Machine Vision and Applications (2009) 20:35–46
DOI 10.1007/s00138-007-0103-1

ORIGINAL PAPER

Complete discriminant evaluation and feature extraction in kernel space for face recognition

Xudong Jiang · Bappaditya Mandal · Alex Kot

Received: 20 December 2006 / Accepted: 21 July 2007 / Published online: 15 November 2007

Abstract This work proposes a method to decompose the kernel within-class eigenspace into two subspaces: a reliable subspace spanned mainly by the facial variation and an unreliable subspace due to limited number of training samples. A weighting function is proposed to circumvent undue scaling of eigenvectors corresponding to the unreliable small and zero eigenvalues. Eigenfeatures are then extracted by the discriminant evaluation in the whole kernel space. These efforts facilitate a discriminative and stable low-dimensional feature representation of the face image. Experimental results on FERET, ORL and GT databases show that our approach consistently outperforms other kernel based face recognition methods.

Keywords Face recognition · Kernel discriminant analysis · Feature extraction · Subspace methods

1 Introduction

Face recognition has drawn considerable attention in biometric society because of its property of non-intrusiveness, which is recognizing a person from a distance. Numerous linear and nonlinear subspace based face recognition methods are proposed in the last two decades [25,30,35]. The most popular nonlinear methods are kernel principal component analysis (KPCA) [29] and kernel Fisher discriminant analysis (KFDA) [22]. These kernel based methods and their variations encode pattern information based on higher order dependencies and are tactful to the dependencies among pixels in the samples. The kernel mappings can capture the nonlinearities and complex relationships among the input data that exist due to the expression, illumination and pose variations.

The basic idea of these kernel methods is to apply nonlinear mapping $\Phi : X \in \mathbb{R}^n \rightarrow \Phi(X) \in \mathbb{H}$ in the image space $\mathbb{R}^n$, followed by linear subspace methods like PCA and FDA in the mapped feature space $\mathbb{H}$. Examples include KPCA [29] and KFDA [22,23]. Since the feature space $\mathbb{H}$ can be very high or possibly infinite dimensional and the orthogonality needs to be characterized in such a space, it is reasonable to view $\mathbb{H}$ as a Hilbert space. It is difficult to compute the dot products in the high dimensional feature space $\mathbb{H}$. Instead of mapping the data explicitly, the feature space can be computed by using the kernel trick, in which the inner products $\langle\Phi(X_{ij}), \Phi(X_{st})\rangle$ in $\mathbb{H}$ can be replaced with a kernel function $K(X_{ij}, X_{st})$, where $K(X_{ij}, X_{st}) = \langle\Phi(X_{ij}), \Phi(X_{st})\rangle$ and X_{ij}, X_{st} are sample vectors in the image space $\mathbb{R}^n$. So, the nonlinear mapping Φ can be performed implicitly in image space $\mathbb{R}^n$ [28,31]. Numerous studies [8,30,34] demonstrate that these kernel based approaches are effective in some real-world applications. However, the basic subspace analysis has still outstanding challenging problems when applied to the face recognition due to the high dimensionality of the face image and the finite number of training samples in practice.

Most of the kernel subspace based face recognition methods perform dimensionality reduction or discard a subspace before the discriminant evaluation. A popular method called kernel Fisherface [34] applies PCA first for dimensionality reduction so as to make the within-class scatter matrix

X. Jiang (✉) · B. Mandal · A. Kot
School of Electrical and Electronic Engineering,
Nanyang Technological University,
Singapore 639798, Singapore
e-mail: exdjiang@ntu.edu.sg

B. Mandal
e-mail: bapp0001@ntu.edu.sg

A. Kot
e-mail: eackot@ntu.edu.sg

nonsingular before the application of LDA. However, applying PCA for dimensionality reduction may lose important discriminative information [3,4,12,32]. The null space approach, NKDA [17] eliminates the principal subspace and extracts eigenfeatures only from the eigenvectors corresponding to the zero eigenvalues. Therefore, NKDA assumes that the null space contains the most discriminative information which is contradictory to KFDA.

