International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:6, No:2, 2012

Face Recognition using a Kernelization of Graph Embedding

Pang Ying Han, Hiew Fu San, Ooi Shih Yin

***Abstract*—**Linearization of graph embedding has been emerged as an effective dimensionality reduction technique in pattern recognition. However, it may not be optimal for nonlinearly distributed real world data, such as face, due to its linear nature. So, a kernelization of graph embedding is proposed as a dimensionality reduction technique in face recognition. In order to further boost the recognition capability of the proposed technique, the Fisher's criterion is opted in the objective function for better data discrimination. The proposed technique is able to characterize the underlying intra-class structure as well as the inter-class separability. Experimental results on FRGC database validate the effectiveness of the proposed technique as a feature descriptor.

***Keywords*—**Face recognition, Fisher discriminant, graph embedding, kernelization.

I. Introduction

RECENT studies claimed that intrinsic data geometrical structures may possess inherited discriminating power since high dimensional data is treated as a set of geometrically associated points lying on or nearly on a low dimensional manifold [1-3]. Graph embedding techniques, which seek data embedding via data neighbourhood preservation, are able to disclose the intrinsic manifold of a data. Representative instances that are widely implemented in face recognition include Laplacianface (or Locality Preserving Projection, LPP) optimally preserves the neighbourhood structure of a data set based on heat kernel nearest neighbour graph [4] and Neighbourhood Preserving Embedding (NPE) restricts neighbouring points in the high dimensional image space to be located within the same neighbourhood in the low dimension feature space in a similar relative spatial situation [5].

The inherited discriminating capability of these algorithms cannot be assured since real world data is too complicated to measure. To further enhance the discriminating capability of the graph embedding algorithms, a discriminant criterion is explicitly integrated. For examples, Marginal Fisher Analysis (MFA) [6], Locality Sensitive Discriminant Analysis (LSDA) [7] and Neighbourhood Preserving Discriminant Embedding (NPDE) [8] incorporate Fisher criterion (FC) to optimize the algorithm objective functions.

Pang Ying Han is with Faculty of Information Science and Technology, Multimedia University, Malaysia (phone: 606-252-3193; fax: 606-231-8840; e-mail: yhpang@mmu.edu.my).

Hiew Fu San is with Infineon Technologies Sdn. Bhd, Free Trade Zone, Batu Berendam, Melaka, Malaysia (email: fusan.hiew@infineon.com).

Ooi Shih Yin is with Faculty of Information Science and Technology, Multimedia University, Malaysia (phone: 606-252-3053; fax: 606-231-8840; e-mail: syooi@mmu.edu.my).

However, these discriminant techniques encode pattern information based on second order dependencies. But, those higher order dependencies in an image (e.g. the correlations among three or more pixels of an edge) have been neglected [9]. This information might capture pertinent data features. Hence, a nonlinear mapping could be used to map the data to a higher dimensional feature space to "unfold" the data manifold. With this, those discriminative nonlinear data structures can emerge under this new representation. Kernel trick allows this unfolding implicitly [9].

In this paper, a kernelization of graph embedding technique is proposed. To achieve superior discriminating capability, the proposed technique incorporates three mechanisms: a kernel trick, a Graph Embedding (GE) criterion and the Fisher's criterion (FC). The technique is namely as Kernel Discriminant Embedding (KDE). In KDE, the input data is first mapped into a higher dimensional feature space via the kernel trick for unfolding the data manifold to release the underlying nonlinear features. Then, the released underlying features are learned by GE and represented in GE coefficients. By optimizing FC, an optimal projection is sought to characterize the intra-class compactness while maximizing the inter-class separability.

This proposed technique overcomes the limitation of the traditional linear subspace techniques, i.e. Principal Component Analysis (PCA) [11] and Linear Discriminant Analysis (LDA) [12], for the data distribution assumption. Besides that, KDE also overcomes the limited success of the ordinary linearization of graph embedding due to its linear nature by incorporating kernel trick.

