

# Extensions and Analysis of Local Non-Linear Techniques

Rashmi Gupta

Ambedkar Institute of Advanced Communication Technologies and Research affiliated to Guru Gobind Singh Indraprastha University, Delhi, India

Rajiv Kapoor

Delhi Technological University, Bawana Road, (Formerly Delhi College of Engineering), Delhi, India

## ABSTRACT

The techniques Conformal Eigenmap and Neighborhood Preserving Embedding (NPE) have been proposed as extensions of local non-linear techniques. Many of the commonly used non-linear dimensionality reduction, such as Local Linear Embedding (LLE) and Laplacian eigenmap are not explicitly designed to preserve local features such as distances or angles. In first proposed Conformal Eigenmap technique, a low dimensional embedding is constructed that maximally preserves angles between nearby data points. The embedding is derived from the bottom eigenvectors of LLE by solving an additional problem in Semidefinite Programming (SDP). In second proposed method, NPE 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. The idea is to modify the LLE by introducing a linear transform matrix. The effectiveness of the proposed methods is demonstrated on synthetic datasets. Experimental results on several data sets demonstrate the merits of proposed techniques.

## Keywords

Dimension reduction; Manifold Learning; Conformal Eigenmap; Neighborhood Preserving Embedding; Local Linear Embedding; Laplacian Eigenmap

## 1. INTRODUCTION

Dimensionality reduction is important in many domains, since it mitigates the curse of dimensionality and other undesired properties of high-dimensional spaces [1-2]. 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), Linear Discriminant Analysis (LDA) and Multidimensional scaling (MDS). These techniques generate faithful low dimensional representations when the high dimensional input patterns are mainly confined to a low dimensional subspace. PCA, the most frequently used dimension reduction method seeks a projection that best represent the data in a least-squares sense [3]. MDS finds an embedding that preserves the interpoint distances, equivalent to PCA when those distances are Euclidean [4]. LDA, a supervised learning algorithm selects a transform matrix in such a way that the ratio of the between-class scatter and the within-class scatter is maximized [3]. If the input patterns are distributed more or less throughout this subspace, the Eigen value spectra from these methods also reveal the data set's intrinsic dimensionality. A more interesting case arises,

however, when the input patterns lie on or near a low dimensional sub manifold of the input space. In this case, the structure of the data set may be highly non-linear, and linear methods are bound to fail.

In the last decade, a large number of non-linear techniques for dimensionality reduction have been proposed [5-10]. Recently, several manifold-embedding-based nonlinear approaches were developed such as locally linear embedding (LLE) [11], isometric feature mapping (Isomap) [12] and Laplacian Eigenmap (LEM) [13]. They all utilized local neighborhood relation to learn the global structure of nonlinear manifolds. But they have quite different motivations and objective functions. In contrast to the traditional linear techniques, the non-linear techniques have the ability to deal with complex non-linear data. On the other hand, such approaches also has several limitations: (i) the solutions do not yield an estimate of the underlying manifold's dimensionality; (ii) the geometric properties preserved by these embedding are difficult to characterize; (iii) the resulting embeddings sometimes exhibit an unpredictable dependence on data sampling rates and boundary conditions. Moreover, the original LLE, Isomap and Laplacian Eigenmap cannot deal with the out-of-sample problem [14] directly. Out-of-sample problem states that only the low dimensional embedding map of training samples can be computed but the samples out of the training set (i.e. testing samples) cannot be calculated at all.

An extension of Isomap was proposed to learn conformal transformations [15]; like Isomap, however, it relies on the estimation of geodesic distances, which can lead to spurious results when the underlying manifold is not isomorphic to a convex region of Euclidean space [16]. Hessian LLE is a variant of LLE that learns isometrics, or distance-preserving embeddings, with theoretical guarantees of asymptotic convergence [17].