Kernel Direct-LDA (KDDA) method [18] first removes the null space of the between-class scatter matrix and then extracts the eigenvectors corresponding to the smallest eigenvalues of the within-class scatter matrix. It is an open question of how to scale the extracted features as the smallest eigenvalues are very sensitive to noise. A common problem of KFDA, NKDA and KDDA approaches is that they all lose some discriminative information, either in the principal or in the null space because they perform the discriminant evaluation in a subspace.

In fact, the discriminative information resides in both subspaces. Recently, Yang et al. [33] proposed a complete kernel Fisher discriminant framework (CKFD), where features extracted from the two complementary subspaces are combined by a summed distance measures in the recognition phase [33]. Dual-space based LDA approach is proposed in [32], where features are scaled in the complementary subspace by an average eigenvalue of the within-class scatter matrix over this subspace. As eigenvalues in this subspace are not well estimated, their average may not be a good scaling factor relative to those in the principal subspace. Open questions of these approaches are how to divide the space into the principal and the complementary subspaces and how to apportion a given number of features to the two subspaces. Furthermore, as the discriminative information resides in the both subspaces, it is inefficient and only suboptimal to extract features separately from the two subspaces.

In this paper, we propose a method which utilizes the ratios of the successive eigenvalues of the eigenspectrum (shown in Fig. 1) to decompose the kernel within-class eigenspace into two subspaces: a reliable subspace spanned mainly by the facial variation and an unreliable subspace due to limited number of training samples. A weighting function (shown in Fig. 2) is proposed which circumvents undue scaling of projection vectors corresponding to the unreliable small and zero eigenvalues. Finally, features are extracted based on the discriminant evaluation in the whole kernel eigenspace. In the next section, we first study the behavior of the unreliable small eigenvalues of within-class variation matrix, then propose a methodology to decompose the eigenspace into principal and unreliable subspaces. Eigenfeature scaling and extraction are presented in Sect. 3. Experimental results and discussions are presented in Sect. 4 before drawing conclusions in Sect. 5.

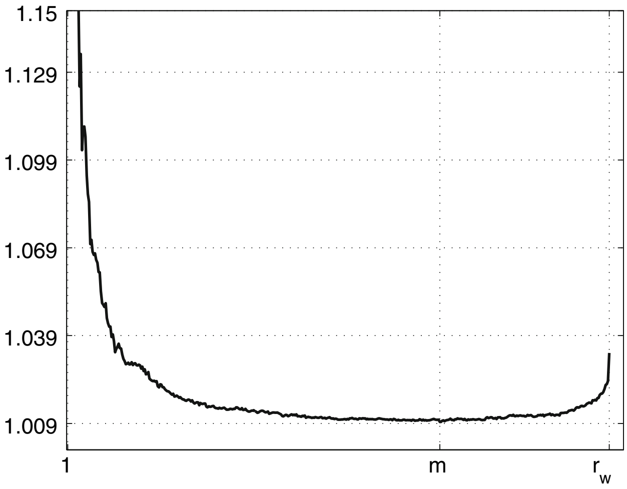

Fig. 1 Eigenratiospectrum (13) from a typical real kernel eigenspectrum of $\mathbf{S}_{\Phi}^{w}$

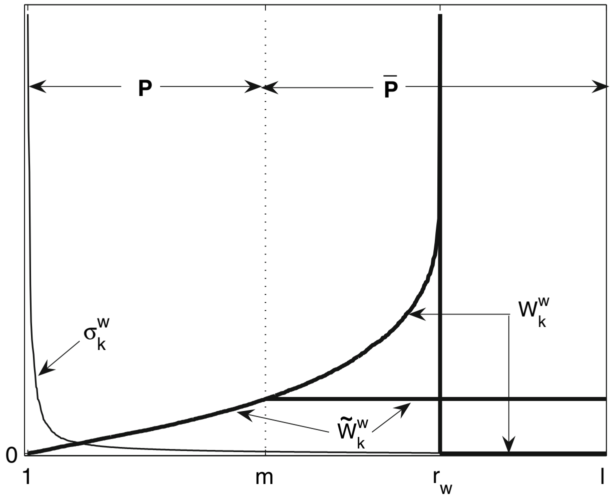

Fig. 2 Weighting functions of (12) and (16) in the principal- and unreliable-subspaces based on a typical real kernel eigenspectrum

