SELECTING KERNEL EIGENFACES FOR FACE RECOGNITION WITH ONE TRAINING SAMPLE PER SUBJECT

Jie Wang, K.N.Plataniotis and A.N.Venetsanopolous

The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto
10 King's College Road, Toronto, M5A 3G4, ONTARIO, CANADA
{jwang,kostas,anv}@dsp.utoronto.ca

ABSTRACT

It is well-known that supervised learning techniques such as linear discriminant analysis (LDA) often suffer from the so called small sample size problem when apply to solve face recognition problems. This is due to the fact that in most cases, the number of training samples is much smaller than the dimensionality of the sample space. The problem becomes even more severe if only one training sample is available for each subject. In this paper, followed by the well-known unsupervised technique, kernel principal component analysis(KPCA), a novel feature selection scheme is proposed to establish a discriminant feature subspace in which the class separability is maximized. Extensive experiments performed on the FERET database indicate that the proposed scheme significantly boosts the recognition performance of the traditional KPCA solution.

1. INTRODUCTION

Face recognition(FR) has received more and more attentions with a wide range of applications such as access control, forensic identification and human computer interface. Although numerous FR algorithms have been proposed in the past two decades with the state-of-the-art reported in the survey of [1],it still remains as a difficult problem far from well solved. This is due to the fact that faces exhibit significant variations in appearance due to illumination, expression, pose and aging factors. At the same time, examples available for training a FR machine are usually limited. To the extreme case, when each subject only has one image sample, the recognition problem becomes even more challenging. In such a case, some well-known supervised learning techniques such as linear discriminant analysis (LDA)[2] even fail to apply since the intrapersonal information can not be obtained from one image sample per subject.

One training sample problem is a realistic problem existing in many applications such as surveillance photo identification. One possible solution to this problem is to apply an unsupervised learning technique on the given samples such as the so-called projection combined principal component analysis $((PC)^2A)$[3] and SVD perturbation method (SVD)[4] which are extensions of the well-known principal component analysis (PCA) solution [5]. The proposals introduced pre-processing schemes followed by a standard PCA. Another possible solution is to artificially generate extra samples for each subject under consideration such as moving the original image in four directions[6]. However, as stated in [7], the generated sam-

ples are usually highly correlated and should not be considered as truly independent training samples.

Different from the above mentioned solutions which only use the given samples to train FR machines, in [8], based on the traditional eigenface (PCA) solution, we proposed a feature selection scheme in a generic learning (GL) framework. Within the GL framework, a PCA machine is built by using a generic database which contains the subjects other than those to be recognized in real operations. This is based on the assumption that human faces exhibit similar intrapersonal variations so that the discriminant information among the specific subjects could be learnt from other subjects. Since PCA is an unsupervised learning technique without considering the class label, inter- and intra- personal variations are coupled together in the extracted PCA space. Therefore, a feature selection scheme was proposed to apply on the extracted eigenfaces. The selected eigenfaces span a feature subspace in which the class separability is maximized. In this paper, we extend the proposed method [8] into a non-linear space by using the so-called kernel machine techniques. By combining the strength of the proposed feature selection scheme in [8] and the kernel techniques[9], a novel kernel eigenface selection scheme is proposed which allows for a non-linear solution to the problem. A KPCA machine[10] is firstly trained on the generic database, and a feature selection procedure is applied on the extracted kernel eigenfaces thereafter. It will be further observed that the method proposed in [8] is a special case of the proposed here algorithm when a linear kernel function is used.

The rest of the paper is organized as follows. We start by briefly reviewing the eigenface selection scheme in section 2. Following that, the kernel eigenface selection procedure is discussed in section 3 with the selection criterion and procedure described in details. Experimentations on the FERET [11] database are presented in section 4 followed by a conclusion drawn in section 5.