II. Kernel Discriminant Embedding

KDE utilizes kernel trick to project the input data onto a higher dimensional feature space, denoted as kernel space. The main purpose is to reveal the underlying intrinsic data structures in this new representation. In addition, KDE employs neighbourhood preserving criterion to learn local features of the data. Furthermore, KDE utilizes Fisher criterion to construct a discriminant projection by making the projected intra-class samples as compact as possible, while the projected samples from different classes are far apart.

A. Computation of Kernel Trick

Let $\{\boldsymbol{x}_i \in \mathbf{R}^d \mid i=1,\ldots, n\}$ be a set of d-dimensional vectors of face images. This input data is projected into a higher dimensional feature space, denoted as $\boldsymbol{F}$, via a nonlinear mapping, $\Phi : \boldsymbol{x}_i \in \mathbf{R}^d \rightarrow \boldsymbol{f}_i \in \boldsymbol{F} \left(= \mathbf{R}^t\right)$.

The inner product between the two mapped samples $\Phi(\boldsymbol{x}_i)$ and $\Phi(\boldsymbol{x}_j)$ in $\boldsymbol{F}$ can be computed via a kernel function:

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = [\Phi(\boldsymbol{x}_i).\Phi(\boldsymbol{x}_j)] \quad (1)$$

Since the dot product of the vectors can be computed as $\Phi(\boldsymbol{x}_i).\Phi(\boldsymbol{x}_j) = \Phi(\boldsymbol{x}_j)^{\mathrm{T}}\Phi(\boldsymbol{x}_i)$, alternatively, we can present the kernel in matrix form,

$$\boldsymbol{K} \equiv \Phi(\mathbf{X})^{\mathrm{T}}\Phi(\mathbf{X}) \quad (2)$$

where $\mathbf{X} = \{\boldsymbol{x}_i \in \mathbf{R}^d \mid i=1,\ldots, n\}$.

B. Formulation of Intra-class Coefficients Modelling

Let the mapped samples be a set of *t*-dimensional vectors in the feature space $\boldsymbol{F}$,$\{\Phi(\boldsymbol{x}_i) \in \mathbf{R}^t \mid i=1,\ldots, n$. The intra-class coefficients ω_{ij}^w reflect the contribution of the j^{th} neighbours to the reconstruction of the i^{th} data. $\omega_{ij}^w \neq 0$ if the pair of samples is from the same class, known as local neighbours; and $\omega_{ij}^w = 0$, otherwise. The intra-class coefficients matrix $\mathbf{W}^w$ can be calculated by minimizing the objective function,

$$\varepsilon^i(\mathbf{W}^w) = \left| \sum_{j=1}^{r} \omega_{ij}^w \left(\Phi(\boldsymbol{x_i}) - \Phi(\boldsymbol{x_j}) \right) \right|^2 \quad (3)$$

where $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ are from the same class.

Let $\boldsymbol{\Upsilon} \in \mathbf{R}^{t \times t'}$ be a transformation matrix and $\left\{ \boldsymbol{y}_i = \boldsymbol{\Upsilon}^{\mathrm{T}}\Phi(\boldsymbol{x}_i) \mid \boldsymbol{y}_i \in \mathbf{R}^{t'} \right\}$ are projected face vectors of $\left\{ \Phi(\boldsymbol{x}_i) \mid \Phi(\boldsymbol{x}_i) \in \mathbf{R}^t \right\}$, where *t'*<< *t*. In order to preserve the data local geometry, the following cost function is defined,

$$\begin{aligned}
\varepsilon_w(\boldsymbol{y}) &= \sum_{i=1} \left| \boldsymbol{y}_i - \sum_{j=1} \omega_{ij}^w \boldsymbol{y}_j \right|^2 \\
&= \sum_{i=1} \left(\boldsymbol{y}_i \boldsymbol{y}_i - 2\sum_{j=1} \omega_{ij}^w \boldsymbol{y}_i \boldsymbol{y}_j + \sum_{j=1}\sum_{k=1} \omega_{ij}^w \omega_{ik}^w \boldsymbol{y}_i \boldsymbol{y}_k \right) \\
&= \sum_{i=1} \left(\sum_{j=1} \delta_{ij} \boldsymbol{y}_i \boldsymbol{y}_j - 2\sum_{j=1} \omega_{ij}^w \boldsymbol{y}_i \boldsymbol{y}_j + \sum_{j=1}\sum_{k=1} \omega_{ki}^w \omega_{kj}^w \boldsymbol{y}_i \boldsymbol{y}_j \right) \\
&= \sum_{i=1}\sum_{j=1} \mathbf{M}_{ij} \boldsymbol{y}_i \boldsymbol{y}_j \\
&= \mathbf{Y}\mathbf{M}\mathbf{Y}^{\mathrm{T}}
\end{aligned} \quad (4)$$