2 Kernel feature scaling and subspace decomposition

2.1 Overview of kernel discriminant analysis

For a nonlinear mapping Φ, the image data space $\mathbb{R}^n$ can be mapped into the feature space $\mathbb{H}$

$$\Phi : X \in \mathbb{R}^n \rightarrow \Phi(X) \in \mathbb{H}. \tag{1}$$

Consequently, a pattern in the original image space $\mathbb{R}^n$ is mapped into a potentially much higher dimensional feature vector in the feature space $\mathbb{H}$. Given a set of properly aligned and normalized h-by-w face images, we can form a training set of column vectors $\{X_{ij}\}$, where $X_{ij} \in \mathbb{R}^{n=hw}$ is called image vector, by lexicographic ordering the pixel elements of image j of person i. Let the training set contain p persons and q_i sample images for person i. The number of total training sample is $l = \sum_{i=1}^{p} q_i$. The within-class scatter matrix is defined by

$$\mathbf{S}^w = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{q_i}\sum_{j=1}^{q_i}(\Phi(X_{ij}) - \overline{\Phi(X_i)})(\Phi(X_{ij}) - \overline{\Phi(X_i)})^T, \tag{2}$$

where $\overline{\Phi(X_i)} = \frac{1}{q_i}\sum_{j=1}^{q_i}\Phi(X_{ij})$. The between-class scatter matrix $\mathbf{S}^b$ is defined by

$$\mathbf{S}^b = \frac{1}{p}\sum_{i=1}^{p}(\overline{\Phi(X_i)} - \overline{\Phi(X)})(\overline{\Phi(X_i)} - \overline{\Phi(X)})^T, \tag{3}$$

where $\overline{\Phi(X)} = \frac{1}{p}\sum_{i=1}^{p}\overline{\Phi(X_i)}$, assuming all classes have equal prior probability.

The well known Fisher objective function [7] can be written in the mapped space $\mathbb{H}$ as

$$J(\mathbf{\Omega}) = \arg\max_{\Omega}\frac{\mid \mathbf{\Omega}^T\mathbf{S}^b\mathbf{\Omega} \mid}{\mid \mathbf{\Omega}^T\mathbf{S}^w\mathbf{\Omega} \mid}. \tag{4}$$

Because any solution $\mathbf{\Omega} \in \mathbb{H}$ must lie in the span of all the samples in $\mathbb{H}$, there exist coefficients ψ_{ij}, such that

$$\mathbf{\Omega} = \sum_{i=1}^{p}\sum_{j=1}^{q_i}\psi_{ij}\Phi(X_{ij}). \tag{5}$$

Combining (4) and (5), we have [14]

$$\mathbf{\Omega}^T\mathbf{S}^w\mathbf{\Omega} = \mathbf{\Psi}^T\mathbf{S}_{\Phi}^w\mathbf{\Psi}, \tag{6}$$

$$\mathbf{\Omega}^T\mathbf{S}^b\mathbf{\Omega} = \mathbf{\Psi}^T\mathbf{S}_{\Phi}^b\mathbf{\Psi}, \tag{7}$$

where $\mathbf{\Psi} = \{\psi_{ij}\}$ and

$$\begin{cases}\mathbf{S}_{\Phi}^w = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{q_i}\sum_{j=1}^{q_i}(\zeta_{ij} - \mu_i)(\zeta_{ij} - \mu_i)^T, \\ \zeta_{ij} = (K(X_{11}, X_{ij}), K(X_{12}, X_{ij}), \ldots, K(X_{pq_p}, X_{ij}))^T,\end{cases} \tag{8}$$

$$\begin{cases}\mu_i = \frac{1}{q_i}\sum_{j=1}^{q_i}\zeta_{ij}, \\ \mathbf{S}_{\Phi}^b = \frac{1}{p(p-1)}\sum_{i=1}^{p}\sum_{j=1}^{p}(\mu_i - \mu_j)(\mu_i - \mu_j)^T.\end{cases} \tag{9}$$