2. REVIEW OF EIGENFACE SELECTION

In this section, the idea of selecting eigenfaces will be briefly reviewed. Let $\mathcal{G}$ be the gallery set containing G subjects to be recognized. Each subject is represented by a face image $\mathbf{g}_i, i = 1, 2, ..., G$. Let $\mathcal{Z} = \{\mathcal{Z}_i\}_{i=1}^C$ be the generic data database, containing C subjects with each subject $\mathcal{Z}_i = \{\mathbf{z}_{ij}\}_{j=1}^{C_i}$, consisting of C_i samples $\mathbf{z}_{ij}$ with a total of $N = \sum_{i=1}^C C_i$ samples, where $\mathbf{z}_{ij} \in \mathcal{R}^S$. Please note, there is no overlapping between $\mathcal{G}$ and $\mathcal{Z}$. PCA is then applied on the generic training set obtaining at most $N - 1$ meaningful eigenvectors with non zero eigenvalues when the number of training samples is smaller than the dimensionality of the sample space.

PCA is an unsupervised linear technique which produces the most expressive subspace for face representation but not necessary

be the most discriminating one. This is due to the fact that the eigenspace computed from total scatter $C = \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{C_i}(\mathbf{z}_{ij} - \bar{\mathbf{z}})(\mathbf{z}_{ij} - \bar{\mathbf{z}})^T$ includes both intrapersonal and interpersonal variations. Therefore, in order to make PCA more attractive to classification applications, we proposed a selection scheme to select a subset of eigenfaces spanning a subspace in which the interpersonal variations are maximized and intrapersonal variations are minimized[8]. Let (A_m^{opt}) be the optimal feature subset with cardinality of m which is obtained by maximizing the ratio of the trace of between- and within- class scatter. The proposed selection criterion is as follows:

$$A_m^{opt} = \arg_{A_m} \max J(A_m) \qquad (1)$$

$$J(A_m) = \frac{tr(Sb)}{tr(Sw)} = \frac{tr(\eta Sb_{Gen}(A_m) + (1-\eta)Sb_{Gal}(A_m))}{tr(Sw_{Gen}(A_m))} \qquad (2)$$

where A_m is any feature subset with cardinality of m. Sb_{Gen} and Sb_{Gal} are the between class scatters calculated from generic samples and gallery samples while Sw_{Gen} is the within class scatter calculated from generic samples. It can be observed that Sb is estimated by using both the generic samples (Sb_{Gen}) and the gallery samples (Sb_{Gal}). If only gallery samples are used ($\eta = 0$), due to the limit sample size for each gallery subject, the estimation of Sb is unreliable, giving rise to high variance[12]. On the contrary, if generic samples of other subjects are included to reduce the estimation variance ($\eta = 1$), the estimation becomes biased from the optimal one which aims at the discrimination of the gallery subjects only. Therefore, a regularization parameter η is applied to balance the estimation bias and variance.

With the proposed selection criterion, a so-called Forward Selection procedure[13] is applied to select the optimal m combinations to optimize the criterion J.

3. SELECTING KERNEL EIGENFACES

3.1. Kernel Feature Space

Compared to linear solutions, non-linear algorithms provide better ways to solve the recognition problem when the sample distribution is no longer Gaussian or convex. Kernel machine technique is the key component in the available nonlinear FR algorithms. The basic idea behind the kernel machine is to find a nonlinear transform from the original sample space ($\mathcal{R}^S$) to a so-called kernel feature space $\mathcal{F}$ by using a nonlinear function $\phi(\cdot)$,

$$\phi : \mathbf{x} \in \mathcal{R}^S \;\; \to \;\; \phi(\mathbf{x}) \in \mathcal{F} \qquad (3)$$

It is expected that in the feature space $\mathcal{F}$, the sample distribution becomes Gaussian and convex so that a traditional linear methodology can be applied[9].Without determining the explicit nonlinear function $\phi(\cdot)$, the kernel machine technique provides an implicit solution to the problem. The nonlinear mapping is implemented implicitly by using a kernel function. It is revealed that the dot products in the feature space $\mathcal{F}$ can be replaced by a kernel function $k(\cdot)$ defined in the original sample space as long as the kernel function satisfies Mercer's condition[9], i.e., for any two vectors in the sample space $\mathbf{x}_i, \mathbf{x}_j \in \mathcal{R}^S$, $k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)$. Therefore, the key task of designing a kernel-based algorithm is to select an appropriate kernel function and represent the linear procedure by using dot product forms.

