The Maximum Non-Linear Feature Selection of Kernel Based on Object Appearance

Mauridhi Hery Purnomo[1], Diah P. Wulandari[1], I. Ketut Eddy Purnama[1] and Arif Muntasa[2]
[1]*Electrical Engineering Department – Industrial Engineering Faculty, Institut Teknologi Sepuluh Nopember, Surabaya*
[2]*Informatics Engineering Department – Engineering Faculty, Universitas Trunojoyo Madura*
Indonesia

1. Introduction

Principal component analysis (PCA) is linear method for feature extraction that is known as Karhonen Loove method. PCA was first proposed to recognize face by Turk and Pentland, and was also known as eigenface in 1991 [Turk, 1991]. However, PCA has some weaknesses. The first, it cannot capture the simplest invariance of the face image [Arif et al., 2008b] , when this information is not provided in the training data. The last, the result of feature extraction is global structure [Arif, 2008]. The PCA is very simple, has overcome curse of dimensionality problem, this method have been known and expanded by some researchers to recognize face such as Linear Discriminant Analysis (LDA)[Yambor, 2000; A.M. Martinez, 2003; J.H.P.N. Belhumeur 1998], Linear Preserving Projection that known Lapalacianfaces [Cai, 2005; Cai et al, 2006; Kokiopoulou, 2004; X. He et al., 2005], Independent Component Analysis, Kernel Principal Component Analysis [Scholkopf et al., 1998; Sch¨olkopf 1999], Kernel Linear Discriminant Analysis (KLDA) [Mika, 1999] and maximum feature value selection of nonlinear function based on Kernel PCA [Arif et al., 2008b]. As we know, PCA is dimensionality reduction method based on object appearance by projecting an original *n-dimensional* (row*column) image into k eigenface where $k<<n$. Although PCA have been developed into some methods, but in some cases, PCA can outperform LDA, LPP and ICA when it uses small sample size.

This chapter will explain some theoretical of modified PCA that derived from Principal Component Analysis. The first, PCA transforms input space into feature space by using three non-linear functions followed by selection of the maximum value of kernel PCA. The feature space is called as kernel of PCA [Arif et al., 2008b]. The function used to transform is the function that qualifies Mercer Kernel and generates positive semi-definite matrix. Kernel PCA as been implemented to recognize face image[Arif et al., 2008b] and has been compared with some method such as Principal Component Analysis, Principal Linear Discriminant Analysis, and Linear Preserving Projection. The last, the maximum value selection has been enhanced and implemented to classify smiling stage by using Kernel Laplacianlips [Mauridhi et al., 2010]. Kernel Laplacianlips transform from input space into feature space on the lips data, followed by PCA and LPP process on feature space. Kernel Laplacianlips yield local structure in feature space. Local structure is more important than global structure. The experimental results show that, Kernel Laplacianlips using selection of non-linear function maximum value outperforms another methods [Mauridhi et al., 2010], such as Two Dimensional Principal Component Analysis (2D-PCA) [Rima et al., 2010], PCA+LDA+Support Vector Machine [Gunawan et al., 2009]. This chapter is composed as follows:

1. Principal Component Analysis in input space
2. Kernel Principal Component Analysis
3. Maximum Value Selection of Kernel Principal Component Analysis as Feature Extraction in Feature Space
4. Experimental Results of Face Recognition by Using Maximum Value Selection of Kernel Principal Component Analysis as Feature Extraction in Feature Space
5. The Maximum Value Selection of Kernel Linear Preserving Projection as Extension of Kernel Principal Component Analysis
6. Experimental Results of Simile Stage Classification Based on Maximum Value Selection of Kernel Linear Preserving Projection
7. Conclusions

2. Principal component analysis in input space

Over the last two decades, many subspace algorithms have been developed for feature extraction. One of the most popular is Principal Component Analysis (PCA) [Arif et al., 2008a, Jon, 2003; A.M. Martinez and A.C. Kak, 2001; M. Kirby and L. Sirovich, 1990; M. Turk and A. Pentland, 1991]. PCA has overcome Curse of Dimensionality in object recognition, where it has been able to reduce the number of object characteristics fantastically. Therefore, until now PCA is still used as a reference to develop a feature extraction.

Suppose a set of training image containing *m training image* $X^{(k)}$, $\forall k$, $k \in 1..m$, each training image has hxw size where $\forall H$, $H \in 1..h$ and $\forall W$, $W \in 1.w$. Each training image is represented as:

$$X = \begin{pmatrix} X_{1,1}^{(k)} & X_{1,2}^{(k)} & X_{1,3}^{(k)} & \ldots\ldots & X_{1,w-1}^{(k)} & X_{1,w}^{(k)} \\ X_{2,1}^{(k)} & X_{2,2}^{(k)} & X_{2,3}^{(k)} & \ldots\ldots & X_{2,w-1}^{(k)} & X_{2,w}^{(k)} \\ X_{3,1}^{(k)} & X_{3,2}^{(k)} & X_{3,3}^{(k)} & \ldots\ldots & X_{3,w-1}^{(k)} & X_{3,w}^{(k)} \\ X_{4,1}^{(k)} & X_{4,2}^{(k)} & X_{4,3}^{(k)} & \ldots\ldots & X_{4,w-1}^{(k)} & X_{4,w}^{(k)} \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ X_{h-1,1}^{(k)} & X_{h-1,2}^{(k)} & X_{h-1,3}^{(k)} & \ldots\ldots & X_{h-1,w-1}^{(k)} & X_{h-1,w}^{(k)} \\ X_{h,1}^{(k)} & X_{h,2}^{(k)} & X_{h,3}^{(k)} & \ldots\ldots & X_{h,w-1}^{(k)} & X_{h,w}^{(k)} \end{pmatrix} \quad (1)$$

Equation (1) can be transformed into one dimensional matrix form, by placing (t+1) th row to tth. If $\forall N$, $N \in 1..n$ and n=hxw, then Equation (1) can be changed into the following equation

$$X = \begin{pmatrix} X_{1,1}^{(k)} & X_{1,2}^{(k)} & \ldots\ldots & X_{1,w}^{(k)} & X_{1,w+1}^{(k)} & \ldots & X_{1,2w}^{(k)} & \ldots\ldots & X_{1,n-1}^{(k)} & X_{1,n}^{(k)} \end{pmatrix} \quad (2)$$

To expresses *m* training image set, it is necessary to composed Equation (2) in the following equation:

$$X = \begin{pmatrix} X_{1,1}^{(1)} & X_{1,2}^{(1)} & \ldots\ldots & X_{1,w}^{(1)} & X_{1,w+1}^{(1)} & \ldots\ldots & X_{1,2w}^{(1)} & \ldots\ldots & X_{1,n-1}^{(1)} & X_{1,n}^{(1)} \\ X_{1,1}^{(2)} & X_{1,2}^{(2)} & \ldots\ldots & X_{1,w}^{(2)} & X_{1,w+1}^{(2)} & \ldots\ldots & X_{1,2w}^{(2)} & \ldots\ldots & X_{1,n-1}^{(2)} & X_{1,n}^{(2)} \\ X_{1,1}^{(3)} & X_{1,2}^{(3)} & \ldots\ldots & X_{1,w}^{(3)} & X_{1,w+1}^{(3)} & \ldots\ldots & X_{1,2w}^{(3)} & \ldots\ldots & X_{1,n-1}^{(3)} & X_{1,n}^{(3)} \\ X_{1,1}^{(4)} & X_{1,2}^{(4)} & \ldots\ldots & X_{1,w}^{(4)} & X_{1,w+1}^{(4)} & \ldots\ldots & X_{1,2w}^{(4)} & \ldots\ldots & X_{1,n-1}^{(4)} & X_{1,n}^{(4)} \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ X_{1,1}^{(m-1)} & X_{1,2}^{(m-1)} & \ldots\ldots & X_{1,w}^{(m-1)} & X_{1,w+1}^{(m-1)} & \ldots\ldots & X_{1,2w}^{(m-1)} & \ldots\ldots & X_{1,n-1}^{(m-1)} & X_{1,n}^{(m-1)} \\ X_{1,1}^{(m)} & X_{1,2}^{(m)} & \ldots\ldots & X_{1,w}^{(m)} & X_{1,w+1}^{(m)} & \ldots\ldots & X_{1,2w}^{(m)} & \ldots\ldots & X_{1,n-1}^{(m)} & X_{1,n}^{(m)} \end{pmatrix} \quad (3)$$

The average of training image set of (Equation (3)) can be obtained by column-wise summation. It can be formulated by using the following equation

$$\overline{X} = \frac{\sum_{k=1}^{m} X_{1,N}^{(k)}}{m} \quad (4)$$

And $\forall N, N \in 1..n$.

