An Analytical Mapping for LLE and Its Application in Multi-Pose Face Synthesis

Jun Wang, Zhang Changshui, Kou Zhongbao
State Key Lab of Intelligent Technology and Systems
Dept. of Automation, Tsinghua University
Beijing, 100084, China
wangjun00@mails.tsinghua.edu.cn
zcs@mail.tsinghua.edu.cn
kzb98@mails.tsinghua.edu.cn

Abstract

Locally Linear Embedding (LLE) is a nonlinear dimensionality reduction method proposed recently. It can reveal the intrinsic manifold of data, which can't be provided by classical linear dimensionality reduction methods. The application of LLE, however, is limited because of its lack of a mapping between the observation and the low-dimensional output. In this paper, we propose a method to establish an analytical mapping for LLE and validate its efficiency with the application in multi-pose face synthesis. Furthermore, through learning the similarity for the same kind of pose change mode of different persons, we generalize our method to small set cases with methods of statistical learning theory. The experiments of multi-pose face synthesis on small sets prove that our idea and method are correct.

1 Introduction

Dimensionality reduction is an important and necessary preprocessing of multidimensional data, such as face images. There are four purposes to reduce dimensionality of observation data: (1) to compress the data to reduce storage requirements; (2) to eliminate noise; (3) to extract features from data for recognition; (4) to project data to a lower-dimensional space, especially a visualized space, so as to discern the distribution of data. For face images, classical dimensionality reduction methods include Eigenface [1] , Independent Component Analysis (ICA) [2, 3], Linear Discriminate Analysis [4], and Local Feature Analysis (LFA) [5, 6], etc. The linear methods have their limitations. Firstly, they cannot reveal the intrinsic distribution of a given data set. Secondly, if there are changes in pose, facial expression and illumination, the projections may not be appropriate and the corresponding reconstruction error may be much higher.

Compared with linear dimensionality reduction methods, Nonlinear Dimensionality Reduction (NDR) techniques yield better results. Kernel-based methods are important nonlinear techniques. Among these, Kernel Principal Component Analysis [7], Kernel Independent Component Analysis [8] and Kernel Discriminate Analysis [9] have been investigated and proven effective.

BMVC 2003 doi:10.5244/C.17.1

LLE (Locally Linear Embedding) and ISOMAP are two other recently developed nonlinear dimensionality reduction techniques. LLE maps the observation data to a single global coordinate system of lower dimensionality preserving the neighbor relationships [10]. ISOMAP computes pair-wise distances in the geodesic space of the manifold, and then performs classical Multidimensional Scaling (MDS) to map data points from their high-dimensional input space to low-dimensional coordinates of a nonlinear manifold [11]. Both LLE and ISOMAP can reveal the underlying structure of data.

Although LLE and ISOMAP are effective dimensionality reduction methods, there are open problems. These two methods both lack a process for mapping between observations and embedded space. Yet mapping is very necessary for dimensionality reduction in many real-world problems, such as data compression. The first problem to solve is how to establish the mapping. Furthermore, in order to reveal the underlying manifold of data, the two methods must be applied on large sets. The intuitive description of revealed manifold is indistinct and inconvincible in small set cases. In many examples of machine learning, however, data is limited so that we must face small-set problems. The second problem is the issue of how to learn the analytical description of manifold and the mapping of a given high-dimensional small set.

In this study, we make use of multi-pose face images to research the two problems above based on LLE. First, we propose an analytical method based on nonlinear regression to establish the process for mapping. Second, we develop our method to adapt to small sets by a priori learning and methods from statistical learning theory. The method is based on the idea that statistical learning theory is an important small-set learning method, which is effective in many applications. If some structural information in high-dimensional space can be obtained as transcendental knowledge, the small-set learning method can be used to reveal the true manifold with fewer samples. In the case of multi-pose face images, we believe that the multi-pose face image sets of different persons share a similar structure in observation space when their heads rotate in the same direction. Therefore we can obtain the structural information in high-dimensional space by learning large sets of data for a few persons and then apply it to the learning in small set cases. We describe the learning method in detail and then validate it through reconstruction and synthesis of multi-pose face images based on small sets.