In this paper, an extended analysis has been provided to remedy the key deficiencies of LLE and Laplacian Eigenmap. It is shown how to construct a more robust, angle-preserving embedding from the spectral decompositions of these algorithms as well as linear approximations to local non-linear techniques.

The rest of this paper is organized as follows: Section 2 described the proposed methods, Conformal Eigenmap and NPE. Experimental results are shown in Section 3. Finally, conclusions are drawn in Section 4.

## 2. EXTENSIONS OF LOCAL NON-LINEAR TECHNIQUES

The capability of local non-linear techniques to successfully identify complex data manifolds has led to the proposal of its several extensions. The original LLE, Isomap and Laplacian Eigenmap cannot deal with out of sample problem and cannot preserve local feature such as angle. To overcome the limitations of existing methods, extensions of local non-linear techniques have been discussed in this section. The difference between these methods lies in their different motivations and objective functions.

### 2.1 Conformal Eigenmap

A conformal mapping is a transformation that preserves the angles between neighboring datapoints when reducing the dimensionality of the data [18]. Conformal Eigenmap are based on the observation that local non-linear techniques for dimensionality reduction do not employ information on the geometry of the data manifold that is contained in discarded eigenvectors that correspond to relatively small Eigen values. Conformal Eigenmap initially perform LLE (or alternatively, another local nonlinear technique for dimensionality reduction) to reduce the high-dimensional data  $D$  to a dataset of dimensionality  $m$ . Conformal Eigenmap use the resulting intermediate solution in order to construct a  $d$ -dimensional embedding (where  $d < m < D$ ) that is maximally angle-preserving.

A conformal map is a low-dimensional embedding where the angles formed by three neighboring points in the original high dimensional dataset are equal to the angles between those same three points in the embedding. Consider the point  $x_i$  and its neighbors  $x_j$  and  $x_k$  in  $d$ -dimensional space. Also, consider  $z_i$ ,  $z_k$  and  $z_j$  to be the images of those points in the final embedding. If the transformation were a conformal map then the triangle formed by the  $x$  points would have to be similar to that formed by the  $z$  points. In the triangle formed by the  $x$  points the expression  $|x_j - x_k|$  represents the length of one side of the triangle while the expression  $|z_j - z_k|$  represents the corresponding side in the embedding. Since the triangles are similar there must exist  $\sqrt{s_i}$  such that:

$$\sqrt{s_i} = \frac{|x_j - x_k|}{|z_j - z_k|} = \frac{|x_i - x_k|}{|z_i - z_k|} = \frac{|x_i - x_j|}{|z_i - z_j|} \quad (1)$$

$$s_i = \frac{|x_j - x_k|^2}{|z_j - z_k|^2} = \frac{|x_i - x_k|^2}{|z_i - z_k|^2} = \frac{|x_i - x_j|^2}{|z_i - z_j|^2} \quad (2)$$

It is usually not possible to find a perfect embedding where all of the triangles are exactly similar to each other. Therefore, the goal is to find a set of  $z$  coordinates such that the triangles are as similar as possible. This leads to the following minimization:

$$\min_{z, s_i} \sum_{j, k} \left( |z_j - z_k|^2 - s_i |x_j - x_k|^2 \right)^2 \quad (3)$$

Where the  $x_i$  represent the initial points and the  $z_i$  denote the points in the final embedding. Let  $y_i$  represent the points in the embedding produced by LLE (or LEM). The  $y_i$  points represent an intermediate step in the algorithm and so go part

of the way to solving for  $z_i$ . Once LLE has produced, goal of the algorithm becomes a search for a transformation matrix  $L$  such that  $z = Ly_i$  where the  $z_i$  value satisfy the minimization in Eq. (4)

$$\min_{L, s_i} \sum_{j, k} \left( |Ly_i - Ly_k|^2 - s_i |x_j - x_k|^2 \right)^2 \quad (4)$$

This should be done for all points  $x_i$  and with the condition that the points  $x_j$  and  $x_k$  are the neighbors of  $x_i$ .