where the matrix $\mathbf{M} \in \mathbb{R}^{n \times n}$,

$$M_{ij} = \delta_{ij} - \omega_{ij}^w - \omega_{ji}^w + \sum_{k=1} \omega_{ki}^w \omega_{kj}^w \text{ with } \delta_{ij} = \begin{cases} 0 & \text{for } i \neq j \\ 1 & \text{for } i = j \end{cases}$$

The matrix $\mathbf{M}$ is sparse matrix, where $\mathbf{M} = \left(\mathbf{I} - \mathbf{W}^w\right)^{\mathrm{T}} \left(\mathbf{I} - \mathbf{W}^w\right)$ with $\mathbf{I}$ is an identity matrix.

Referring to (4), we can have alternative expression for the objective function of calculating the intra-class coefficients,

$$\begin{aligned}
\varepsilon(\mathbf{W}^w) &= trace\left[\Phi(\mathbf{X}) \left(\mathbf{I} - \mathbf{W}^w\right)^{\mathrm{T}} \left(\mathbf{I} - \mathbf{W}^w\right) \Phi(\mathbf{X})^{\mathrm{T}} \right] \\
&= trace\left[\left(\mathbf{I} - \mathbf{W}^w\right) \Phi(\mathbf{X})^{\mathrm{T}} \Phi(\mathbf{X}) \left(\mathbf{I} - \mathbf{W}^w\right)^{\mathrm{T}} \right] \\
&= trace\left[\left(\mathbf{I} - \mathbf{W}^w\right) \boldsymbol{K} \left(\mathbf{I} - \mathbf{W}^w\right)^{\mathrm{T}} \right]
\end{aligned} \quad (5)$$

C. Formulation of Inter-class Coefficients Modelling

Let ω_{ij}^b denotes the inter-class coefficients where $\omega_{ij}^b \neq 0$ if the j^{th} sample is one of the *K* nearest neighbours of i^{th} sample with different class label, i.e. j^{th} sample is the inter-class neighbour of i^{th} sample, known as between-class neighbour; otherwise, $\omega_{ij}^b = 0$. The inter-class coefficients matrix $\mathbf{W}^b$ of the inter-class neighbour (j^{th} sample) of i^{th} sample can be sought by minimizing the following objective function,

$$\varepsilon(\mathbf{W}^b) = \sum_{i=1} \left| \Phi(\boldsymbol{x}_i) - \sum_{j=1}^{K} \omega_{ij}^b \Phi(\boldsymbol{x}_j) \right|^2 \quad (6)$$

Without loss of generality, the weights sum up to one for each point. In order to keep the projected samples of different classes far from each other, we maximize the following cost function,

$$\begin{aligned}
\varepsilon_b(\boldsymbol{y}) &= \sum_{i=1} \left| \boldsymbol{y}_i - \sum_{j=1} \omega_{ij}^b \boldsymbol{y}_j \right|^2 \\
&= \mathbf{Y}\mathbf{D}\mathbf{Y}^{\mathrm{T}}
\end{aligned} \quad (7)$$

where $\mathbf{D} = \left(\mathbf{I} - \mathbf{W}^b\right)^{\mathrm{T}} \left(\mathbf{I} - \mathbf{W}^b\right)$ where $\mathbf{I}$ is an identity matrix.