Matthew Brand established a nonlinear mapping from high-dimensional sample space to low-dimensional vector space, recovering a Cartesian coordinate system for the manifold from which the data is sampled [12]. The principle of his method is to first decompose the data into locally linear low-dimensional patches then merge them into a single low-dimensional coordinate system and finally compute forward and reverse mappings between samples and coordinate spaces. For constructing the mapping, the method need large sets to accomplish the training.

The method we propose in this paper is an appearance-based method. In face recognition field, there are appearance-based methods published, such as SLAM [13], Eigenfaces [1] and Illumination Cone method [14]. They demonstrated the power of appearance-based methods both in ease of implementation and in accuracy. Especially Illumination Cone method can synthesize faces after it learns a face 3-D geometry from a few face images. However, there are three assumptions in Illumination Cone method: The surface is Lambertian, The object is convex, and the pose is fixed and frontal. The important difference of our method with Illumination Cone method is that we focus on the common law when different faces move. So we learn structural information in high-dimensional space with large data sets of a few faces. Then using the structural information we learned and a small data set of a different face we synthesize new face images for this different face. As a result, we need more face images before we synthesize new face images for a different face than Illumination Cone method does. However, this is a new way to study reconstruction of multi-pose face images. We conclude that there are indeed common laws when different faces move. And the common law can be shared with other faces reconstruction. In our experiments, we firstly used many multi-pose images of person A to learn the knowledge of head movement. Then we applied the obtained knowledge to construct analytical model for person B with fewer samples. (In our experiments, about 20 samples are enough to construct new analytical model for person B.). Our objective is to establish analytical mapping to describe the manifold of data set on small sets. Furthermore, there are no Lambertian and convex assumptions in our method. It makes our method easy to use.

Other sections of the paper are organized as follows. Section 2 of this paper gives an overview of NDR. We mainly introduce the LLE algorithm and some previous works based on it. The analysis of dimensionality reduction of multi-pose face images applying LLE is given in Section 3. This section also achieves a discussion of the method. In Section 4, a nonlinear regression method based on Support Vector Regression (SVR) is given to obtain the parametric mapping in small set cases. The results of face image synthesis based on small sets are shown. Our conclusions are provided in the final section of the paper.

2 An Effective Nonlinear Dimensionality Reduction Method

Tenenbaum, Silva and Langford presented the nonlinear dimensionality reduction method ISOMAP in 2000 [11]. ISOMAP computes geodesic distance along a manifold and then applies Multidimensional Scale (MDS) to reduce dimensionality. In fact, this method maintains the data distribution manifold by preserving the geodesic distance between data. Sam Roweis and Lawrence Saul implemented nonlinear dimensionality reduction by Locally Linear Embedding (LLE) [10]. LLE maps the high-dimensional data to a single global coordinate system in a manner that preserves the neighboring relationships. The two methods have been successfully applied to reduce the dimensionalities of artificial data and real-world data, such as “Swiss roll” data and face images. In Reference [10], an experiment of dimensionality reduction of face images is demonstrated. The intrinsic dimension and structure of the data set, which includes changes of facial expression and pose, were discovered in the experiment. Other work, in which the true manifold structure of data is revealed, has also been conducted [15, 16]. Dennis DeCoste then developed the Kernel Locally Linear Embedding method [17]. All these previous applications of nonlinear dimensionality reduction mainly focused on data analysis.

As further background, the standard LLE algorithm is given as follows:

Suppose the observation data consist of N real-valued vectors Y_i, $i = 1, \ldots, N$, $Y_i \in R^D$, sampled from some smooth underlying manifold.

Step 1: To compute the neighbors of each data point, we should define k as the number of nearest neighbors or the number of points in the super sphere of fixed radius$r$.

Step 2: Define neighbors representation cost function as Equation 1:

$$\varepsilon(W) = \sum_i \left| Y_i - \sum_j W_{ij} Y_j \right|. \tag{1}$$

If Y_j does not belong to the neighbors of Y_i, $W_{ij} = 0$. Considering translation invariance, we add this constraint:

$$\sum_j W_{ij} = 1. \tag{2}$$

We can get the weights W_{ij} that best linearly reconstructs Y_i from its neighbors by solving the constrained least-squares problem:

$$W = \arg\min_W \varepsilon(W). \tag{3}$$

Step 3: Fixing weights, compute the d-dimensional embedding vectors $\{X_i\}_{i=1\ldots,N} \in R^d$, which are best reconstructed by W_{ij}. This can be achieved by minimizing the following cost function:

$$\varepsilon(X) = \sum_i \left| X_i - \sum_j W_{ij} X_j \right|. \tag{4}$$

So solve X:

$$X = \arg\min_X \varepsilon(X). \tag{5}$$

The details and proofs of this algorithm can be found in Reference [18].