3.2. Kernel Principal Component Analysis (KPCA)

KPCA [10] is actually a standard PCA solution performed in $\mathcal{F}$. For any input pattern $\mathbf{x}$, its feature representation in the KPCA feature subspace $\varphi_{KPCA}(\mathbf{x})$ can be obtained by using the dot product computed implicitly through the kernel function:

$$\varphi_{KPCA}(\mathbf{x}) = \Gamma_{KPCA}^T K_{\mathbf{x}}, \;\; K_{\mathbf{x}} = (k(\mathbf{z}_{ij}, \mathbf{x}))_{i=1,...,C}^{j=1,...,C_i} \qquad (4)$$

where $\Gamma_{KPCA} = [\gamma_1, ..., \gamma_F]$ consists of the first F significant eigenvectors of the gram matrix $\mathbf{K}$ ($N \times N$) corresponding to F largest eigenvalues, where $\mathbf{K} = (k(\mathbf{z}_{im}, \mathbf{z}_{jn}))_{i,j=1,...,C}^{m=1,...,C_i;n=1,...,C_j}$ and $K_{\mathbf{x}}$ is a $N \times 1$ vector.

3.3. Calculating Selection Criterion

Within the KPCA space, the corresponding feature selection criterion becomes

$$\Gamma_m^{opt} = \arg_{\Gamma_m} \max J(\Gamma_m) \qquad (5)$$

$$J(\Gamma_m) = \frac{tr(\tilde{Sb})}{tr(\tilde{Sw})} = \frac{tr(\eta \tilde{Sb}_{Gen}(\Gamma_m) + (1-\eta)\tilde{Sb}_{Gal}(\Gamma_m))}{tr(\tilde{Sw}_{Gen}(\Gamma_m))}$$

where Γ_m is any $m-$ dimensional KPCA feature subset and $\tilde{Sb}, \tilde{Sw}$ are the corresponding between- and within-class scatters defined in the kernel space $\mathcal{F}$. In the following, the detailed calculation of the selection criterion will be discussed. Let $\mathbf{y}_{ij} = \Gamma_m^T \mathbf{K}_{\mathbf{z}_{ij}}$ and $\mathbf{s}_i = \Gamma_m^T \mathbf{K}_{\mathbf{g}_i}$ be the feature representations of the generic sample $\mathbf{z}_{ij}$ and gallery sample $\mathbf{g}_i$ in the feature subspace spanned by Γ_m. Then,

$$
\begin{aligned}
tr(\tilde{Sb}_{Gen}) &= \frac{1}{N}\sum_{i=1}^{C} C_i(\bar{\mathbf{y}}_i - \bar{\mathbf{y}})^T(\bar{\mathbf{y}}_i - \bar{\mathbf{y}}) \\
&= \frac{1}{N}\sum_{i=1}^{C} C_i \bar{\mathbf{y}}_i^T \bar{\mathbf{y}}_i - \bar{\mathbf{y}}^T \bar{\mathbf{y}} \\
&= \frac{1}{N}\sum_{i=1}^{C} \frac{1}{C_i} \mathbf{1}_{C_i}^T \mathbf{K}_{ii}^y \mathbf{1}_{C_i} - \frac{1}{N^2}\mathbf{1}_N^T \mathbf{K}^y \mathbf{1}_N
\end{aligned} \qquad (6)
$$

$$
\begin{aligned}
tr(\tilde{Sw}_{Gen}) &= \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{C_i}(\mathbf{y}_{ij} - \bar{\mathbf{y}}_i)^T(\mathbf{y}_{ij} - \bar{\mathbf{y}}_i) \\
&= \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{C_i} \mathbf{y}_{ij}^T \mathbf{y}_{ij} - \frac{1}{N}\sum_{i=1}^{C} C_i \bar{\mathbf{y}}_i^T \bar{\mathbf{y}}_i \\
&= \frac{1}{N}tr(\mathbf{K}^y) - \frac{1}{N}\sum_{i=1}^{C} \frac{1}{C_i}\mathbf{1}_{C_i}^T \mathbf{K}_{ii}^y \mathbf{1}_{C_i}
\end{aligned} \qquad (7)
$$

$$
\begin{aligned}
tr(\tilde{Sb}_{Gal}) &= \frac{1}{G}\sum_{i=1}^{G}(\mathbf{s}_i - \bar{\mathbf{s}})^T(\mathbf{s}_i - \bar{\mathbf{s}}) \\
&= \frac{1}{G}\sum_{i=1}^{G} \mathbf{s}_i^T \mathbf{s}_i - \bar{\mathbf{s}}^T \bar{\mathbf{s}} \\
&= \frac{1}{G}tr(\mathbf{K}^s) - \frac{1}{G^2}\mathbf{1}_G^T \mathbf{K}^s \mathbf{1}_G
\end{aligned} \qquad (8)
$$