Hence, the cost function in (6) can be alternatively represented as,

$$\begin{aligned}
\varepsilon(\mathbf{W}^b) &= trace\left[\Phi(\mathbf{X}) \left(\mathbf{I} - \mathbf{W}^b\right)^{\mathrm{T}} \left(\mathbf{I} - \mathbf{W}^b\right) \Phi(\mathbf{X})^{\mathrm{T}} \right] \\
&= trace\left[\left(\mathbf{I} - \mathbf{W}^b\right) \Phi(\mathbf{X})^{\mathrm{T}} \Phi(\mathbf{X}) \left(\mathbf{I} - \mathbf{W}^b\right)^{\mathrm{T}} \right] \\
&= trace\left[\left(\mathbf{I} - \mathbf{W}^b\right) \boldsymbol{K} \left(\mathbf{I} - \mathbf{W}^b\right)^{\mathrm{T}} \right]
\end{aligned} \quad (8)$$

D. Discriminant Projection

KDE optimizes its objective function via Fisher criterion for a better discriminant projection. KDE minimizes $\varepsilon(\mathbf{W}^w)$ and maximizes $\varepsilon(\mathbf{W}^b)$ for calculating the optimized projection,

$$J_{KDE}(\boldsymbol{\Upsilon}_{opt}) = \arg\max_{\boldsymbol{\Upsilon}} \frac{|\varepsilon_b(y)|}{|\varepsilon_w(y)|}$$
$$= \arg\max_{\boldsymbol{\Upsilon}} \frac{\left|\boldsymbol{\Upsilon}^{\mathrm{T}}\Phi(\mathbf{X})\left(\mathbf{I}-\mathbf{W}^{b}\right)^{\mathrm{T}}\left(\mathbf{I}-\mathbf{W}^{b}\right)\Phi(\mathbf{X})^{\mathrm{T}}\boldsymbol{\Upsilon}\right|}{\left|\boldsymbol{\Upsilon}^{\mathrm{T}}\Phi(\mathbf{X})\left(\mathbf{I}-\mathbf{W}^{w}\right)^{\mathrm{T}}\left(\mathbf{I}-\mathbf{W}^{w}\right)\Phi(\mathbf{X})^{\mathrm{T}}\boldsymbol{\Upsilon}\right|}$$
$$= \arg\max_{\boldsymbol{\Upsilon}} \frac{\left|\boldsymbol{\Upsilon}^{\mathrm{T}}\Psi_{b}\boldsymbol{\Upsilon}\right|}{\left|\boldsymbol{\Upsilon}^{\mathrm{T}}\Psi_{w}\boldsymbol{\Upsilon}\right|} \quad (9)$$

where $\Psi^{b} = \boldsymbol{K}\left(\mathbf{I}-\mathbf{W}^{b}\right)\left(\mathbf{I}-\mathbf{W}^{b}\right)^{\mathrm{T}}\boldsymbol{K}$ and

$\Psi^{w} = \boldsymbol{K}\left(\mathbf{I}-\mathbf{W}^{w}\right)\left(\mathbf{I}-\mathbf{W}^{w}\right)^{\mathrm{T}}\boldsymbol{K}$, $\boldsymbol{K} = \Phi(\mathbf{X})^{\mathrm{T}}\Phi(\mathbf{X})$.

III. Justification

In face recognition, it is desired to construct a projection that maximizes the inter-class samples separability, while minimizing the intra-class samples compactness for better data discrimination. An example of a two-class classification problem is discussed in this section. Figure 1 illustrates the data distribution, as well as the optimal projections of PCA, LDA and KDE, represented as solid lines. The lines that are orthogonal to each projection direction are the optimal classification lines of each method, represented as dotted lines. From the figure, we observe that KDE is able to derive a discriminative projection for the data. The inter-class data are not overlapping on the KDE projection. In addition, the decision boundary of KDE can better separate the two data clusters compared with other techniques.