The result of Equation (4) is in the row vector form, it has 1xN dimension. It can be re-written in the following equation

$$\overline{X} = [\overline{X}_{1,1}\ \overline{X}_{1,2}\ \overline{X}_{1,3}\ \overline{X}_{1,4}\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ \overline{X}_{1,n-1}\ \overline{X}_{1,n}] \quad (5)$$

The result of Equation (5) can be re-formed as original training image. To illustrate Equation (1), (2), (3) and (4), it is important to give an example of image average of the Olivetty Research Laboratory (ORL) face image database as seen In Figure 1

Fig. 1. average of ORL Face Image Database Using 3, 5 and 7 Face Image for Each Person

The zero mean matrix can be calculated by subtracting the face image values of training set with Equation (5). In order to perform the subtraction, both face image and Equation (5) must have the same size.

Therefore, Equation (5) can be replicated as many as *m* row size. The zero mean matrix can be formulated by using the following equation

$$\Phi_M = X_M - \overline{X} \tag{6}$$

$\forall M$, $M \in 1..m$. Furthermore, the covariance value can be computed by using the following equation

$$C = (X_M - \overline{X}).\left(X_M - \overline{X}\right)^T \tag{7}$$

As shown in Equation (7), *C* has *mxm* size and the value of *m*<<*n*. To obtain the principal components, the eigenvalues and eigenvectors can be computed by using the following equation:

$$\begin{aligned} &C.\Lambda = \lambda.\Lambda \\ &C.\Lambda = \lambda.I.\Lambda \\ &(\lambda I - C).\Lambda = 0 \\ &Det(\lambda I - C) = 0 \end{aligned} \tag{8}$$

The values of λ and Λ represent eigenvalues and eigenvectors of *C* respectively.

$$\lambda = \begin{pmatrix} \lambda_{1,1} & 0 & 0 & 0 & 0 \\ 0 & \lambda_{2,2} & 0 & 0 & 0 \\ 0 & 0 & .... & 0 & 0 \\ 0 & 0 & 0 & \lambda_{M-1,M-1} & 0 \\ 0 & 0 & 0 & 0 & \lambda_{M,M} \end{pmatrix} \tag{9}$$

$$\Lambda = \begin{pmatrix} \Lambda_{1,1} & \Lambda_{1,2} & ..... & \Lambda_{1,m-1} & \Lambda_{1,m} \\ \Lambda_{2,1} & \lambda_{2,2} & ...... & \Lambda_{2,m-1} & \Lambda_{2,m} \\ ...... & ...... & .... & ...... & ...... \\ \Lambda_{m-1,1} & \Lambda_{m-1,2} & ...... & \Lambda_{m-1,m-1} & \Lambda_{m-1,m} \\ \Lambda_{m,1} & \Lambda_{m,2} & ...... & \Lambda_{m,m-1} & \lambda_{m,m} \end{pmatrix} \tag{10}$$

Equation (9) can be changed into row vector as seen in the following equation

$$\lambda = \left[\lambda_{1,1}\ \lambda_{2,2}\ \lambda_{3,3}\ \ldots\ldots\ldots\ldots\ldots\ldots.\lambda_{m-1,m-1}\ \lambda_{m,m}\right] \tag{10}$$

To obtain the most until the less dominant features, the eigenvalues were sorted in descending order $\left(\lambda_1 > \lambda_2 > \lambda_3 > \ldots\ldots\ldots\ldots\ldots\ldots > \lambda_{m-1} > \lambda_m\right)$ and followed by corresponding eigenvectors.

3. Kernel principal component analysis

Principal Component Analysis has inspired some researchers to develop it. Kernel Principal Component Analysis (KPCA) is Principal Component Analysis in feature space [Sch¨olkopf et al., 1998; Sch¨olkopf et al., 1999; Arif et al., 2008b; Mauridhi et al., 2010]. Principally, KPCA works in feature space [Arif et al., 2008b]. Input space of training set is transformed into feature space by using Mercer Kernel that yields positive semi definite matrix as seen in the Kernel Trick [Sch¨olkopf et al., 1998; Sch¨olkopf et al., 1999]

$$k(X,Y) = (\phi(X), \phi(Y)) \tag{11}$$

Functions that can be used to transform are *Gaussian*, *Polynomial*, and *Sigmoidal* as seen in the following equation

$$k(X,Y) = \exp\left(\frac{-\,||X-Y||^2}{\sigma}\right) \tag{12}$$

$$k(X,Y) = (a(X.Y)+b)^d \tag{13}$$

$$k(X,Y) = \tanh(a(X.Y)+b)^d \tag{14}$$

4. Maximum value selection of kernel principal component analysis as feature extraction in feature space

The results of Equation (12), (13) and (14) will be selected as object feature candidates [Arif et al., 2008b, Mauridhi, 2010]. The biggest value of them will be employed as feature space in the next stage, as seen in the following equation

$$F = \max(\phi_i : R_i \xrightarrow[k(x,y)]{} F_i) \tag{15}$$

For each kernel function has yielded one matrix feature, so we have 3 matrix of feature space from 3 kernel functions. For each corresponding matrix position will be compared and will be selected the maximum value (the greatest value). The maximum value will be used as feature candidate. It can be represented by using the following equation

$$\phi(X) = \begin{pmatrix} \phi\left(X_{1,1}^{(k)}\right) & \phi\left(X_{1,2}^{(k)}\right) & \phi\left(X_{1,3}^{(k)}\right) & \ldots\ldots & \phi\left(X_{1,w-1}^{(k)}\right) & \phi\left(X_{1,m}^{(k)}\right) \\ \phi\left(X_{2,1}^{(k)}\right) & \phi\left(X_{2,2}^{(k)}\right) & \phi\left(X_{2,3}^{(k)}\right) & \ldots\ldots & \phi\left(X_{2,w-1}^{(k)}\right) & \phi\left(X_{2,m}^{(k)}\right) \\ \phi\left(X_{3,1}^{(k)}\right) & \phi\left(X_{3,2}^{(k)}\right) & \phi\left(X_{3,3}^{(k)}\right) & \ldots\ldots & \phi\left(X_{3,w-1}^{(k)}\right) & \phi\left(X_{3,m}^{(k)}\right) \\ \phi\left(X_{4,1}^{(k)}\right) & \phi\left(X_{4,2}^{(k)}\right) & \phi\left(X_{4,3}^{(k)}\right) & \ldots\ldots & \phi\left(X_{4,w-1}^{(k)}\right) & \phi\left(X_{4,m}^{(k)}\right) \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ \phi\left(X_{m-1,1}^{(k)}\right) & \phi\left(X_{m-1,2}^{(k)}\right) & \phi\left(X_{m-1,3}^{(k)}\right) & \ldots\ldots & \phi\left(X_{m-1,m-1}^{(k)}\right) & \phi\left(X_{m-1,m}^{(k)}\right) \\ \phi\left(X_{m,1}^{(k)}\right) & \phi\left(X_{m,2}^{(k)}\right) & \phi\left(X_{m,3}^{(k)}\right) & \ldots\ldots & \phi\left(X_{m,m-1}^{(k)}\right) & \phi\left(X_{m,m}^{(k)}\right) \end{pmatrix} \tag{16}$$

