diffusion maps, spectral clustering and the reaction coordinate of dynamical systems. Applied and Computational Harmonic Analysis: Special Issue on Diffusion Maps and Wavelets, 21:113–127, 2006.
[20] S. De Backer, A.Naud, P.Scheunders. "Non-linear dimensionality reduction techniques for unsupervised feature extraction". Original Research Article Pattern Recognition Letters, Volume 19, Issue8, June 1998, Pages711-720.
[21] J.A. Lee and M. Verleysen. Nonlinear dimensionality reduction. Springer, New York, NY, USA, 2007.
[22] M. Belkin, P. Niyogi. Towards a Theoretical Foundation for Laplacian based Manifold Methods. Proceedings of the 18th Annual ConferencLearning Theory (COLT), Springer, New York, 2005, pp 486-500.
[23] J.A. Cook, I. Sutskever, A. Mnih, and G.E. Hinton. Visualizing similarity data with a mixture of maps. Proceedings of the 11th International Conference on Artificial Intelligence and Statistics, volume 2, pages 67–74, 2007.
[24] He Xiaofei, Deng Cai, Shuicheng Yan, Hong-Jiang Zhang. Neighborhood preserving embedding. Tenth IEEE International Conference on Computer Vision, Volume 2, Issue , 17-21 Oct. 2005
[25] D.L. Donoho and C. Grimes. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. PNAS, 100: 5591-5596. 2003.
[26] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput., 15: 1373-1396. 2003.
[27] L.J.P. vanderMaaten. An introduction to dimensionality reduction using Matlab. Technical Report 07-06, MICC-IKAT, Maastricht University, Maastricht, The Netherlands, 2007.
[28] S.R. Baofeng Guo Gunn, R.I. Damper, J.D.B. Nelson. Band Selection for Hyperspectral Image Classification Using Mutual Information. Geoscience and Remote controle sensing Letters, IEEE, vol. 3, Issue, 4, pp522-526, 2006.
[29] M. Eskicioglu et P.S. Fisher. A survey of quality measures for gray scale image compression. 9th Computing in Aerospace Conference, 93-4514. AIAA, oct. 1993.