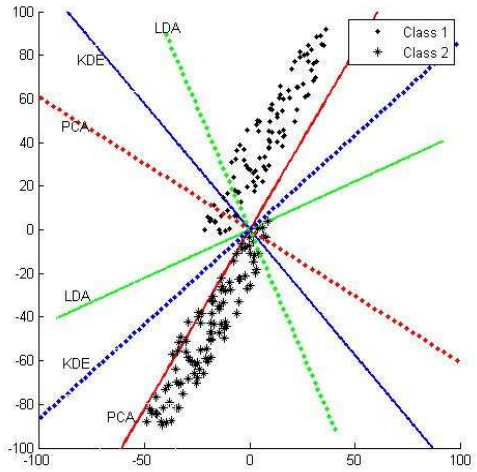

Fig. 1 Optimal projections and decision boundaries of PCA, LDA and KDE

IV. Experimental Results and Discussions

The performance of the proposed technique is assessed by using Face Recognition Grand Challenge Database (FRGC) [12]. Sample images of FRGC Database were collected at the University of Notre Dame. The FRGC data corpus contains high resolution still images taken under controlled lighting conditions and with unstructured illumination, 3D scans, and contemporaneously collected still images. The controlled images were taken under a studio setting, they are full frontal facial images taken under two lighting conditions (two or three studio lights) and with two facial expressions (smiling and neutral). On the other hand, the uncontrolled images were taken under varying illumination conditions; e.g., hallways, atria, or outdoors. Fig. 2 illustrates face images of FRGC database.

Fig. 2 Face image samples of FRGC database

The recognition performance of the proposed KDE is compared with other existing techniques, such as PCA, LDA, LPP, supervised LPP (SLPP), NPE and supervised NPE (SNPE). Note that the difference between LPP and supervised LPP is the neighbourhood assignment. In LPP, *k* nearest samples of a specific sample is assigned as its neighbours; these neighbours may be from the same class or the different classes. On the other hand, in SLPP, the same class samples of a specific sample are treated as its neighbours. Similar neighbourhood assignment is performed on NPE and SNPE.

FRGC database is partitioned into two sets: training and testing sets. The training set is used to establish the projection space for PCA, LDA, LPP, SLPP, NPE, SNPE and KDE; the testing set is used to evaluate the performance of the respective dimensionality reduction technique.

Two test strategies are carried out in this study:

- subject-dependent test. There is no overlapping in subject between the training and testing sets.
- subject-independent test. Both training and testing sets contain same subjects; but, there is no overlapping in sample between the training and testing sets.

In subject-dependent test, we are using a subset of FRGC database consisting 100 subjects with six training samples and six testing samples of each subject. In subject-independent test, 480 images (from 80 subjects with six samples of each) are employed as training set; whereas, another 480 images (from another 80 subjects with six samples of each) are adopted as testing set. The average error rates (AERs) (that is the average value of false accept rate (FAR) and false reject rate (FRR)) measured in this experiment serve as a performance measurement metric for the quality of the dimensionality reduction techniques.

We evaluate the effectiveness of KDE with polynomial and Gaussian kernels, as shown in Table I. Fig. 3 and 4 show the optimal results corresponding to the optimal parameter of each kernel. Gaussian kernel with parameter sigma, σ=10 demonstrates the best results among the kernels in both subject-dependent and subject-independent tests

TABLE I
PARAMETER RANGES USED IN THE EXPERIMENT

Kernel	Parameter Ranges	
Polynomial $k(x,y) = (x^{\mathrm{T}}y)^{\mathrm{d}}$	Degree (d)	Gamma (G)
	1~2	N/A
Gaussian $k(x,y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$	N/A	1, 10, 20

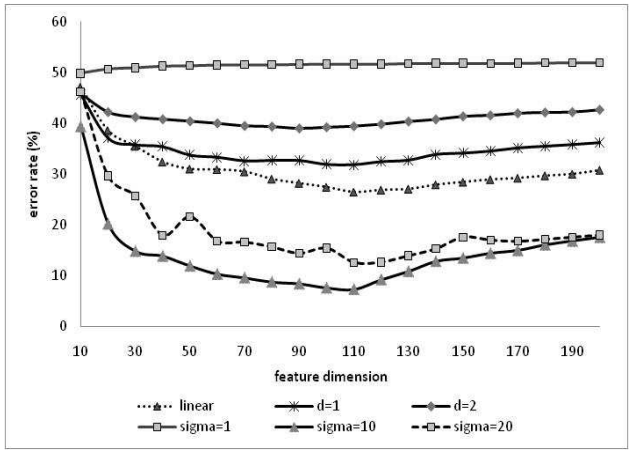

Fig. 3 Recognition error rates of KDE with different kernels in subject-dependent test