The biggest value of feature space is the most dominant feature value. As we know, feature space as seen on equation (16) is yielded by using kernel (in this case, training set is transformed into feature space using equation (12), (13) and (14) and followed by selection of the biggest value at the same position using equation (15). where feature selection in kernel space will be used to determine average, zero mean, covariance matrix, eigenvalue and eigenvector in feature space. These values are yielded by using kernel trick as nonlinear component. Nonlinear component is linear component (principal component) improvement. So, it is clear that the biggest value of these kernels is improvement of the PCA performance. The average value of Equation (16) can be expressed in the following equation

$$\phi\left(\overline{X}\right)=\frac{\sum_{k=1}^{m}\phi\left(X_{1,N}^{(k)}\right)}{m} \tag{17}$$

So, zero mean in the feature space can be found by using the following equation

$$\phi(\Phi_M)=\phi(X_M)-\phi\left(\overline{X}\right) \tag{18}$$

Where $\forall M, M\in 1..m$. The result of Equation (18) has mxm. To obtain the eigenvalues and the eigenvectors in feature space, it is necessary to calculated the covariance matrix in feature space. It can be computed by using the following equation

$$\phi(C)=(\phi(X)-\phi\left(\overline{X}\right)).\left(\phi(X)-\phi\left(\overline{X}\right)\right)^T \tag{19}$$

Based on Equation (19), the eigenvalues and the eigenvectors in feature space can be determined by using the following equation

$$\begin{aligned}&\phi(C).\phi(\Lambda)=\phi(\lambda).\phi(\Lambda)\\&\phi(C).\phi(\Lambda)=\phi(\lambda).I.\phi(\Lambda)\\&\left(\phi(\lambda)I-\phi(C)\right).\phi(\Lambda)=0\\&Det\left(\phi(\lambda)I-\phi(C)\right)=0\end{aligned} \tag{20}$$

The eigenvalues and eigenvectors yielded by Equation (20) can be written in following matrices

$$\phi(\lambda)=\begin{pmatrix}\phi(\lambda_{1,1}) & 0 & 0 & 0 & 0\\ 0 & \phi(\lambda_{2,2}) & 0 & 0 & 0\\ 0 & 0 & .... & 0 & 0\\ 0 & 0 & 0 & \phi(\lambda_{M-1,M-1}) & 0\\ 0 & 0 & 0 & 0 & \phi(\lambda_{M,M})\end{pmatrix} \tag{21}$$

$$\phi(\Lambda)=\begin{pmatrix}\phi(\Lambda_{1,1}) & \phi(\Lambda_{1,2}) & ..... & \phi(\Lambda_{1,m-1}) & \phi(\Lambda_{1,m})\\ \phi(\Lambda_{2,1}) & \phi(\lambda_{2,2}) & ...... & \phi(\Lambda_{2,m-1}) & \phi(\Lambda_{2,m})\\ ...... & ...... & .... & ...... & ......\\ \phi(\Lambda_{m-1,1}) & \phi(\Lambda_{m-1,2}) & ...... & \phi(\Lambda_{m-1,m-1}) & \phi(\Lambda_{m-1,m})\\ \phi(\Lambda_{m,1}) & \phi(\Lambda_{m,2}) & ...... & \phi(\Lambda_{m,m-1}) & \phi(\lambda_{m,m})\end{pmatrix} \tag{22}$$

To obtain the value of the most until the less dominant feature, the Equation (21) will be sorted decreasingly and followed by Equation (20) [Arif et al., 2008b, Mauridhi, 2010]. The bigger value of the eigenvalue in the feature space, the more dominant the corresponding eigenvector in feature space. The result of sorting Equation (21) can be shown in the following equation

$$\phi(\lambda)=\left[\phi(\lambda_{1,1})\ \phi(\lambda_{2,2})\ \phi(\lambda_{3,3})\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\phi(\lambda_{m-1,m-1})\ \phi(\lambda_{m,m})\right] \qquad (23)$$

5. Experimental results of face recognition by using maximum value selection of kernel principal component analysis as feature extraction in feature space

In this chapter, the experimental results of "The Maximum Value Selection of Kernel Principal Component Analysis for Face Recognition" will be explained. We use Olivetti-Att-ORL (ORL) [Research Center of Att, 2007] and YALE face image databases [Yale Center for Computational Vision and Control, 2007] as experimental material.

5.1 Experimental results using the ORL face image database

ORL face image database consist of 40 persons, 36 of them are men and the other 4 are women. Each of them has 10 poses. The poses were taken at different time with various kinds of lighting and expressions (eyes open/close, smiling/not smiling) [Research Center of Att, 2007]. The face position is frontal with 10 up to 20% angles. The face image size is 92x112 pixels as shown in Figure 2.

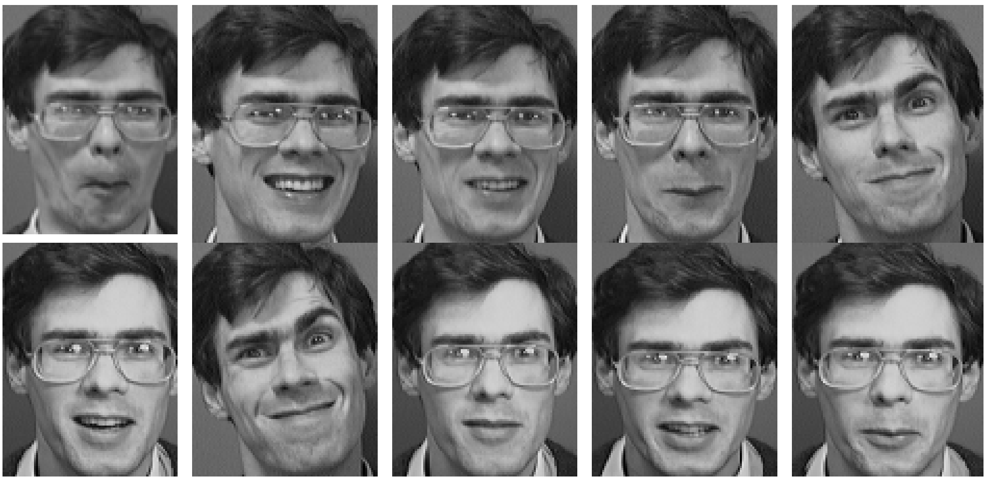

Fig. 2. Face Images of ORL Database

The experiments are employed for 5 times, and for each experiment 5, 6, 7, 8 and 9 poses for each person are used. The rest of training set, i.e. 5, 4, 3, 2 and 1, will be used as the testing [Arif et al., 2008b] as seen in Table 1

Scenario	Data Quantity			
	For Each Person		Total	
	Training	Testing	Training	Testing
1st	5	5	200	200
2nd	6	4	240	160
3rd	7	3	280	120
4th	8	2	320	80
5th	9	1	360	40

Table 1. The Scenario of ORL Face Database Experiment

In this experiment, each scenario used different dimension. The 1st, 2nd, 3rd, 4th, and 5th scenarios used 200, 240, 280, 320, and 360 dimensions respectively. The result of the 1st scenario can be seen on the Figure 3 [Arif et al., 2008b]

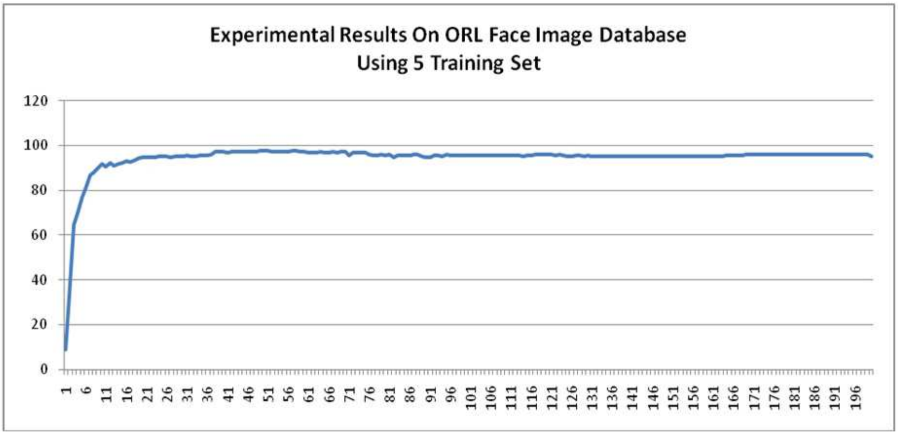

Fig. 3. Experimental Results on ORL Face Image Database Using 5 Training Set

Figure 3 shows that the more number of dimensional used, the higher recognition rate, but the recognition decreased on the certain dimension. As seen in Figure 3, recognition rate decreased into 95% when 200 dimensions were used. The first maximum recognition rate, which was 97.5%, occurred when 49 dimensions were used in this experiment [Arif et al., 2008b].