where $\bar{\mathbf{y}}_i = \frac{1}{C_i}\sum_{j=1}^{C_i} \mathbf{y}_{ij}, \bar{\mathbf{y}} = \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{C_i}\mathbf{y}_{ij}, \bar{\mathbf{s}} = \frac{1}{G}\sum_{i=1}^{G}\mathbf{s}_i$ and $\mathbf{1}_m$ is a $m \times 1$ vector with all elements equal to 1. $\mathbf{K}^y(N \times N), \mathbf{K}_{ii}^y(C_i \times C_i)$ and $\mathbf{K}^s(G \times G)$ are the gram matrices of generic samples and gallery samples in the KPCA subspace spanned by Γ_m, which are defined as follows:

$$
\begin{aligned}
\mathbf{K}^y &= (\mathbf{y}_{ip}^T \mathbf{y}_{jq})_{i,j=1,...,C}^{p=1,...,C_i;q=1,...,C_j} \quad \mathbf{K}_{ii}^y = (\mathbf{y}_{ip}^T \mathbf{y}_{iq})^{p,q=1,...,C_i} \\
\mathbf{K}^s &= (\mathbf{s}_i^T \mathbf{s}_j)_{i,j=1,...,G}
\end{aligned} \qquad (9)
$$

It can be further observed that $\mathbf{K}^y$ and $\mathbf{K}^s$ can be obtained from the gram matrix defined in the original sample space with the predefined kernel function, i.e.,

$$
\begin{aligned}
\mathbf{y}_{ip}^T \mathbf{y}_{jq} &= \mathbf{K}_{\mathbf{z}_{ip}}^T \Gamma_m \Gamma_m^T \mathbf{K}_{\mathbf{z}_{jq}} \\
\mathbf{K}^y &= \mathbf{K}\Gamma_m \Gamma_m^T \mathbf{K}^T
\end{aligned} \qquad (10)
$$

$$
\begin{aligned}
\mathbf{s}_i^T \mathbf{s}_j &= \mathbf{K}_{\mathbf{g}_i}^T \Gamma_m \Gamma_m^T \mathbf{K}_{\mathbf{g}_j} \\
\mathbf{K}^s &= \mathbf{K}_G \Gamma_m \Gamma_m^T (\mathbf{K}_G)^T \\
\mathbf{K}_G &= (k(\mathbf{g}_i, \mathbf{z}_{mn}))_{m=1,...,C;n=1,...,C_m}^{i=1,...,G}
\end{aligned} \qquad (11)
$$

$\mathbf{K}_{ii}^y$ is the submatrix of $\mathbf{K}^y$ corresponding to the samples belonging to ith subject, $\mathbf{K}_G$ is a $G \times N$ gram matrix, $\mathbf{K}_{\mathbf{z}_{ip}}$ and $\mathbf{K}_{\mathbf{g}_i}$ are

$N \times 1$ vectors such that $\mathbf{K}_{\mathbf{z}_{ip}} = (k(\mathbf{z}_{mn}, \mathbf{z}_{ip}))_{m=1,...,C}^{n=1,...,C_m}$, $\mathbf{K}_{\mathbf{g}_i} = (k(\mathbf{z}_{mn}, \mathbf{g}_i))_{m=1,...,C}^{n=1,...,C_m}$.

With the calculations of the selection criterion described above, Forward Selection procedure[13] is also applied to select the optimal m combinations in Γ to maximize J defined in Eq.(5). Similar to [8], other than selecting from all $N-1$ obtained kernel eigenfaces, the complete kernel eigenfaces set Γ contains the first M kernel eigenfaces corresponding to the largest eigenvalues, i.e., $\Gamma = [\gamma_1, ..., \gamma_M]$. Thus the optimal m kernel eigenfaces Γ_m are selected from Γ. The reason for exclusion of trailing kernel eigenfaces is due to the fact that the kernel eigenfaces corresponding to small eigenvalues are usually unreliable since only small number of samples are used to estimate the sample covariance in the high dimensional kernel space $\mathcal{F}$. In addition, in order to further keep the most discriminating information, we start with the most significant feature γ_1 with the largest eigenvalue.The detailed selection procedure is listed as follows