$$\min_{L, s_i} \sum_{i, j, k} \eta_{ij} \eta_{ik} \left( |Ly_i - Ly_k|^2 - s_i |x_j - x_k|^2 \right)^2 \quad (5)$$

Where  $\eta$  is an indicator variable and  $\eta_{ij} = 1$  only if  $x_j$  is neighbor of  $x_i$  otherwise  $\eta_{ij} = 0$ . The value for  $s_i$  can be calculated via least squares and the initial minimization (5) can be rewritten as:

Minimize  $t$

$$\begin{aligned} \text{s.t. } & P \succeq 0, \\ & \text{trace}(P) = 1, \\ & \begin{pmatrix} 1 & R \text{Vec}(P) \\ R \text{Vec}(P)^T & t \end{pmatrix} \succeq 0, \end{aligned} \quad (6)$$

Where  $P = L^T L$ ,  $t$  is an unknown scalar,  $I$  and  $R$  are  $m^2 \times m^2$  matrices:  $I$  denote the identity matrix, while  $R$  depends on  $\{x_i, y_i\}_{i=1}^n$ , but is independent of optimization variables  $P$  and  $t$ . The condition  $\text{trace}(P) = 1$  is added to avoid the trivial solution where  $P=0$ . The optimization is an instance of SDP over elements of unknown matrix  $P$  [19]. After solving the SDP, the matrix can be decomposed back into  $L^T L$  and the final embedding can be found by  $z = Ly_i$  for all  $i$ . Conformal Eigenmap introduced the interesting idea of using spectral methods like LLE and LEM to find the low dimensional manifold and further modifying the output to produce a conformal map. The effectiveness of conformal map is limited by the computational complexity of SDP solvers.

Another extension of local non-linear dimension reduction technique is discussed in the section that follows.

### 2.2 Neighborhood Preserving Embedding

Neighborhood Preserving Embedding (NPE) is the linear approximation to local non-linear technique [20]. In contrast to traditional linear techniques such as PCA, local non-linear techniques for dimensionality reduction are capable of successful identification of complex data manifolds such as Swiss roll. This capability is due to the cost functions that are minimized by local non-linear 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 was constructed by the dimensionality reduction technique. NPE is a technique that aims at combining the benefits of linear techniques and local non-

linear techniques for dimensionality reduction by finding a linear mapping that minimizes the cost function of LLE [11]. NPE minimizes the cost function of a local non-linear technique for dimensionality reduction under the constraint that the mapping from the high-dimensional to the low-dimensional data representation is linear.

Similar to LLE, NPE starts with the construction of a nearest neighbor graph in which every datapoint is connected to its nearest neighbors. The weights of the edges in the graph are computed and subsequently solves the generalized eigen problem.

Given a set of points  $X = \{X_1, X_2, \dots, X_N\}$ , in  $R^D$ , NPE attempts to seek an optimal transformation matrix  $P$  to map high-dimensional data  $X$  onto a low-dimensional data  $Y = \{Y_1, Y_2, \dots, Y_N\}$ , in  $R^d$  ( $d \ll D$ ) in which the local neighborhood structure of  $X$  can be preserved, namely,

$$Y = P^T X .$$

NPE algorithm can be stated in three steps.

Step 1: Compute the neighbors of each data point  $\mathbf{X}_i$ . There are two ways to compute neighbors:

- K nearest neighbors (KNN): Put a directed edge from node  $i$  to  $j$  if  $\mathbf{X}_j$  is among the K nearest neighbors of  $\mathbf{X}_i$ .
- $\varepsilon$  neighborhood: Put an edge between nodes  $i$  and  $j$  if  $\|\mathbf{x}_i - \mathbf{x}_j\| \leq \varepsilon$

The graph constructed by the first method is a directed graph, while the one constructed by the second method is an undirected graph. In many real world applications, it is difficult to choose a good  $\varepsilon$ . In this work, the KNN method is adopted to construct the adjacency graph. When computational complexity is a major concern, one may switch to  $\varepsilon$  neighborhood.