In the 2nd scenario, the maximum dimension used was 240 (240=40*6) training set. The first maximum recognition rate occurred when 46 dimensions were used, this was 99.375%. When 1 until 46 dimensions were used, recognition rate increased proportionally to the number of dimension used, but when 47 until the 240 dimensions were used, the recognition rate tended to be stable, with insignificant fluctuations as seen in Figure 4 [Arif et al., 2008b].

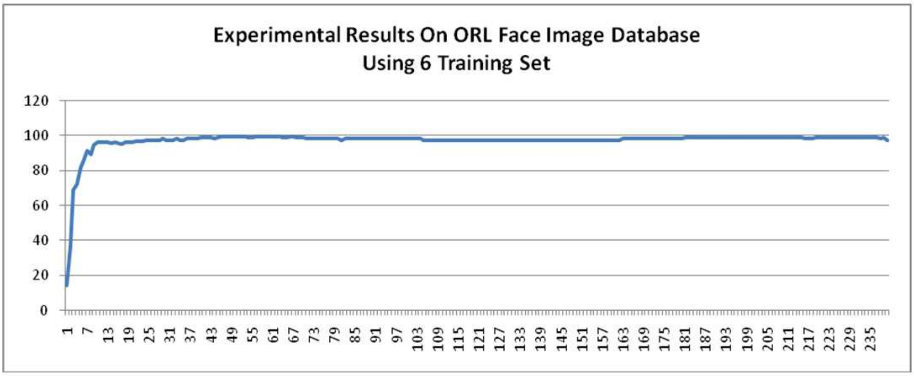

Fig. 4. Experimental Results on ORL Face Image Database Using 6 Training Set

In the 3rd scenario, training set used for each person was 7, whereas the number of dimensions used was 280. The more number of training set used, the number of dimension is increased. In this scenario, the maximum recognition rate was 100%, it occurred when 23 until 53 dimensions were used, whereas when more than 53 dimensions were used, recognition rate decreased to be 99.67% as seen in Figure 5 [Arif et al., 2008b].

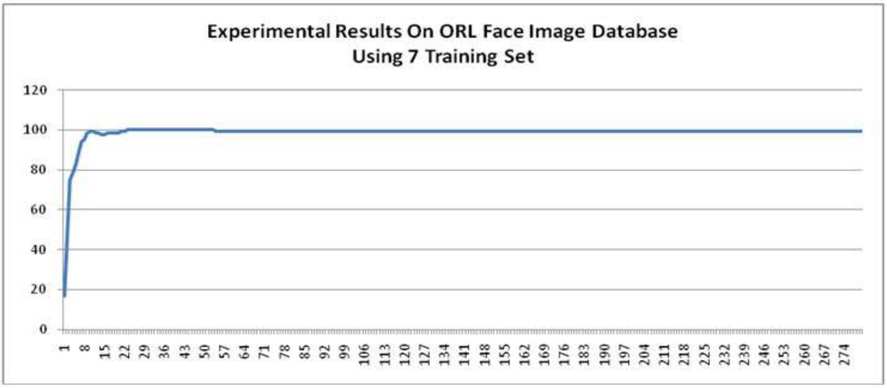

Fig. 5. Experimental Results on ORL Face Image Database Using 7 Training Set

Figure 6 is the experimental results of the 4th scenario. In this scenario, 8 training sets for each person were used, whereas the number of dimensions used was 320. Figure 6 shows that the recognition rate tended to increase significantly for experimental results using less than 23 dimensions, whereas 100% recognition rate occurred for experimental results using more than 24 dimensions [Arif et al., 2008b].

In the last scenario, 9 training sets were used, whereas the number of dimension used was 360, as seen in Figure 7. Similarly to the previous scenario, the recognition rate tended to increase when experimental used less than 6 dimensions, while using 7 dimensions resulted in 100% recognition rate, using 8 dimension resulted in 97% recognition rate, and 100% recognition rate was yielded from experimental results using more than 9 dimensions, as shown in Figure 7 [Arif et al., 2008b].

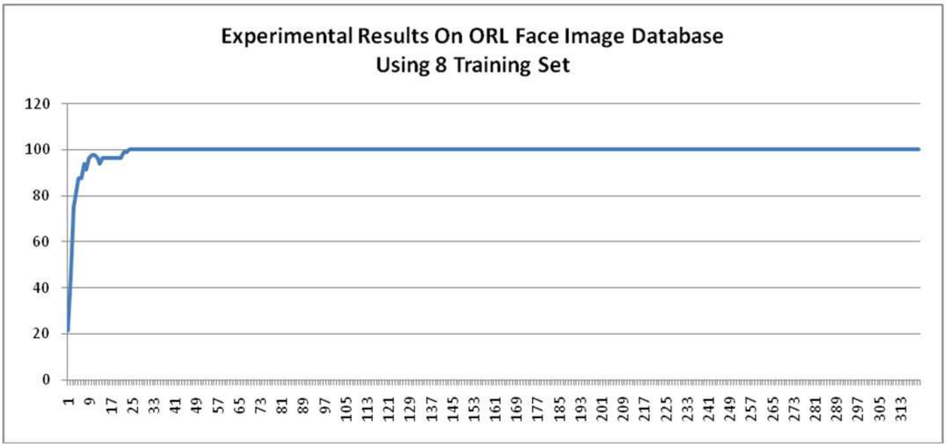

Fig. 6. Experimental Results on ORL Face Image Database Using 8 Training Set

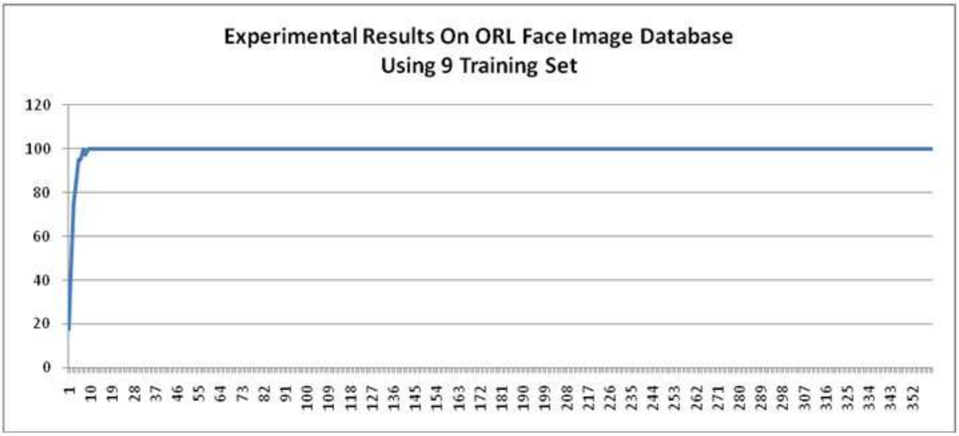

Fig. 7. Experimental Results on ORL Face Image Database Using 9 Training Set

The maximum recognition rate for all scenarios can be seen in Table 2. This table shows that the more number of training set used, the higher recognition rate achieved, whereas the first maximum recognition rate tended to occur on the lower dimension inversely proportional to the number of dimensions used [Arif et al., 2008b].