So the solution of function (4) can be obtained by maximizing

$$J(\mathbf{\Psi}) = \arg\max_{\Psi}\frac{\mid \mathbf{\Psi}^T\mathbf{S}_{\Phi}^b\mathbf{\Psi} \mid}{\mid \mathbf{\Psi}^T\mathbf{S}_{\Phi}^w\mathbf{\Psi} \mid} \tag{10}$$

and the problem of kernel discriminant analysis is converted into finding the leading eigenvectors of $\mathbf{S}_{\Phi}^{w\,-1}\mathbf{S}_{\Phi}^b$ [5]. However, in practice, the inversion of $\mathbf{S}_{\Phi}^w$ is impossible as it is often singular due to the limited number of training samples.

Let $\mathbf{S}_{\Phi}^g$, $g \in \{w, b\}$ represent one of the above scatter matrices. If we regard the elements of the kernel vector or the class mean vector as features, these preliminary features will be de-correlated by solving the eigenvalue problem

$$\mathbf{\Lambda}^g = \mathbf{\Psi}^{g\,T}\mathbf{S}_{\Phi}^g\mathbf{\Psi}^g, \tag{11}$$

where $\mathbf{\Psi}^g = [\psi_1^g, \ldots, \psi_l^g]$ is the eigenvector matrix of $\mathbf{S}_{\Phi}^g$, and $\mathbf{\Lambda}^g$ is the diagonal matrix of eigenvalues $\lambda_1^g, \ldots, \lambda_l^g$ corresponding to the eigenvectors. We assume that the eigenvalues are sorted in descending order $\lambda_1^g \geq, \cdots, \geq \lambda_l^g$. The plot of eigenvalues λ_k^g against the index k is called eigenspectrum of the training data in the nonlinear plane. It plays a critical role in the subspace methods as the eigenvalues are used to scale and extract features.

2.2 Problems in feature scaling and extraction of KFDA

If we compute all the eigenvalues $\mathrm{diag}(\mathbf{\Lambda}^w) = [\lambda_1^w, \ldots, \lambda_l^w]$ and eigenvectors $\mathbf{\Psi}^w = [\psi_1^w, \ldots, \psi_l^w]$ of the l-by-l dimensional matrix $\mathbf{S}_{\Phi}^w$ using (11), the projection matrix $\bar{\mathbf{\Psi}}^w = [\psi_1^w/\sigma_1^w, \ldots, \psi_l^w/\sigma_l^w]$ is so called whitened eigenvector matrix of $\mathbf{S}_{\Phi}^w$ with $||\psi_k^w|| = 1$ and $\sigma_k^w = \sqrt{\lambda_k^w}$. This implies that if any one of the eigenvalues in (11) of these matrices is zero then the corresponding eigenvector (10) gets an infinite weighting factor. Since the scatter matrices (8) and (9) are singular [7,18,19], in practice, most of the subspace based algorithms circumvent this problem by ignoring the eigenvectors corresponding to zero eigenvalues. However, as pointed out earlier that the null space of $\mathbf{S}_{\Phi}^w$ contains indispensable discriminative information essential for improving recognition accuracy.