There is no doubt that LLE is an efficient dimensionality reduction method. LLE, however, lacks a process for mapping between the observations and the embedded space. Thus, dimensionality reduction can only be implemented in training sets. The test data can be projected to the embedded space. Because of this disadvantage, the applications of LLE are confined to the area of data analysis. In addition, as a kind of local linear method, LLE has the drawback that it should be based on large sample sets to achieve an intuitive presentation of the nonlinear manifold.

3 Dimensionality Reduction of Multi-pose Face Images by LLE

As a result of head rotation, the human face exhibits variety. Multi-pose face images, especially images formed as a result of horizontal and vertical rotation, cause difficulty in face recognition. Some examples of multi-pose face images are shown in Figure 1. A sequence of face images are obtained when the human head finishes a rotation with the model from left to right or up to down. Figure 2 exhibits an image sequence with head rotation from left to right.

Although the observation data are high dimensional, the intrinsic dimensionality of some face images may be very low. In lower-dimensional space, especially visualized space, it is easier for us to find the underlying structure through observation. LLE is applied to reduce the dimensionality of face images in our experiment. In the experiment,

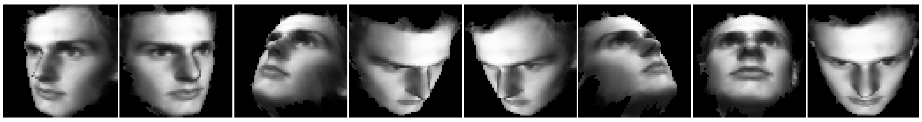

Figure 1: Examples: multi-pose face images.

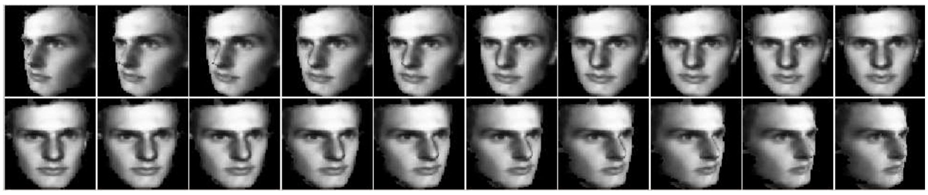

Figure 2: A sequence of face images with head rotating from left to right.

a sequence includes 160 images, sampled when the head rotates. Figure 3 shows the result of dimensionality reduction by LLE on the sequence.

In Figure 3, the projected points by LLE algorithm scatter along a smooth curve and the arrangement of points on the curve associates with the rotation trend.

Although LLE is an efficient dimensionality reduction method, it lacks a process for mapping between the observations and the embedded space. In addition, the revealed underlying manifold can only be observed subjectively. Thus the applications of LLE are confined to the area of data analysis. Furthermore, as a kind of local linear method, the observation of manifold can bring imprecision and obscureness in small set cases. Figure 4 shows a result of dimensionality reduction by LLE on a small set with 20 samples.

In Figure 4, regression method is used to obtain the three-dimensional curve of the manifold and eight new points are sampled along the regressed curve in embedding. Because of the lack of mapping, we cannot know the corresponding source images of these generated samples.

4 Establishing the Analytical Mapping for LLE Based on Nonlinear Regression

In real-world applications, however, the samples are always sparse. How can the high-dimensional structure be learned and the mapping established in small set cases? We conjecture that the mappings and the manifold of multi-pose face images in high-dimensional space are similar for different persons when the same type of head rotation occurs. Based on this, we propose solving the problem as followings. High-density multi-pose images of some persons are sampled and used as the training set for extracting the common information of distributional structure and obtaining some knowledge of the mapping. Utilizing the form of the high-dimensional curve as transcendental knowledge, the definite mapping can be obtained with fewer multi-pose images of other persons. The detailed learning method and algorithm are given as follows.