$$
\begin{aligned}
&(1) \quad \Gamma_m(1) = [\gamma_1] \\
&(2) \quad for \quad k = 1 \ to \ m \\
&\qquad\quad for \quad \gamma_i \in \Gamma - \Gamma_m(k) \\
&\qquad\qquad\quad with \ Equation.5, \quad calculate \\
&\qquad\qquad\quad C(\gamma_i) = J(\Gamma_m(k) \oplus \gamma_i) \\
&\qquad\quad end \\
&\qquad\quad \Gamma_m(k+1) = \Gamma_m(k) \oplus arg\max_{\gamma_i} C(\gamma_i) \\
&\quad end
\end{aligned}
$$

It can be observed, if linear kernel function is used, i.e., $k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$, KPCA is reduced to the traditional PCA solution and the described above method is reduced to the eigenface selection scheme proposed in [8]. Therefore the proposed here solution is a generalized algorithm which includes [8] as a special case.

4. EXPERIMENT

4.1. Experiment Setup

In order to demonstrate the effectiveness of the proposed algorithm, the well-known FERET database is used for experimentations. We use the same experiment setup as the one proposed in [8]. Among all 1200 subjects included in the FERET database, a subset of 226 subjects with 3 images per subject is used as a generic training set. In addition, 1703 images of 256 subjects with at least 4 images per subject are used to form the gallery and probe sets. There is no overlapping of the subjects in the generic training set and gallery/probe set. We randomly select 256 frontal images (one per subject) to form the gallery set. The remaining 1447 images are considered to be the probe set. We further partition the probe set into three subsets. Set $\mathcal{P}_1$ contains 914 frontal images of 256 subjects. The camera time difference between $\mathcal{P}_1$ probe images and their corresponding gallery matches is less than half year($\leq 180 days$), mostly taken in the same session. Set $\mathcal{P}_2$ consists of 226 frontal images of 75 subjects whose camera time difference to their corresponding gallery matches is greater than one and half year ($\geq 540 days$). Set $\mathcal{P}_3$ contains 227 non frontal images of 48 subjects with no particular consideration with respect to camera time. It can be observed that set $\mathcal{P}_1$ contains small intrapersonal variations. However, $\mathcal{P}_2$ and $\mathcal{P}_3$ includes larger intrapersonal variations due to aging and pose factors which is often encountered in real applications.

All images are preprocessed according to the recommendation of the FERET protocol, which includes (1) images are rotated and scaled so that the centers of the eyes are placed on specific pixels and the image size is 150×130; (2) a standard mask is applied to remove nonface portions; (3) histogram equalized and image normalized to

have zero mean and unit standard deviation. Then each image is finally represented as a vector of length 17154.

KPCA is firstly applied on the generic training set obtaining 677 meaningful kernel eigenfaces. The Gaussian kernel is selected as the kernel function with $\sigma^2 = 5 \times 10^6$, i.e., $k(\mathbf{x}, \mathbf{y}) = \exp(-(\mathbf{x} - \mathbf{y})^T(\mathbf{x} - \mathbf{y})/\sigma^2)$. The first 246 KPCA features are kept to form the complete kernel eigenface set Γ which captures 95% of the energy. The cardinality of the selected feature subset Γ_m is up to $m = 100$. The recognition is performed by comparing the Euclidean distance between the probe and each gallery image in the feature space spanned by Γ_m, i.e., $d(\mathbf{p}, \mathbf{g}_i) = ||\Gamma_m^T(\mathbf{K}_\mathbf{p} - \mathbf{K}_{\mathbf{g}_i})||$ A probe is in the top H if the distance to its corresponding gallery match is among the H smallest distances for the gallery. Thus the recognition rate at rank H is the number of probe images in the top H divided by probe size.

For comparison purposes, the traditional KPCA solution, the $(PC)^2A$ solution proposed by [3] ($(PC)^2A$) and SVD solution proposed by [4] (SVD) are also performed.