The above argument can be viewed as an l-dimensional pattern vector ζ_{ij} is first represented by an l-dimensional eigenfeature vector $Y_{ij} = \mathbf{\Psi}^{w\,T}\zeta_{ij}$, and then multiplied by a weighting function

$$w_k^w = \begin{cases} 1/\sqrt{\lambda_k^w}, & k \leq r_w \\ 0, & r_w < k \leq l \end{cases}, \tag{12}$$

as shown in Fig. 2, where r_w is the rank of $\mathbf{S}_{\Phi}^w$. It is apparent from (12) that the eigenvectors $\{\psi_k^w\}_{k=r_w+1}^{l}$ or the null space of $\mathbf{S}_{\Phi}^w$ are weighted by zero and thus the corresponding eigenvectors fail to contribute to the whole space discriminant evaluation, which is done in the later portion of the algorithm. This is unreasonable because features in the null space have zero within-class variances based on the training data and hence should be more heavily weighted. It seems anomalous that the weighting function increases with the decrease of the eigenvalues and then suddenly has a big drop from the maximum value to zero as shown in Fig. 2. Furthermore, weights determined by the inverse of σ_k^w is, though optimal in terms of the ML estimation, dangerous when σ_k^w is small ($m < k \leq r_w$). The small and zero eigenvalues are training-set-specific and very sensitive to different training sets [9]. Adding new samples to the training set or using different training set may easily change some zero eigenvalues to nonzero and make some very small eigenvalues several times larger. Therefore, these eigenvalues of the within-class scatter matrix are unreliable.

2.3 Eigenratiospectrum and subspace decomposition

In order to alleviate the above problem we first work on the eigenspectrum of the within-class variation matrix. It is not difficult to estimate the rank of $\mathbf{S}_{\Phi}^{w}$, which is $r_w \leq (l-p)$. The eigenvalues whose indices are close to $r_w (m < k \leq r_w)$ are very small or close to zero, their inverses give undue overemphasis to the eigenvectors corresponding to this region (or indices) as shown in Fig. 2.

Jiang et al. [10] regularize eigenvalues in the linear space by modelling the eigenspectrum using a $1/f$ function to fit the most reliable portion of the eigenspectrum. The less reliable portion of the eigenspectrum is then replaced by the model. Consistent improvement of the face recognition performance is reported in [10]. However, in the kernel space, the model of $1/f$ function cannot fit the kernel eigenspectrum well due to the nonlinear transform of face images. A very clear fact in our experiments is that the eigenspectrum in the kernel space decays much faster than that in the linear space. Thus, in this work, we shall employ different approach to find the unreliable eigenvalues in the kernel space and replace them by a constant instead of the $1/f$ function.

To differentiate the unreliable eigenvalues from the larger ones we employ the ratios of the successive eigenvalues of the eigenspectrum to decompose the whole eigen-space into two subspaces: a principal or reliable subspace spanned mainly by the facial variation, $\mathbf{P} = \{\psi_k^w\}_{k=1}^m$ and an unreliable or noise dominating subspace due to limited number of training samples, $\bar{\mathbf{P}} = \{\psi_k^w\}_{k=m+1}^l$. For a clearer illustration, we first define the eigenratios as $\mathbf{\Gamma}_{\Phi}^{w} = \{\gamma_1^w, \ldots, \gamma_{r_w-1}^w\}$, such that

$$\gamma_k^w = \frac{\lambda_k^w}{\lambda_{k+1}^w}, \quad 1 \leq k < r_w. \tag{13}$$

The plot of eigenratios γ_k^w of a typical real eigenspectrum against the index k is called kernel eigenratiospectrum of the training data as shown in Fig. 1.

For a robust training, the database size should be significantly larger than the (face or reliable) dimensionality m. We examined several different face databases, the eigenratio plots shown in Fig. 1 is a general behavioral pattern that all the eigenratios of different databases portray. It is apparent from the graph that the eigenratios first decreases very rapidly, then stabilizes and finally increases. The increase of the eigenratios should not be the behavior of the true variances but occurs due to the limited number of training samples. The corresponding eigenvalues are therefore unreliable. Thus, one robust way of finding such a point would be finding the minimum of the eigenratios. The start point of the unreliable region $m+1$ is estimated by

$$\gamma_{m+1}^w = \min\{\forall \gamma_k^w, \quad 1 \leq k < r_w\}. \tag{14}$$

A typical such m value of a real kernel eigenspectrum is shown in Fig. 1.