Suppose the observation data consist of $Y_j(j = 1, 2..N)$, where j is the index of sample and N is the number of samples, $Y_j \in R^D$, $Y_j = [y_1, y_2, ..., y_D,]^T$, where $y_i \in R(i = 1, 2..D)$.

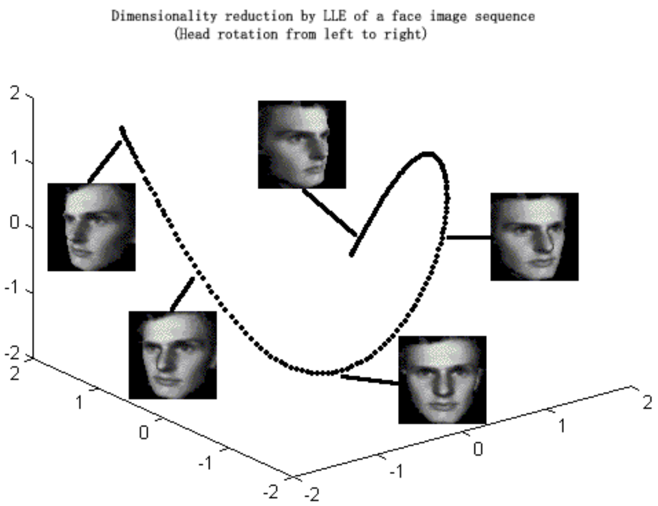

Figure 3: Dimensionality reduction by LLE of a face image sequence sampled when the head rotates from left to right.

The projection of Y_j by LLE algorithm is X_j, $X_i \in R^d$, where d is the dimensionality of embedded space.

For every dimensionality of observation the data Y_j, it can be regressed as:

$$y_i = f_i(X). \tag{6}$$

Thus the analytical mapping is described as:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_D \end{bmatrix} = \begin{bmatrix} f_1(X) \\ f_2(X) \\ \cdots \\ f_D(X) \end{bmatrix} = F(X). \tag{7}$$

Equation (7) gives the method of reconstruction and synthesis of high-dimensional data. The following steps describe the learning model for $f_i(\bullet)$ in detail.

Step 1: On a large training set, apply LLE to project face images to a d-dimensional space. How to determine the value of d is question need to consider firstly. The LLE algorithm itself can only give the range of the value of d, but not the precise value. Sam T. Roweis has indicated that the intrinsic value of d can be estimated by analyzing a reciprocal cost function, in which reconstruction weights derived from the embedding vectors X_i are applied to the data points Y_i [10]. The detailed method, however, is not described in that paper. Marzia Polito and Pietro Perona proposed in their paper that the value of d should be known in advance [19]. They believe that the value of d can be determined by using other dimensionality reduction methods. We use ISOMAP to get the value of the intrinsic dimensionality of a given data set. Figure 5 shows the relationship

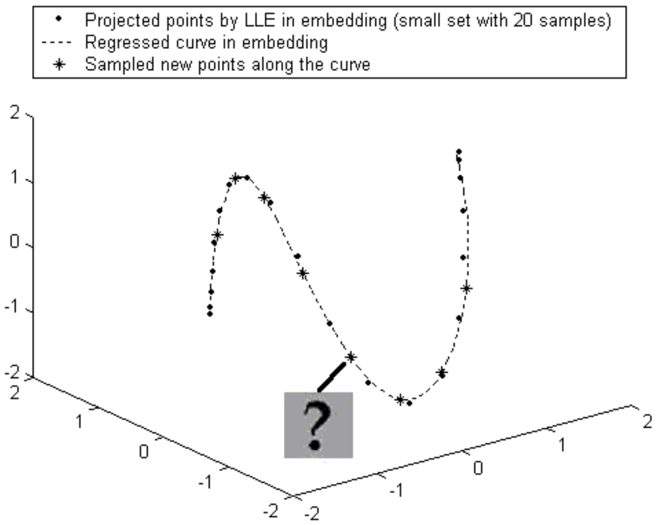

Figure 4: Dimensionality reduction by LLE of a multi-pose face image set with 20 samples. Points denote true samples and asterisks denote generated samples obtained along the regressed curve in embedding.

between dimensionality and residual variance of a given set applying ISOMAP algorithm. For more precision, the value of d is chosen as 3 in our experiments.