Step 2: Compute the weights  $W_{ij}$  that best reconstruct each data point from its neighbors. In this step, the weights on the edges are computed. Let  $W$  denote the weight matrix with  $W_{ij}$  having the weight of the edge from node  $i$  to node  $j$ , and 0 if there is no such edge. The weights on the edges can be computed by minimizing the following objective function,

$$\min \sum_i \left\| \mathbf{X}_i - \sum_j W_{ij} \mathbf{X}_j \right\|^2 \quad (7)$$

Where  $\sum_j W_{ij} = 1, j = 1, 2, \dots, m$ , is enforced to make sure

that the reconstruction weight matrix  $W$  is invariant to translation of the data point.

Step 3: Computing the projections  $Y_i$  of each data points  $X_i$  in space  $\mathbf{R}^D$  can be reconstructed by  $W$ , then the corresponding point by  $Y_i$  in space  $\mathbf{R}^d$  can be reconstructed by  $W$  also. Therefore, the mapping transformation matrix  $P$  can be obtained by solving the following minimization problem:

$$P_{opt} = \arg \min_P \left[ \sum_i \left\| \mathbf{Y}_i - \sum_{j=1}^k W_{ij} \mathbf{Y}_j \right\|^2 \right] \quad (8)$$

$$\arg \min_A \left( P^T X M X^T P \right)$$

$$s.t. P^T X M X^T P = 1, M = (I - W)^T (I - W)$$

Where  $I$  represent the  $n \times n$  identity matrix. Using Lagrange multiplier, the minimization problem of equation (8) becomes a generalized eigenvalue problem:

$$X M X^T P = \lambda X X^T P \quad (9)$$

Then, the column vectors of the transformation matrix  $P$  are given by the bottom  $d$  eigenvectors of (9), which are ordered according to their eigenvalues  $\lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{d-1}$ . Therefore, and the projection can be obtained through the formula  $Y = P^T X$ .

### 3. RESULTS

In this section, some results of proposed algorithms are presented for a number of synthetic datasets. The datasets are specifically selected to investigate how the dimensionality reduction techniques deal with data that lies on a low-dimensional manifold.

The synthetic datasets on which the algorithms are implemented are the swissroll dataset, the helix dataset and twinpeaks dataset. Figure 1 shows plots of the three artificial datasets. All artificial datasets consist of 5,000 samples. The experiments are run for parameter  $k$  (nearest neighbors of data point) ranges from 5 to 15.

Figures 2-4 show the results of Conformal Map and NPE dimension reduction techniques on Swiss roll, helix and twinpeaks dataset respectively. From the results, it is clear that the angle preserving embedding more faithfully preserves the shape of the underlying manifold's boundary. The maximally angle preserving embedding exploits structure in the few bottom eigenvectors of LLE, not just in the bottom two eigenvectors. The semidefinite programming (SDP) in Eq. (6) mixes all of the bottom eigenvectors from LLE or Laplacian Eigenmap to obtain the maximally angle-preserving embedding.

Conformal transformations cast a new light on older algorithms, such as LLE and Laplacian Eigenmap. Viewing these bottom eigenvectors as a partial basis for functions on the data set, it is shown how to compute a maximally angle-preserving embedding by solving an additional problem in SDP. At little extra computational cost, Conformal Eigenmap significantly extends the utility of LLE and Laplacian Eigenmap, yielding more faithful embeddings as well as a global estimate of the data's intrinsic dimensionality.