5.2 Experimental results using the YALE face image database

In this last experiment, the YALE face database was used. It contains 15 people, each of them were doing 11 poses. The poses were taken in various kinds of lighting (left lighting and center lighting), various expressions (normal, smiling, sad, sleepy, surprising, and wink) and accessories (wearing or not wearing glasses) [Yale Center for Computational Vision and Control, 2007] as shown in Figure 8.

Scenario	Number of Training Sample for Each Person	The First Maximum Recognition Rate	Dimension
1st	5	97.5	49
2nd	6	99.375	46
3rd	7	100	23
4th	8	100	24
5th	9	100	7

Table 2. The ORL Face Database Recognition Rate using Maximum Feature Value Selection Method of Nonlinear Function based on KPCA

Fig. 8. Face Sample of Images of YALE Database

The experiments were conducted for 6 scenarios, for each scenario, 5, 6, 7, 8, 9, and 10 training set were used. The rest of each data sample for every experiment, i.e. 6, 5, 4, 3, 2 and 1, were used as testing set as listed in Table 3 [Arif et al., 2008b].

Scenario	Data Quantity			
	For Each Person		Total	
	Training	Testing	Training	Testing
1st	5	6	75	90
2nd	6	5	90	75
3rd	7	4	105	60
4th	8	3	120	45
5th	9	2	135	30
6th	10	1	150	15

Table 3. The Scenario of the YALE Face Database Experiment

In the first scenario, 5 training sets were used, where the rest of the YALE data experiment was used for testing. In this scenario, the number of dimensions used was 75. The completed experimental results can be seen in Figure 8. This figure shows that the number of recognition rate increased significantly when less than 9 dimensions were used, which were 16.67% until 92.22%. Whereas the maximum recognition rate occurred when 13, 14, and 15 dimensions were used, that was 94.44% [Arif et al., 2008b]. For experimental results using more than 16 dimensions, the recognition rate fluctuated insignificantly as seen in Figure 8.

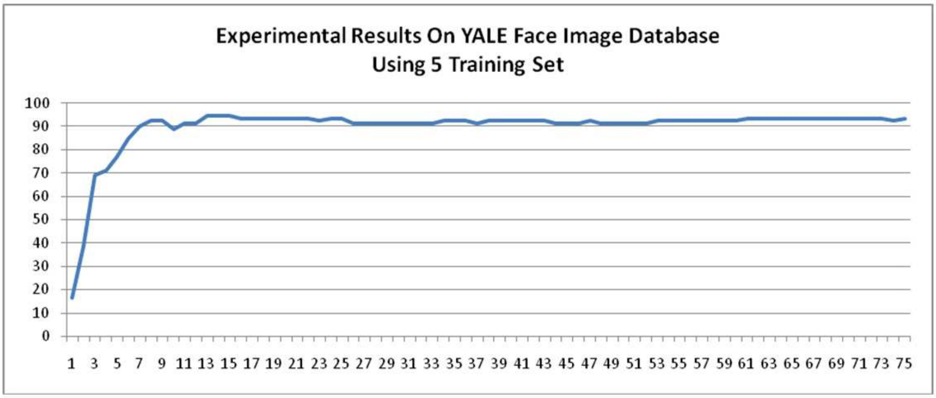

Fig. 8. Experimental Results on YALE Face Image Database Using 5 Training Set

The experimental results of the 2nd scenario were shown in Figure 9. The recognition rate increased from 22.67% until 97.33% when using less than 10 dimensions, recognition rate decreased insignificantly when using 16 dimensions, and recognition rate tended to be stable around 97.33% when experiments used more than 17 dimensions, [Arif et al., 2008b].

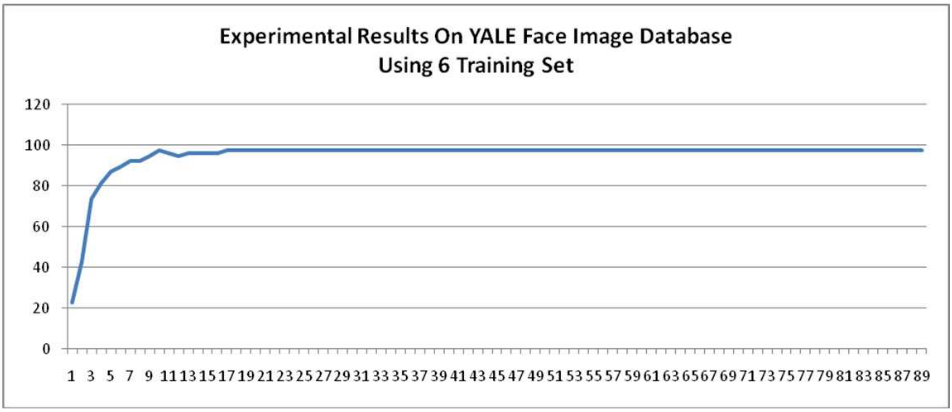

Fig. 9. Experimental Results on YALE Face Image Database Using 6 Training Set

Similarly, it occurred in the 3rd scenario. In this scenario, the recognition rate increased significantly when the number of dimensions was less than 13, though on the certain number of dimensions the recognition rate decreased. But when the number of dimensions used was more than 14, experimental results yielded its maximum rate, which is 98.33% as seen in Figure 10 [Arif et al., 2008b].

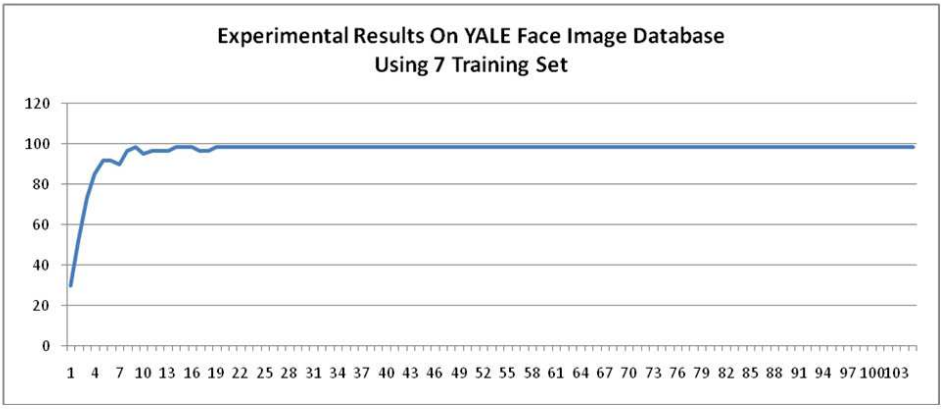

Fig. 10. Experimental Results on YALE Face Image Database Using 7 Training Set

In the last three scenarios as seen in Figure 11, 12, and 13, experimental results have shown that the recognition rate also tended to increase when the number of dimensions used was less than 7, whereas experimental results that used more than 8 dimensions achieved 100% recognition rate [Arif et al., 2008b].