We believe that the multi-pose face image sets of different persons share a similar structure in observation space when their heads rotate in the same direction. Thus the function $f_i(\bullet)$ describing the process of dimensionality reduction of different persons has similarities, such as the function form.

Step 2: The function $f_i(\bullet)$ is defined as

$$y_i = f_i(X) = \sum_{j=1}^{l} \alpha_{ij} k(X, X_j) + b, \tag{8}$$

where $X_j(j = 1, 2, ..., l)$ is the training sample and l is the number of samples in training set. The kernel function $k(X, X_i)$ decides the property of $f_i(\bullet)$. The parameters are a_{ij}. In this step, we have enough samples of different persons to learn the proper form of $f_i(\bullet)$. Appropriate selection of the kernel function and its parameters will exhibit a good fit not only on the training set but also on a test set with fewer samples.

In order to get useful information for function and parameter selection, it is necessary to define a cost function to evaluate the performance of regression. We classify the l samples in training set as two types ω_1, ω_2, l_1 samples in ω_1 and l_2 samples in ω_2 ($l_1 + l_2 = l$). Samples in ω_1 are used for training to estimate the function form and parameters. Samples in ω_2 are used to evaluate the performance of regression. We give the squared error on the training set as follows:

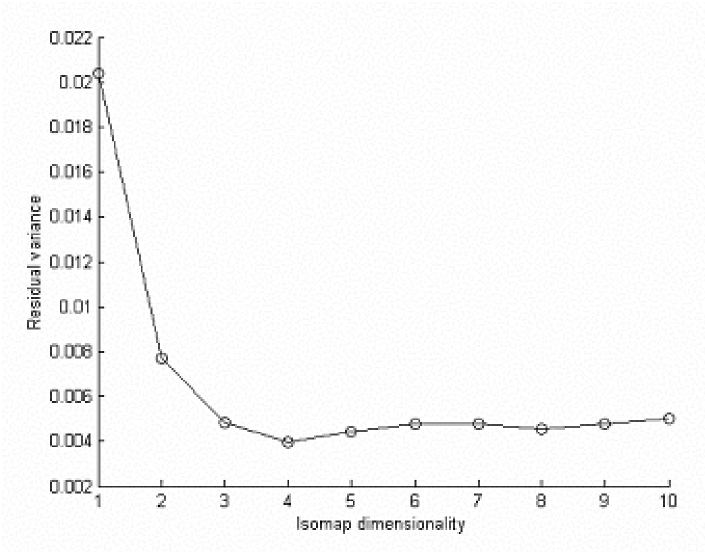

Figure 5: The relationship between dimensionality and residual variance applying ISOMAP.

$$\varphi(F) = \sum_{j=1}^{l_2} \left\| Y_j - F(X_j) \right\| \quad Y_j \in \omega_2, j = 1,2,...,l_2. \tag{9}$$

Equation (9) can be used as a standard to choose proper kernel function and parameters. By experiments, polynomial function and radial base function (RBF) are suitable kernels for function $f_i(\bullet)$. The two kinds of kernels are shown in (10) and (11).

$$k(X,X_i) = < X,X_i >^p . \tag{10}$$

$$k(X,X_i) = e^{-\frac{\|X-X_i\|^2}{2\sigma^2}}. \tag{11}$$

In Equation (10), $< X,X_i >^p$ means p-norm of X and X_i. By this step, some useful information about the function $f_i(\bullet)$ is obtained.

Step 3: After the form of the function and the proper kernel parameters are determined, these are regarded as a priori knowledge and the function learning is generalized to the small sets. In this step, data in a small test set are projected to embedded space by LLE. As a result, we have two small test sets. Y_{test} is the data of face images, which is in high-dimensional space. X_{test} is the corresponding point in lower-dimensional embedding, which is the projection of Y_{test} by LLE. Utilizing the form of the function $f_i(\bullet)$ obtained in step 2, the definite mapping for specific person can be learned by applying the method of statistical learning theory on the small sets X_{test} and Y_{test}. In this experiment, SVR is an effective method for determining the parameters a_{ij} and b in function $f_i(\bullet)$ [20] [21] [22]. The series of functions $f_i(\bullet)$, denoting the reverse transform from low-dimensional space to high-dimensional space, is obtained.