The proposed NPE is able to search a direction projected onto which neighborhood relations are preserved along the curve of the manifold. From Fig. 2(c), Fig. 3(c) and Fig.4(c), it is

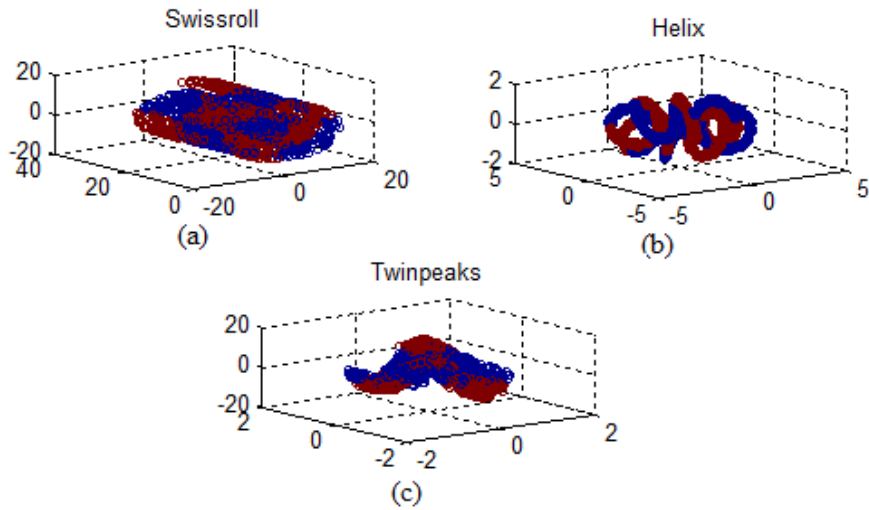


Fig 1: Four artificial datasets (a) Swissroll (b) Helix (c) Twinpeaks

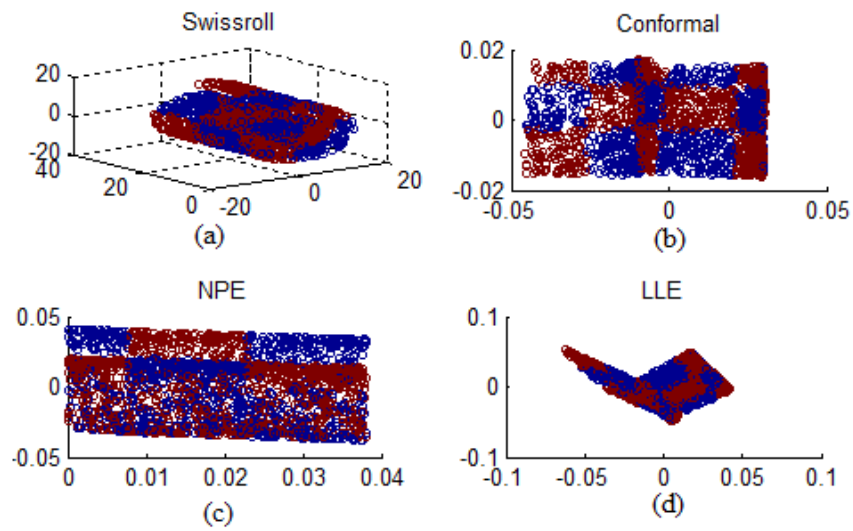


Fig 2: Results of dimensionality reduction on Swiss roll dataset (a) Swiss roll (b) Conformal (c) NPE (d) LLE