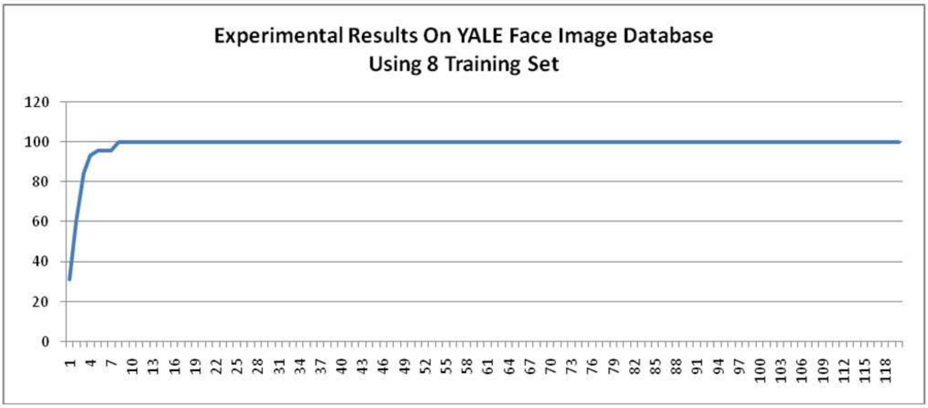

Fig. 11. Experimental Results on YALE Face Image Database Using 8 Training Set

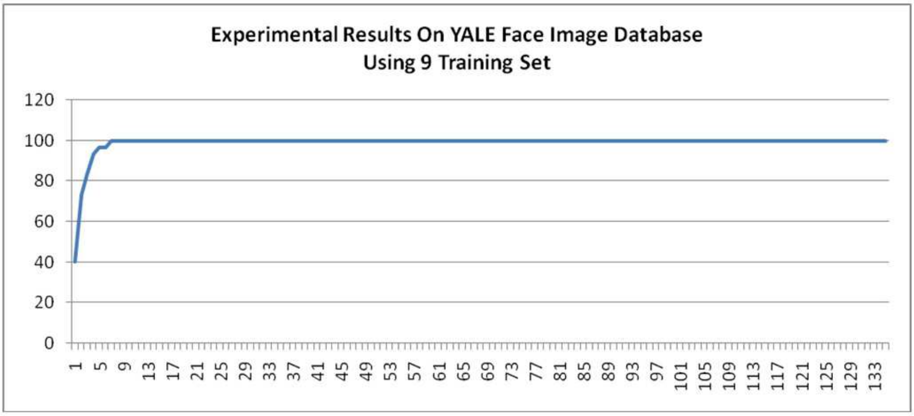

Fig. 12. Experimental Results on YALE Face Image Database Using 9 Training Set

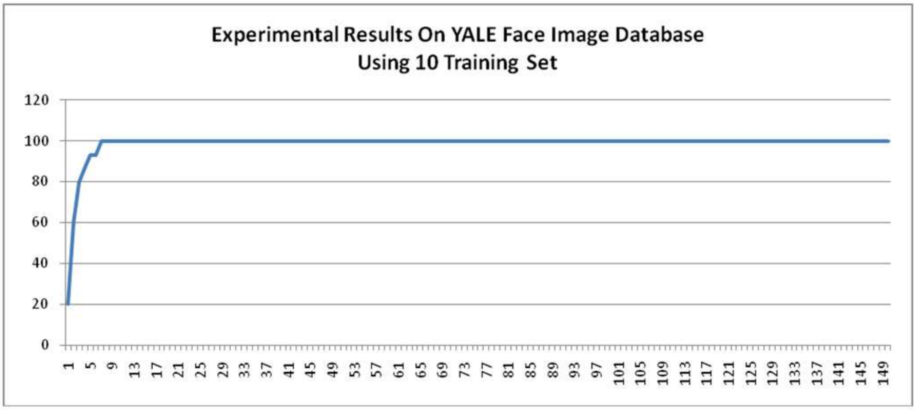

Fig. 13. Experimental Results on YALE Face Image Database Using 10 Training Set

As seen in Table 4, the maximum recognition rate for 5, 6, 7, 8, 9, and 10 training sets were 94.444%, 97.333%, 98.333%, 100%, 100% and 100% respectively. Based on Table 4, the maximum recognition rate increased proportionally to the number of training sets used. The more number of training set used, the faster maximum recognition rate is reached [Arif et al., 2008b].

The experimental results of the 1st, 2nd, and 3rd scenarios were compared to other methods, such as PCA, LDA/QR, and LPP/QR as seen in Table 5, whereas for the 4th and 5th scenarios were not compared, since they have achieved maximum result (100%). The recognition rate of 5, 6, and 7 training set, for both on the ORL and the YALE face database, "The Maximum Value Selection of Kernel Principal Component Analysis", outperformed the other methods.

Scenario	Number of Training Sample for Each Person	The First Maximum Recognition Rate	Dimension
1st	5	94.444	13
2nd	6	97.333	10
3rd	7	98.333	9
4th	8	100	8
5th	9	100	7
6th	10	100	7

Table 4. The YALE Face Database Recognition Rate using Maximum Feature Value Selection Method of Non linear Function based on KPCA

Database	Number of Training Set	The Maximum Recognition Rate (%)			
		PCA	LDA/QR	LPP/QR	The Maximum Value Selection of Kernel Principal Component Analysis
ORL	5	79.50	86.5	94.00	97.50
	6	83.13	91.25	94.37	99.38
	7	85.00	92.50	95.83	100.00
YALE	5	81.11	84.44	86.67	94.44
	6	85.33	86.67	94.67	97.33
	7	95.00	95.00	96.67	98.33

Table 5. The Comparative Results for Face Recognition Rate

6. The maximum value selection of kernel linear preserving projection as extension of kernel principal component analysis

Kernel Principal Component Analysis as appearance method in feature space yields global structure to characterized an object. Besides global structure, local structure is also important. Kernel Linear Preserving Projection as known as KLPP is method used to preserve the intrinsic geometry of the data and local structure in feature space [Cai et al., 2005; Cai et al.,, 2006; Kokiopoulou, 2004; Mauridhi et al., 2010]. The objective of LPP in feature space is written in the following equation [Mauridhi et al., 2010]

$$\min \sum_{ij} (\phi(y)_i - \phi(y)_j)^2 . \phi(S_{ij}) \tag{24}$$

In this case the value of $S_{i,j}$ can be defined as

$$\phi(S_{ij}) = \begin{cases} e^{\frac{(\phi(x_i)-\phi(x_j))^2}{t}} & ||\phi(x_i)-\phi(x_j)|| < \varepsilon \\ 0 & otherwise \end{cases} \tag{25}$$

Where ε >0, but it is sufficiently small compared to the local neighborhood radius. Minimizing the objective function ensures the closeness between points that is located in the same class. If neighboring points of $\phi(x_i)$ and $\phi(x_j)$ are mapped far apart in feature space and if $(\phi(y_i) - \phi(y_j))$ is large, then $\phi(S_{ij})$ incurs a heavy penalty in feature space. Suppose a set of data and a weighted graph $G = (V, E)$ is constructed from data points where the data points that are closed to linked by the edge. Suppose maps of a graph to a line is chosen to minimize the objective function of KLPP in Equation (24) on the limits (constraints) as appropriate. Suppose a represents transformation vector, whereas the i^{th} column vector of X is symbolized by using x_i. By simple algebra formulation step, the objective function in feature space can be reduced in the following equation [Mauridhi et al., 2010]

$$
\begin{aligned}
&\frac{1}{2}\sum_{ij}(\phi(y_i)-\phi(y_j))^2\phi(S)_{ij}\\
&=\frac{1}{2}\sum_{ij}(\phi(a^T)\phi(x_i)-\phi(a^T)\phi(x_j))^2\phi(S_{ij})\\
&=\frac{1}{2}\sum_{ij}((\phi(a^T)\phi(x_i))^2-2\phi(a^T)\phi(x_i)\phi(a^T)\phi(x_j)+(\phi(a^T)\phi(x_j))^2)\phi(S_{ij})\\
&=\frac{1}{2}\sum_{ij}(\phi(a^T)\phi(x_i)\phi(x_i{}^T)\phi(a)-2\phi(a^T)\phi(x_i)\phi(x_j{}^T)\phi(a)+\phi(a^T)\phi(x_j)\phi(x_j{}^T)\phi(a))\phi(S_{ij})\\
&=\frac{1}{2}\sum_{ij}(2\phi(a^T)\phi(x_i)\phi(x_i{}^T)\phi(a)-2\phi(a^T)\phi(x_i)\phi(x_j{}^T)\phi(a))\phi(S_{ij})\\
&=\sum_{ij}\phi(a^T)\phi(x_i)\phi(S_{ij})\phi(x_i^T)\phi(a)-\sum_{ij}\phi(a^T)\phi(x_i)\phi(S_{ij})\phi(x_j^T)\phi(a)\\
&=\sum_{i}\phi(a^T)\phi(x_i)\phi(D_{ii})\phi(x_i^T)\phi(a)-\phi(a^T)\phi(X)\phi(S)\phi(X^T)\phi(a)\\
&=\phi(a^T)\phi(X)\phi(D)\phi(X^T)\phi(a)-\phi(a^T)\phi(X)\phi(S)\phi(X^T)\phi(a)\\
&=\phi(a^T)\phi(X)(\phi(D)-\phi(S))\phi(X^T)\phi(a)\\
&=\phi(a^T)\phi(X)\phi(L)\phi(X^T)\phi(a)
\end{aligned}
\tag{26}
$$