4.2. Results and Analysis

The best found the correct recognition rate(BstCRR) at rank 1 and 20 are listed in table.1 for our proposed method (FS), KPCA, $(PC)^2A$ and SVD. It is well-known that CRR is a function of the feature number and the best found CRR is the one with the peak value corresponding to the optimal feature number ($Nopt$) which is obtained by exhaustively searching all possible feature numbers. Bold case indicates the best performance. In addition, Fig.1 depicts the corresponding recognition rate at rank 20 of these 4 methods with various number of features.

It can be observed from table.1 that no obvious improvement of BstCRR can be observed from the proposed method in $\mathcal{P}_1$. However, in $\mathcal{P}_2$ and $\mathcal{P}_3$, the BstCRR of the proposed method at rank 20 (74.30% and 71.81%) are the highest among all four algorithms while all these methods fail in rank 1 evaluation ($< 50\%$). The similar result can be observed from fig.1. The rank 20 recognition rate of the proposed method (FS) is similar to that of the traditional KPCA solution in $\mathcal{P}_1$ which is slightly better than that of $(PC)^2A$ and SVD solutions. When evaluated in $\mathcal{P}_2$, the proposed solution outperforms significantly other three methods. While in $\mathcal{P}_3$, the proposed solution also demonstrates the best performance. These observations can be explained as follows. In $\mathcal{P}_1$, probe images are quite similar to their corresponding gallery templates, i.e., the intrapersonal variations are considerably small. Thus interpersonal variations dominate recognition performance. In such a case, the traditional KPCA solution provides a better performance. This is due to the fact that the traditional KPCA selects the features with large eigenvalues, i.e., the large total variations. Therefore, when actual intraperosnal variations are small, it is equivalent to select a feature subspace which contains large interpersonal variations. However, in $\mathcal{P}_2$ and $\mathcal{P}_3$, intrapersonal variations are much larger, throwing more effect on the recognition performance. Therefore, the proposed here solution which selects the features maximizing the ratio of inter- and intra-personal variations demonstrates a better performance.

Table.2 lists the BstCRR at rank 20 with various η. It can be observed that a larger η value is preferred in $\mathcal{P}_2$ and $\mathcal{P}_3$. This indicates that under the one training sample scenario while intrapersonal variations are considerably large, high estimation variance resulting from the using of gallery samples only is a more serious problem. Therefore, a generic learning framework by including unrelated generic face samples is an appropriate choice. In addition, the best performance is achieved between two extreme cases when $\eta = 0$

Table 1. Best found CRR (%) on $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{P}_3$ at rank 1 and 20 with optimal feature number N_{opt}; For FS $\eta = 0.6$.

	RK($\mathcal{P}_1$)		RK($\mathcal{P}_2$)		RK($\mathcal{P}_3$)	
	1	20	1	20	1	20
FS	72.65	91.90	23.46	**74.30**	19.38	**71.81**
N_{opt}	86	39	91	92	99	96
KPCA	**75.05**	**93.87**	17.88	62.01	21.15	67.84
N_{opt}	95	92	84	75	79	94
PC^2A	65.10	89.06	13.14	58.10	17.62	58.59
N_{opt}	82	44	91	25	33	13
SVD	74.18	91.90	**24.58**	66.48	**25.99**	69.60
N_{opt}	85	63	97	28	74	31

Table 2. Best found CRR (%) of the proposed method (FS) on $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{P}_3$ at rank 20 with different η.

	η					
	0	0.2	0.4	0.6	0.8	1.0
$\mathcal{P}_1$	**92.77**	92.66	91.68	91.90	92.01	92.12
$\mathcal{P}_2$	58.10	63.68	73.74	**74.30**	73.74	73.74
$\mathcal{P}_3$	66.96	67.84	**72.68**	71.80	71.80	70.92

perimentation results indicate that the proposed algorithm outperforms the traditional kernel principal component analysis solution and other state-of-the-art FR algorithms.