The main purpose of finding the value of m using the eigenratios is to distinguish the reliable eigenvalues from the unreliable ones, which facilitates the decomposition of the entire eigenspace into reliable $\mathbf{P}$ and unreliable $\bar{\mathbf{P}}$ subspaces. Eigenvalues in the unreliable subspace $\bar{\mathbf{P}}$, spanned by $\{\psi_k^w\}_{k=m+1}^l$, will be regularized to facilitate the discriminant evaluation and feature extraction from the whole space of $\mathbf{S}_{\Phi}^{w}$ matrix (as described in the next section).

3 Scaling of kernel eigenvectors and feature extraction

3.1 Scaling of kernel eigenvectors

As pointed out in [6], the largest sample-based eigenvalues are biased high and the smallest ones are biased low due to the finite number of training samples. The eigenspectrum in the principal/reliable subspace is dominated by the face structural component, hence, we keep the eigenvalues in the principal subspace unchanged. In the unreliable subspace $\bar{\mathbf{P}}$, however, the limited number of training samples results in faster decay of the eigenvalues than the true variances. Therefore, the decay of the eigenvalues should be slowed down to compensate the effect of the finite number of training samples.

From Fig. 2 it is evident that when the inverses of the eigenvalues $\{\lambda_k^w\}_{k=m+1}^l$ are used for feature weighting (12), the corresponding eigenvectors get undue over-scaling in this range. We should not trust the eigenvalues, $\lambda_k, k > m$ as they are greatly effected by the finite number of training samples. Moghaddam et al. [24] replaces the small and zero eigenvalues by the average eigenvalue over the unreliable subspace and Wang et al. [32] adopted it in the dual-space approach. However, this may introduce additional overfitting problem in dimensions of the unreliable subspace whose eigenvalues are larger than the average. (There must be some eigenvalues in the unreliable subspace larger than the average being replaced by the smaller average eigenvalue). As eigenvalues in this subspace are biased smaller, their average may not be a good scaling factor relative to those in the principal subspace. Therefore, we propose to replace the unreliable eigenvalues $\{\lambda_k^w\}_{k=m+1}^l$ by the upper bound eigenvalue of the unreliable subspace, i.e.,

$$\lambda_{\text{const}}^w = \max\{\forall \lambda_k^w, \quad k \geq m\}. \tag{15}$$

Thus, the final weighting function can be written as

$$\tilde{w}_k^w = \begin{cases} 1/\sqrt{\lambda_k^w}, & k \leq m \\ 1/\sqrt{\lambda_{\text{const}}^w}, & m < k \leq l \end{cases}. \tag{16}$$

Figure 2 shows the proposed feature weighting function $\tilde{w}_k^w$ calculated by (13), (14), (15) and (16) comparing with that w_k^w of (12). The new weighting function $\tilde{w}_k^w$ is identical to w_k^w in the principal space and remains constant in the unreliable and null subspaces. Note that, different from those in [24,32], no eigenvalue in the unreliable space becomes smaller in our approach.

Using this weighting function and the eigenvectors ψ_k^w, training pattern data are transformed to

$$\tilde{Y}_{ij} = \tilde{\mathbf{\Psi}}_l^{w^T} \zeta_{ij}, \tag{17}$$

where

$$\tilde{\mathbf{\Psi}}_l^w = [\tilde{w}_k^w \psi_k^w]_{k=1}^l = [\tilde{w}_1^w \psi_1^w, \ldots, \tilde{w}_l^w \psi_l^w]. \tag{18}$$

There is no dimension reduction in this transformation as $\tilde{Y}_{ij}$ and ζ_{ij} have the same dimensionality l.