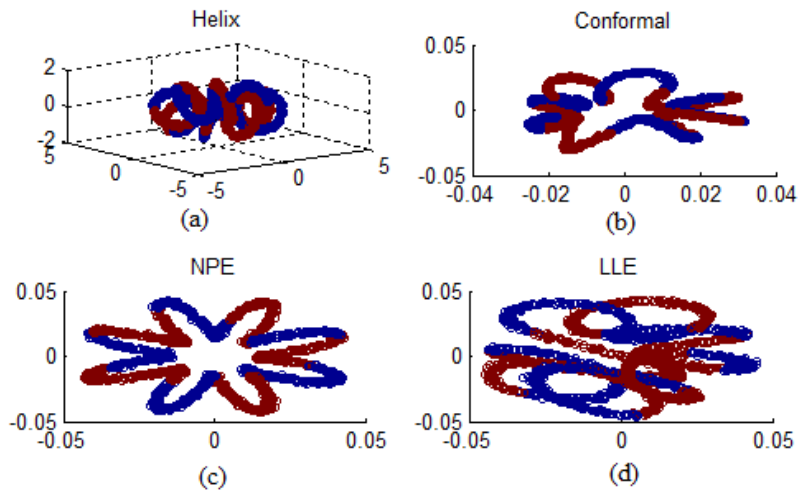


Fig. 3: Results of dimensionality reduction on Helix dataset (a) Helix (b) Conformal (c) NPE (d) LLE

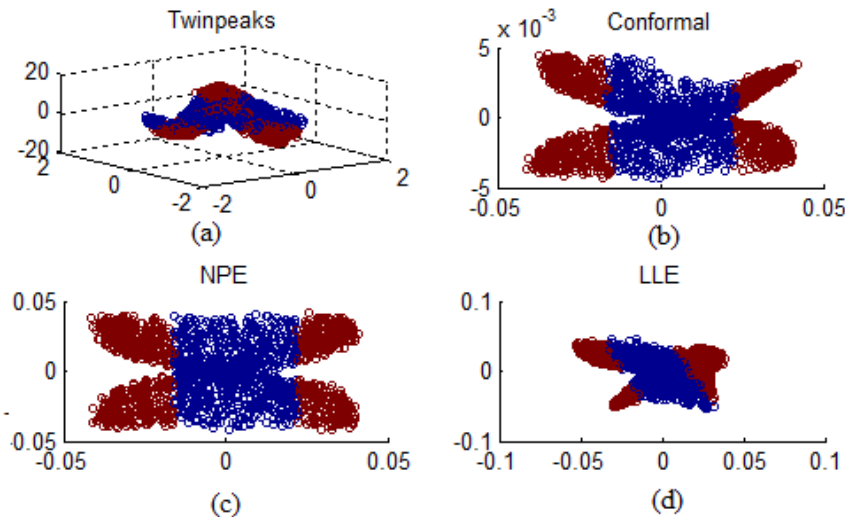


Fig. 4: Results of dimensionality reduction on Twinpeaks dataset (a) Twinpeaks (b) Conformal (c) NPE (d) LLE

	LLE	Conformal	NPE
Speed	Fast	Slow	Very Fast
Handle Curvature	Maybe	Yes	No
Handles noise	No	Yes	No
Preserves angles	No	Yes	No

observed that NPE cannot always unfold the manifold as LLE can. Furthermore, many neighbors are collapsed into a single point in the low dimensional space. The reason is that NPE is a linear transform instead of non-linear one like LLE. Nevertheless, the NPE has favorable properties against other linear transform methods such as PCA. The techniques are compared based on various parameters such as speed, noise, non-convexity, curvature, non-uniform sampling. First, it is observed for manifold geometry swissroll LLE is pretty slow, LLE and Laplacian can't handle this data. Twin Peaks: fold up the corners of a plane. LLE will have trouble because it introduces curvature to plane. LLE distort the mapping the most. Add noise to the Helix sampling. LLE cannot recover the circle. When the sampling rate is changed along the torus, Laplacian starts to mess up and Hessian is completely thrown off. Hessian LLE code crashed frequently on this example.

#### 4. CONCLUSIONS

Conformal Map and NPE build on LLE and Laplacian Eigenmap, thus inheriting their strengths as well as their weaknesses. Results show that the Conformal Map algorithm has the potential to preserve the angles from their high dimensional data. NPE, due to linear in nature, might not outperform as good as non-linear LLE, Isomap and Laplacian Eigenmap. However, it is a novel and useful linear dimension reduction method. However the effectiveness of conformal mapping is limited by the computational complexity of SDP solver. In future work, the algorithm can be developed that does not require SDP solution so that it can be applied to wide variety of datasets.