Step 4: In order to validate the results, we can use the analytical form $Y = F(X)$ to synthesize new face images from generated low-dimensional points. Firstly, the distributive curve of the embedded space samples is learned by regression methods, such as SVR. Then new low-dimensional points can be sampled from the curve. For a new low-dimensional sample X_{new}, the corresponding high-dimensional data in image space can be computed by the function learned before:

$$Y_{new} = F(X_{new}) = \begin{bmatrix} y_{new1} \\ y_{new2} \\ \cdots \\ y_{newD} \end{bmatrix} = \begin{bmatrix} f_1(X_{new}) \\ f_2(X_{new}) \\ \cdots \\ f_D(X_{new}) \end{bmatrix}. \tag{12}$$

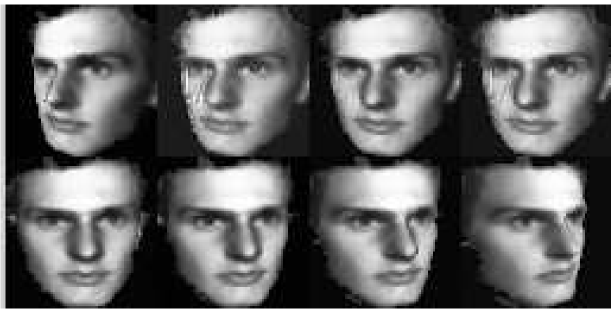

Figure 6: Eight synthesized face images, which is corresponding to the generated three-dimensional points in Figure 4.

Figure 4 shows a training set with only 20 true samples, the regression curve for this set, and 8 generated samples in three-dimensional embedding (Every point represents a projection of true image in the training set and every asterisk represents a generated sample along the regression curve.). The eight synthesized images applying our method are shown in Figure 6.

The experiment above is implemented on simulated face data. For more convictive results, we constructed a real-world face images data, which includes four types (vertical pose rotation, horizontal pose rotation and two kinds of mixed rotations) of images from 20 persons. Our method is used on these real-world face data. More examples of synthesized images are displayed in Figures 7 and Figures 8.

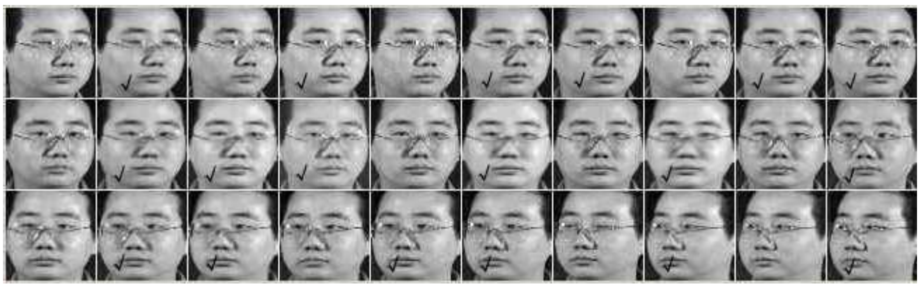

Figure 7: Example1: synthesized images based on 20 real-world face images.

Figure 8: Example2: synthesized images based on 20 real-world face images.

Under practical conditions, many factors can affect the results of experiments. Pose rotations are always accompanied by changes of expression and other uncertain factors. The model, however, only considers the variety of pose, which restricts it from achieving a better performance. As a result, as shown in Figures 7 and Figures 8, several synthesized face images have flaws.

5 Conclusions

As an elegant method for dimensionality reduction, LLE can bring out the underlying manifold of observation data in an embedded space, but lacks an effective mapping between source data and output data. In this paper, a nonlinear method is proposed to obtain a definite mapping for LLE, which also achieves an analytical representation of the manifold of high-dimensional data. By learning the common information from high-density data sets in advance, methods of statistical learning theory are applied to establish the mapping on small sets. Our experiments of synthesis for multi-pose face images prove our idea and algorithm are correct and effective. Our work shows the significance that the underlying manifold of high-dimensionality data can be analytically described with few samples by applying NDR and small-set based learning methods.