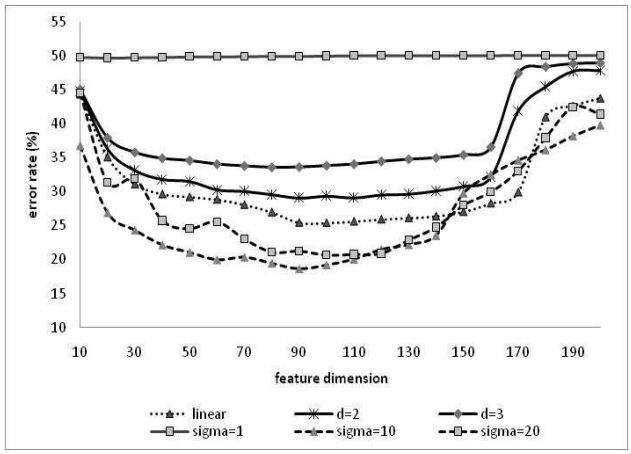

Fig. 4 Recognition error rates of KDE with different kernels in subject-independent test

Fig. 5 and 6 demonstrates the recognition performance of KDE with Gaussian kernel, σ =10 and other existing dimensionality reduction techniques (PCA, LDA, LPP, SLPP, NPE and SNPE) along with different feature dimensions.

Table II shows the optimal recognition performance corresponding to its feature dimension of the techniques. For LDA, all the samples are projected onto a subspace spanned by the *c*-1 largest eigenvectors, where *c* is the number of class, i.e. LDA lengths are 99 in the subject-dependent test and 79 in the subject-independent test, respectively. From the experimental results, it is observed that supervised methods including KDE, LDA, SLPP and SNPE achieve better recognition performance than non-supervised methods, such as PCA, LPP and NPE, in both tests.

SNPE and SLPP are supervised methods in such a way that they seek a projection that preserves the local geometry, formed by neighbours with a similar class label, based on respective objective function. Since SNPE and SLPP consider only the within-class information, their performances are not comparable to that of KDE. Results show that KDE obtains the highest recognition accuracy in both tests. This is because KDE is able to signify nonlinear features of face data and explicitly extract discriminating features via kernel trick, GE and Fisher criteria.

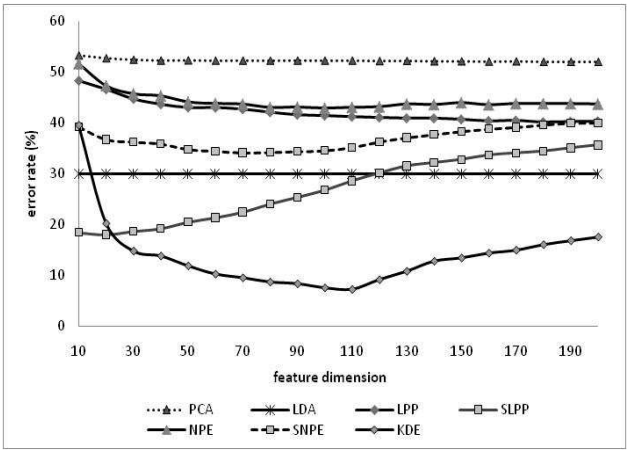

Fig. 5 Recognition error rates of KDE with Gaussian kernel, sigma=10 and other dimensionality reduction techniques in subject-dependent test

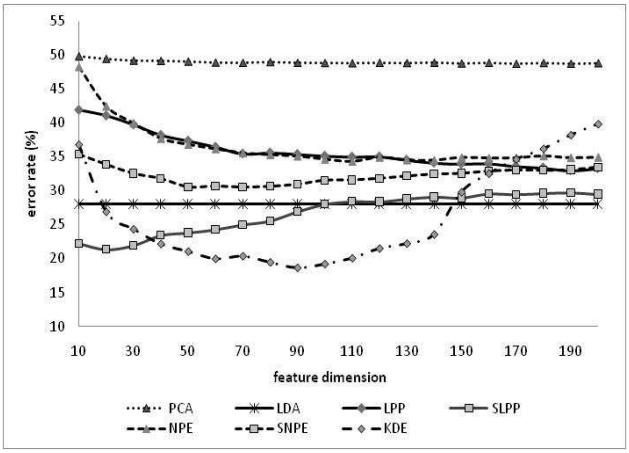

Fig. 6 Recognition error rates of KDE with Gaussian kernel, sigma=10 and other dimensionality reduction techniques in subject-independent test