#### 5. REFERENCES

- [1] Jain, A.K., Duin, R.P.W., and Mao, J. 2000. Statistical Pattern Recognition: A Review, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 1, 4-37.
- [2] Jimenez, L.O. and Landgrebe, D.A. 1997. Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data, IEEE Transactions on Systems, Man and Cybernetics, vol. 28, no.1, 39-54.
- [3] Belhumeur, P.N., Hefanha, J.P., and Kriegman, D.J. 1997. Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE. Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, 711-720.
- [4] Torgerson, W.S. 1952. Multidimensional scaling I: Theory and method, Psychometrika, vol. 17, 401-419.
- [5] Burges, C.J.C. 2005. 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.
- [6] Saul, L.K., Weinberger, K.Q., Ham, J.H., F. Sha, and Lee, D.D 2006. Spectral methods for dimensionality reduction, In Semisupervised Learning, Cambridge, MA: The MIT Press.
- [7] Maaten, L.J.P., Postma, E.O., and Herik, H.J. 2008. Dimensionality reduction: A comparative review.
- [8] Scholkopf, B., Smola, A., and Muller, K.R., 1998. Nonlinear Component Analysis as a Kernel Eigenvalue Problem" Neural Computation, vol. 10, no. 5, 1299-1319.
- [9] Baudat, G. and Anouar, F. 2000. Generalized discriminant analysis using a kernel approach, Neural Computation, vol. 12, 2385-2404.
- [10] He, X. and Niyogi, P. 2003. Locality Preserving Projections, Advances in Neural Information Processing Systems 16, Vancouver, British Columbia, Canada.

- [11] Roweis S.T. and Saul, L.K. 2000. Nonlinear dimensionality reduction by Locally Linear Embedding, *Science*, vol. 290, no. 5500, 2323–2326.
- [12] Tenenbaum, J.B., Silva, V. de and J.C. Langford, 2000. A global geometric framework for nonlinear dimensionality reduction, *Science*, vol. 290, no. 5500, 2319–2323.
- [13] Belkin, M. and Niyogi, P. 2000. Laplacian Eigenmaps and spectral techniques for embedding and clustering, In *Advances in Neural Information Processing Systems*, vol. 14, 585–591, Cambridge, MA: The MIT Press.
- [14] Bengio, Y., Paiement, J., Vincent, P., Dellalau, O., Roux, N.L, Quimet, M. 2003. Out-of sample Extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering, *Neural Information Processing Systems*.
- [15] Silva, V. de and Tenenbaum, J. B. 2003. Global versus local methods in nonlinear dimensionality reduction,” *Advances in Neural Information Processing Systems 15*, 721–728, Cambridge, MA: MIT Press.
- [16] Donoho, D. L. and Grimes, C. E. 2002. When does Isomap recover the natural parameterization of families of articulated images? (Technical Report 2002-27). Department of Statistics: Stanford University.
- [17] Donoho, D. L. and Grimes, C. E. 2003. Hessian eigenmaps: locally linear embedding techniques for high dimensional data, *Proceedings of the National Academy of Arts and Sciences*, 100, 5591–5596.
- [18] Sha, F. and Saul, L.K. 2005. Analysis and Extensions of Spectral Methods for Nonlinear Dimensionality Reduction, *Proceedings of twenty second International Conference on Machine Learning*, Germany.
- [19] Vandenberghe, L. and Boyd, S. 1996. Semidefinite programming. *SIAM Review*, vol. 38, no. 1, 49–95.
- [20] .He, X., Cai, D., Yan, S., and Zhang, H. 2005. J Neighborhood preserving embedding, *Proceedings of Tenth International Conference on Computer Vision*, Piscataway: IEEE Press, 1208-1213.