There are, however, some open problems. In the future, we plan to establish a unified model of pose changes not only for a specific person but also for a variety of human faces. Other factors, such as expression and illumination, will also be considered. LLE has the potential to be used for recognition. How can our method be applied to face recognition? This is an issue that our ongoing research will address.

References

[1] M.Turk, A. Pentland, Eigenfaces for Recognition, Journal of Cognitive Neuroscience, 1991, (3) 71-86.

[2] Marian Stewart Bartlett, Terrence J. Sejnowski, Independent components of face images: A representation for face recognition, Proceedings of the 4^{th} Annual Jount Symposium on Neural Computation, Pasadena, CA, May 17, 1997

[3] Marian Stewart Bartlett, Face image analysis by unsupervised learning and redundancy reduction, Ph.D. Thesis at University of California, San Diego (1998)

[4] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, Eigenfaces vs. Fisherfaces: Recognition using class specfic linear projection, IEEE Trans. on PAMI, Vol. 19, No. 7, 711-720, 1997.

[5] Penio S Penev, Joseph J Atick, Local Feature Analysis: A general statistical theory for object representation, Network: Computation in Neural Systems 7 (3) (1996) 477-500.

[6] Penio S Penev, Local Feature Analysis: A statistical theory for information representation and transmission, Ph.D. Thesis at The Rockefeller University (1998)

[7] B. Scholkopf, A. Smola, and K.-R. Muller. Nonlinear component analysis as a kernel eigenvalue problem. Neural Computuation, 10:1299-1399, 1998.

[8] Francis R. Bach, and Michael I. Jordan, Kernel Independent Component Analysis, Report No. UCB/CSD-01-1166, University of California, Berkeley, June 2002

[9] S.Mika, G.Ratsch, J.Weston, B.Scholkopf, and K.-R.Muller. Fisher discriminant analysis with kernels. In Proceedings of IEEE Neural Networks for Signal Processing Workshop, 1999.

[10] Sam Roweis, Lawrence Saul, *Nonlinear dimensionality reduction by locally linear embedding*, *Science*, v.290 no.5500, Dec.22, 2000. pp. 2323–2326.

[11] Joshua B. Tenenbaum, Vin de Silva, John C. Langford, A Global Geometric Framework for Nonlinear Dimensionality Reduction, *Science*, v.290 no. 5500, Dec.22, 2000. pp.2319–2322

[12] Matthew Brand, Charting a Manifold, In Suzanna Becker, Sebastian Thrun, and Klaus Obermayer, Advances in Neural Information Processing Systems 15, Cambridge, MA, MIT Press

[13] H.Murase and S.Nayar. Visual learning and recognition of 3-D objects from appearance. Int. J. Computer Vision, 14(5-24), 1995.

[14] Athinodoros S. Georghiades, Peter N. Belhumeur, David J. Kriegman, From few to many: Generative Models for recognition under variable pose and illumination. IEEE Trans. on Pattern Analysis and Machine Intelligence. Vol.23, No.6, 643-660.

[15] Lawrence K. Saul, Sam T. Roweis, An introduction to locally linear embedding, http://www.gatsby.ecl.ac.uk/∼roweis/lle/papers/lleintro.pdf

[16] Alice Zheng, Deconstructing motion, EECS Department U. C. Berkeley December 15, (2000), http://www.cs.berkeley.edu/∼alicez/lle_walking_print.ps

[17] Dennis DeCostem, Visualizing mercer kernel feature spaces via kernelized locally-linear embeddings, The 8th International Conference on Neural Information Processing, November (2001)

[18] David Gering, Linear and nonlinear data dimensionality reduction, April 17, (2002), http://www.ai.mit.edu/people/gering/areaexam/

[19] Marzia Polito, Pietro Perona, Grouping and dimensionality reduction by locally linear embedding, Neural Information Processing System (2001)

[20] Steve Gunn, Support vector machines for classification and regression, ISIS Technical Report ISIS-1-98, Image Speech Intelligent System Research Group, University of Southampton, (1998)

[21] A. J. Smola and B. Scholkopf. A tutorial on support vector regression. NeuroCOLT Technical Report NC-TR-98-030, Royal Holloway College, University of London, UK, (1998).

[22] V Vapnik. The nature of statistical learning theory. Springer Verlag, New York, 1995.