and $\eta = 1$. This coincides our claim that balancing bias ($\eta = 1$) and estimation variance ($\eta = 0$) is necessary to further improve the recognition performance.
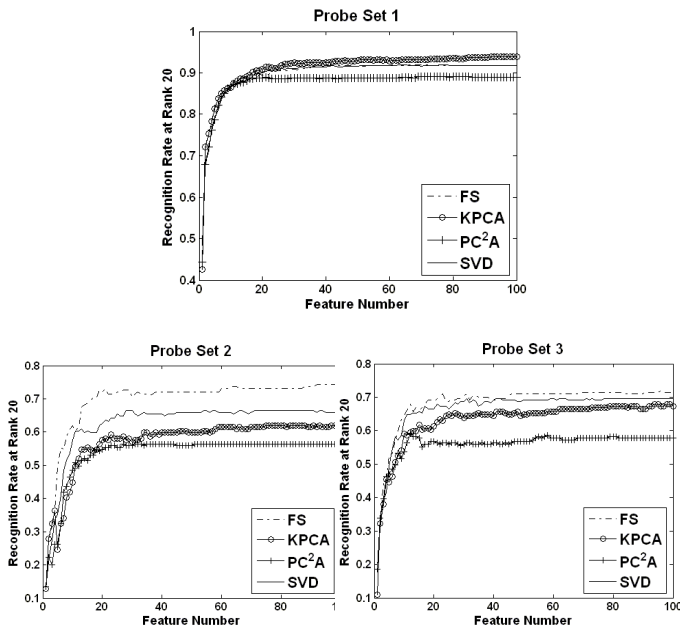

Fig. 1. Correct Recognition Rate at Rank 20 v.s. Feature Number Evaluated on Probe set 1,2 and 3; For FS, $\eta = 0.6$

5. CONCLUSION

In this paper, a novel feature selection scheme is proposed to select kernel eigenfaces. The proposal calls for the establishment of a discriminant feature subspace spanned by the selected kernel eigenfaces in which the interpersonal variations are maximized and intrapersonal variations are minimized. Considering the one training sample per subject situation, the feature selection scheme is performed within the generic learning framework. The interpersonal variations are estimated from both the generic samples and the given gallery samples so that the estimation bias and variance are balanced. Ex-

6. REFERENCES

[1] W.Zhao, R.Chellappa and P.Phillips,"Face recognition: A literature survey (2000)",*UMD CfAR Technical Report CAR-TR-948*,2000.

[2] A.M. Martez and A.C.Kak,"PCA versus LDA", *IEEE Transactions on Pattern Analysis and Machine Intelligence*,vol.23, no.2,pp.228-233,2001.

[3] S.C.Chen, D.Q.Zhang and Z.Zhou,"Enhanced $(PC)^2A$ for face recognition with one training image per person",*Pattern Recognition Letters*,vol.25,no.10,pp.1173-1181,2004.

[4] D.Zhang, S.Chen and Z.Zhou,"A new face recognition method based on SVD perturbation for single example image per person",*Applied Mathematics and Computation*,vol.163,pp.895-907,2005.

[5] M.A.Turk and A.Pentland, "Eigenfaces for recognition", *Journal of Cognitive Neuroscience*, vol.3, no.1, pp.71-86, 1991.

[6] J. Huang, P. C. Yuen, W. S. Chen, and J. H. Lai, "Component-based LDA method for face reocgnition with one training sample", *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*,pp.120-126,2003.

[7] A.M.Martinez, "Recognize Imprecisely Localized, Partially Occluded and Expression Variant Face from a Single Sample per Class", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(6) (2002) 748-763.

[8] J.Wang, K.N.Plataniotis and A.N.Venetsanopoulos,"Selecting Discriminant Eigenfaces for Face Recognition",*Pattern Recognition Letters*,vol.26, no.(10), pp.1470-1482,2005.

[9] K.R.Muller, S.Mika, G.Ratsch, K.Tsuda and B.Scholkopf, "An introduction to Kernel-based learning algorithms",*IEEE Transactions on Neural Networks*,vol.12,no.2,pp 181-200,2001.

[10] B.Scholkopf,A.Smola and K.R.Muller,"Nonlinear component analysisi as a kernel eigenvalue problem",*Neural Computation* vol.10, pp.1299-1319,1999.

[11] P.J.Phillips, H.Moon, S.A.Rizvi and P.Rauss,"The FERET evaluation method for face recognition algorithms",*IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.22,no.10,pp.1090-1104,2000.

[12] R.O.Duda, P.E.Hart and D.G.Stork, "Pattern Classification", *Second ed. John Wiley and Sons*, New York (2000).

[13] S.Kumar,"Modular learning through output space decomposition", *Ph.D dissertation, University of Texas at Austin*, 2000.