TABLE II
RECOGNITION ERROR RATE OF KDE AND OTHER DIMENSIONALITY REDUCTION TECHNIQUES

Subject-dependent Test		
Methods	Error Rate (%)	Feature Dimension
Non-supervised techniques		
PCA	51.9	200
LPP	40.0	180
NPE	42.8	100
Supervised techniques		
LDA	29.8	99
SLPP	18.0	20
SNPE	34.1	70
KDE	7.3	110

Subject-independent Test		
Methods	Error Rate (%)	Feature Dimension
Non-supervised techniques		
PCA	48.7	190
LPP	32.9	190
NPE	34.3	110
Supervised techniques		
LDA	28.1	79
SLPP	21.3	20
SNPE	30.5	50
KDE	18.6	90

To evaluate the computational load of KDE and other techniques, the execution time (in elapsed CPU seconds) for training and recognition/ testing processes are recorded in Table III. These processes are executed in Matlab version 7.2 (R2006a) platform at the workstation of ASUS notebook Duo P8400 CPU with memory capacity of 2GB. The recorded training time (per second) is the time needed to construct projection space(s) from 600 training samples (100 subjects with six images per subject from FRGC database) during training stage; whereas, the recognition time (per second) is the time needed to project one new data onto the optimal projection space for computing optimal feature template.

The computation time of KDE in training is much greater than that of the other techniques. The time taken by KDE is about 4 times higher. The main reason is that the projection of input data onto higher dimensional kernel space consumes more time to retrieve nonlinear features of the data. However, recognition time is crucial in real recognition applications because recognition is an online process. From the table, we observe that the recognition time of RLPDE is only 0.004 seconds.

TABLE III
COMPUTATIONAL TIME (IN ELAPSED CPU SECONDS) OF KDE AND OTHER TECHNIQUES

Methods	Training Time (seconds)	Testing Time (seconds)
Non-supervised techniques		
PCA	5.361986	0.006534
LPP	4.118671	0.006179
NPE	3.012671	0.004524
Supervised techniques		
LDA	3.221592	0.004220
SLPP	4.223864	0.001321
SNPE	3.236567	0.002827
KDE	22.378431	0.004833

REFERENCES

[1] M. Belkin, P. Niyogi, P, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proc. of the Conference on Advances in Neural Information Processing System* 15, pp. 585-591, 2001.
[2] S.T. Roweis, L. Saul, "Nonlinear dimensionality reduction by Locally Linear Embedding," *Science* , vol. 290, no.5500, pp. 2323-2326, 2000.
[3] J. Tenenbaum, V. Silva, J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no.5500, pp. 2319-2323, 2000.
[4] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, "Face recognition using laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328-340, 2005.
[5] X. He, Deng Cai, S. Yan, H.J. Zhang, "Neighborhood Preserving Embedding," in *Proc. of the Tenth IEEE International Conference on Computer Vision*, pp. 1208-1213, 2005.
[6] S. Yan, D. Xu, B. Zhang, H.J. Zhang, Q. Yang, S. Lin, S, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40-51, 2007.
[7] D. Cai, X. He, K. Zhou, J. Han, H. Bao, "Locality sensitive discriminant analysis," in *Proc. of IJCAI*, pp. 708-713, 2007.
[8] Y.H. Pang, B.J. Andrew Teoh, Fazly Salleh Abas, "Neighbourhood Preserving Discriminant Embedding in face recognition, *Elsevier Journal of Visual Communication and Image Representation*, vol. 20, no. 8, pp. 532-542, 2009.
[9] M.H. Yang, "Kernel Eigenfaces vs. Kernel Fisherfaces: face recognition using kernel methods," in *Proc. of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 215-220, 2002.
[10] M. Turk, A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience* , vol. 3, no. 1, pp. 71-86, 1991.
[11] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 711-720, 1997.
[12] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, "Overview of the face recognition grand challenge," in *Proc. The IEEE International Conference on Computer Vision and Pattern Recognition*, CVPR05, pp. 947-954, 2005.