In this case, $\phi(X)=[\ \phi(x_1),\ \phi(x_2),\ .\ .\ .\ .\ .\ \phi(x_M)]$, $\phi(D_{ii}) = \phi(\Sigma_j\ S_{ij})$ and $\phi(L) = \phi(D) - \phi(S)$ represent *Laplacian* matrices in feature space known as *Laplacianlips*, when these are implemented in smiling stage classification. The minimum of the objective function in feature space is given by the minimum eigenvalue solution in feature space by using the following equation

$$
\begin{aligned}
\phi(X)(\phi(D)-\phi(S))\phi(X^T)\phi(w)&=\phi(\lambda)\phi(x)\phi(D)\phi(x^T)\\
\phi(X)\phi(L)\phi(X^T)\phi(w)&=\phi(\lambda)\phi(x)\phi(D)\phi(x^T)
\end{aligned}
\tag{27}
$$

Eigenvalues and eigenvectors in feature space can be calculated by using Equation (27). The most until the less dominant features can be achieved by sorting eigenvalues decreasingly and followed by sorting corresponding eigenvectors in feature space.

7. Experimental results of smile stage classification based on the maximum value selection of kernel linear preserving projection

To evaluate **the Maximum Value Selection of Kernel Linear Preserving Projection Method**, it is necessary to conduct the experiment. In this case, 30 persons were used as experiment. Each person consists of 3 patterns, which are smiling pattern I, III and IV, while smiling pattern II is not used. The image size was 640x 640 pixels and every face image was changed the size into 50x50 pixels (Figure 14). Before feature extraction process, face image had been manually cropped against a face data at oral area to produce spatial coordinate [5.90816 34.0714 39.3877 15.1020] [Mauridhi et al., 2010]. This was conducted to simplify calculation process. In this case, cropped data were used for both training and testing set. This process caused the face data size reduction into 40x16 pixels as seen in Figure 15.

Fig. 14. Original Sample of Smiling Pattern

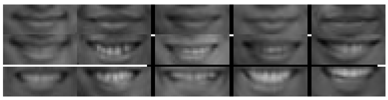

Fig. 15. Cropping Result Sample of Smiling Pattern

Experiments were applied by using 3 scenarios. In the first scenario, the first of 2/3 data (20 of 30 persons) became the training set and the rest (10 persons) were used as testing set. In the second scenario, the first of 10 persons (10 of 30) were used as testing set and the last 20 (20 of 30) persons were used as training set. In the last scenario, the first and the last of 10 persons (20 of 30) were used as training set and the middle 10 persons (10 of 30) were used for testing set. It means, data were being rotated without overlap, thus each of them has got the experience of becoming testing data. Due to smiling pattern III, the numbers of training and testing data were 20*3=60 and 10*3=30 images respectively. In this experiment, 60 dimensions were used. To measure similarity, the angular separation and Canberra were used. The Equation (17) and (18) are similarity measure for the angular separation and Canberra [Mauridhi et al., 2010]. To achieve classification rate percentage, equation (19) was used. The result of classification using the 1st, 2nd, and 3rd scenario can be seen in Figure 16, 17, and 18 respectively [Mauridhi et al., 2010].

The 1st, 2nd, and 3rd scenario had similarity trend as seen in Figure 16, 17, and 18. Recognition rate increased significantly from the 1st until 10th dimension, whereas recognition rate using more than 11 dimensions slightly fluctuated. The maximum and the average recognition rate in the 1st scenario were not different, which was 93.33%. In the 2nd scenario, the maximum recognition rate was 90%, when Canberra similarity measure was used. In the 3rd scenario, the maximum recognition rate was 100%, when angular separation was used. The maximum recognition rate was 93.33%, for both Angular Separation and Canberra Similarity Measure [Mauridhi et al., 2010] as seen in Table 6.

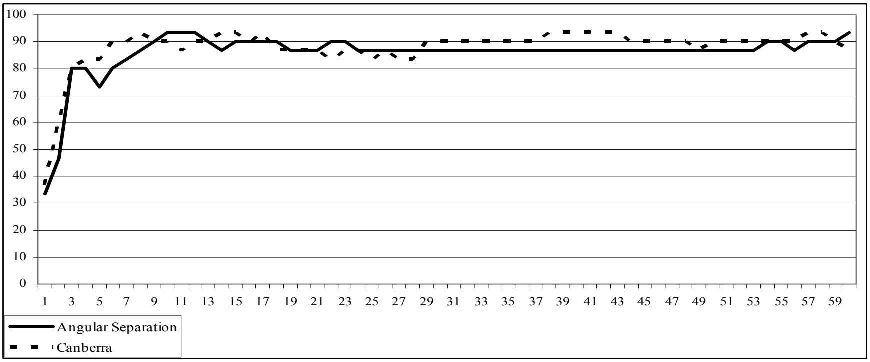

Fig. 16. Smile Stage Classification Recognition Rate Based on the Maximum Value Selection of Kernel Linear Preserving Projection Method Using 1st Scenario

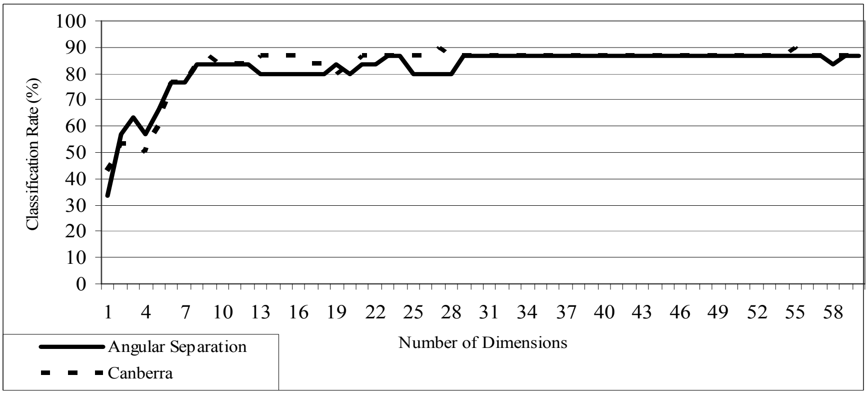

Fig. 17. Smile Stage Classification Recognition Rate Based on the Maximum Value Selection of Kernel Linear Preserving Projection Method Using 2nd Scenario

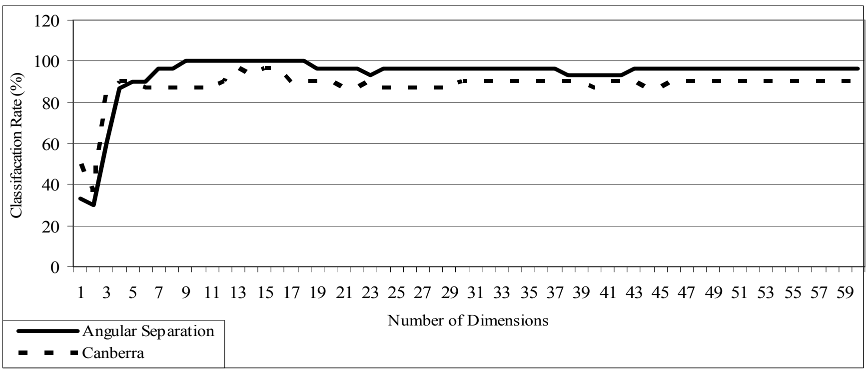

Fig. 18. Smile Stage Classification Recognition Rate Based on the Maximum Value Selection of Kernel Linear Preserving Projection Method Using 3rd Scenario

Similarity Methods	The Maximum Recognition Rate in the Scenario (%)			Average
	1st	2nd	3rd	
Angular Separation	93.33	86.67	100	93.33
Canberra	93.33	90.00	96.67	93.33
Maximum	93.33	90.00	100	94.44
Average	93.33	88.34	98.34	93.33

Table 6. The Smile Stage Classification Recognition Rate using Maximum Feature Value Selection Method of Non linear Kernel Function based on Kernel Linear Preserving Projection

The experimental results of **the Maximum Value Selection of Kernel Linear Preserving Projection Method** have been compared to "Two Dimensional Principal Component Analysis (2D-PCA) and Support Vector Machine (SVM) as its classifier" [Rima et al., 2010] and have been combined with some methods, which were Principal Component Analysis (PCA)+Linear Discriminant Analysis (LDA) and SVM as its classifier [Gunawan et al., 2009] as seen Figure 19

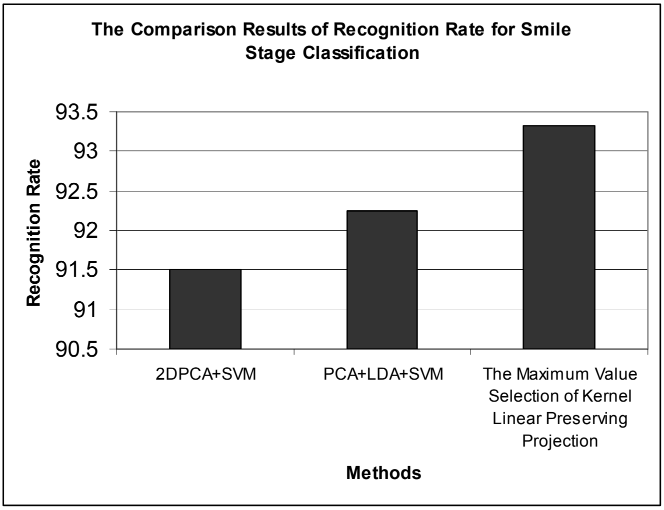

Fig. 19. The Comparison Results of Recognition Rate for Smile Stage Classification

8. Conclusion

For both the maximum non-linear feature selection of Kernel Principal Component Analysis and Kernel Linear Preserving Projection has yielded local feature structure for extraction, which is more important than global structures in feature space. It can be shown that, the maximum non-linear feature selection of Kernel Principal Component Analysis for face recognition has outperformed the PCA, LDA/QR and LPP/QR on the ORL and the YALE face databases. Whereas the maximum value selection of Kernel Linear Preserving Projection as extension of Kernel Principal Component Analysis has outperformed the 2D-PCA+SVM and the PCA+LDA+SVM for smile stage classification.

9. References

A.M. Martinez and A.C. Kak, “PCA versus LDA,” IEEE Trans. Pattern Analysis Machine Intelligence, vol. 23, no. 2, pp. 228-233, Feb. 2001.

Arif Muntasa, Mochamad Hariadi, Mauridhi Hery Purnomo, "*Automatic Eigenface Selection for Face Recognition*", The 9th Seminar on Intelligent Technology and Its Applications (2008a) 29 – 34.

Arif Muntasa, Mochamad Hariadi, Mauridhi Hery Purnomo, “Maximum Feature Value Selection of Nonlinear Function Based On Kernel PCA For Face Recognition”, Proceeding of The 4th Conference On Information & Communication Technology and Systems, Surabaya, Indonesia, 2008b

Cai, D., He, X., and Han, J. Using Graph Model for Face Analysis, University of Illinois at Urbana-Champaign and University of Chicago, 2005.

Cai, D., X. He, J. Han, and H.-J. Zhang. Orthogonal laplacianfaces for face recognition. *IEEE Transactions on Image Processing*, 15(11):3608–3614, 2006.

Gunawan Rudi Cahyono, Mochamad Hariadi, Mauridhi Hery Purnomo, “Smile Stages Classification Based On Aesthetic Dentistry Using Eigenfaces, Fisherfaces And Multiclass Svm”, 2009

J.H.P.N. Belhumeur, D. Kriegman, “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection”, IEEE Trans. on PAMI, 19(7):711–720, 1997..

Jon Shlens, *"A Tutorial On Principal Component Analysis And Singular Value Decomposition"*, http://mathworks.com , 2003

Kokiopoulou, E. and Saad, Y. Orthogonal Neighborhood Preserving Projections, University of Minnesota, Minneapolis, 2004.

M. Kirby and L. Sirovich, "Application of the KL Procedure for the Characterization of Human Faces," IEEE Trans. Pattern Anal. Mach. Intell., vol. 12, no. 1, pp. 103-108, 1990.

M. Turk, A. Pentland, “Eigenfaces for recognition”, Journal of Cognitive Science, pages 71–86, 1991.

Mauridhi Hery Purnomo, Tri Arif, Arif Muntasa, “Smiling Stage Classification Based on Kernel Laplacianlips Using Selection of Non Linier Function Maximum Value”, 2010 IEEE International Conference on Virtual Environments, Human-Computer Interfaces, and Measurement Systems (VECIMS 2010) Proceedings, pp 151-156

Mika, S., Ratsch, G.,Weston, J., Scholkopf, B. and Mller, K.R.: Fisher discriminant analysis with kernels. IEEE Workshop on Neural Networks for Signal Processing IX, (1999) 41-48

Research Center of Att, UK, *Olivetti-Att-ORL FaceDatabase*, http://www.uk.research.att.com/facedabase.html, Accessed in 2007

Rima Tri Wahyuningrum, Mauridhi Hery Purnomo, I Ketut Eddy Purnama, “Smile Stages Recognition in Orthodontic Rehabilitation Using 2D-PCA Feature Extraction”, 2010

Sch¨olkopf, B., Mika, S., Burges, C. J. C., Knirsch, P., Mller, K. R., Raetsch, G. and Smola, A.: Input Space vs. Feature Space in Kernel Based Methods, IEEE Trans. on NN, Vol 10. No. 5, (1999) 1000-1017

Sch¨olkopf, B., Smola, A.J. and Mller, K.R.: Nonlinear Component Analysis as a Kernel Eigen-value Problem, Neural Computation, 10(5), (1998) 1299-1319

X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang. Face recognition using laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):328–340, 2005.

Yale Center for Computational Vision and Control, *Yale Face Database*, http://cvc.yale.edu/projects/yalefaces/yalefaces.html, Accessed 2007

Yambor, W.S . Analysis of PCA-Based and Fisher Discriminant-Based Image Recognition Algorithms, Tesis of Master, Colorado State University, 2000

Principal Component Analysis
Edited by Dr. Parinya Sanguansat

ISBN 978-953-51-0195-6
Hard cover, 300 pages
Publisher InTech
Published online 02, March, 2012
Published in print edition March, 2012

This book is aimed at raising awareness of researchers, scientists and engineers on the benefits of Principal Component Analysis (PCA) in data analysis. In this book, the reader will find the applications of PCA in fields such as image processing, biometric, face recognition and speech processing. It also includes the core concepts and the state-of-the-art methods in data analysis and feature extraction.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Mauridhi Hery Purnomo, Diah P. Wulandari, I. Ketut Eddy Purnama and Arif Muntasa (2012). The Maximum Non-Linear Feature Selection of Kernel Based on Object Appearance, Principal Component Analysis, Dr. Parinya Sanguansat (Ed.), ISBN: 978-953-51-0195-6, InTech, Available from: http://www.intechopen.com/books/principal-component-analysis/maximum-non-linear-feature-selection-of-kernel-based-on-object-appearance

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821