Problems of dimensionality reduction of KFDA were also discussed in [5], where a kernel machine-based regularized Fisher discriminant (K1PRFD) algorithm was proposed. This approach regularizes the within-class scatter matrix by adding a constant to all eigenvalues. As pointed out in [6], the bias of eigenvalues is most pronounced when the eigenvalues tend toward equality, and it is much less severe when their values are highly disparate. For the application of face recognition, it is well-known that the eigenspectrum first decays very rapidly and then stabilizes. Hence, adding a constant to the eigenspectrum may bias back the rapidly changing eigenvalues in principal space too much that introduces additional error source, and bias back the flat eigenvalues in null space too little at the same time [6].

3.2 Kernel eigenfeature extraction

After the feature scaling, a new between-class scatter matrix is formed by vectors $\tilde{Y}_{ij}$ of the training data as

$$\tilde{\mathbf{S}}_{\Phi}^b = \frac{1}{p}\sum_{i=1}^{p}(\overline{\tilde{Y}}_i - \overline{Y})(\overline{\tilde{Y}}_i - \overline{Y})^T, \tag{19}$$

where $\overline{\tilde{Y}}_i = \frac{1}{q_i}\sum_{j=1}^{q_i}\tilde{Y}_{ij}$ and $\overline{Y} = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{q_i}\sum_{j=1}^{q_i}\tilde{Y}_{ij}$. The weighted features $\tilde{Y}_{ij}$ will be de-correlated for $\tilde{\mathbf{S}}_{\Phi}^b$ by solving the eigenvalue problem as (11). Suppose that the eigenvectors in the eigenvector matrix $\tilde{\mathbf{\Psi}}_l^b = [\tilde{\psi}_1^b, \ldots, \tilde{\psi}_l^b]$ are sorted in descending order of the corresponding eigenvalues. The dimensionality reduction is performed here by keeping the eigenvectors with the d largest eigenvalues

$$\tilde{\mathbf{\Psi}}_d^b = [\tilde{\psi}_k^b]_{k=1}^d = [\tilde{\psi}_1^b, \ldots, \tilde{\psi}_d^b], \tag{20}$$

where d is the number of features usually selected by a specific application. Thus, the proposed feature scaling and extraction matrix $\mathbf{U}_{\Phi}$ is given by

$$\mathbf{U}_{\Phi} = \tilde{\mathbf{\Psi}}_l^w \tilde{\mathbf{\Psi}}_d^b. \tag{21}$$

This transforms a face sample vector ζ_{ij} of dimensionality l into a feature vector F of dimensionality d, by

$$F = \mathbf{U}_{\Phi}^T \zeta_{ij}. \tag{22}$$

Note that the dual-space approaches [32,33] extract features separately from two subspaces. As the discriminative information resides in the whole space, it is inefficient and only suboptimal to extract the discriminative features separately from two subspaces by two separate discriminant evaluations in two subspaces. Our method decomposes the kernel space into two subspaces only for the regularization of the eigenvalues in the unreliable subspace. We extract the discriminative features from the whole space by searching the most discriminative features in the full space. Thus, our method is based on the global optimization instead of two local optimizations of the dual-space approaches in [32,33].

3.3 The proposed algorithm

The proposed complete discriminant evaluation and feature extraction in kernel space for face recognition (CDEFE) approach is summarized below:
At the training stage:

1. Given a training set of face image vectors $\{X_{ij}\}$ and a kernel function $K(X_{ij}, X_{st})$, compute ζ_{ij} using (8).
2. Compute $\mathbf{S}_{\Phi}^w$ by (8) and solve the eigenvalue problem as (11).
3. Decompose the kernel eigenspace into principal- and unreliable-spaces by determining the m value using (13) and (14).
4. Transform the training pattern samples represented by ζ_{ij} into $\tilde{Y}_{ij}$ by (17) with the weighting function (16) determined by (13), (14) and (15).
5. Compute $\tilde{\mathbf{S}}_{\Phi}^b$ by (19) with $\tilde{Y}_{ij}$ and solve the eigenvalue problem as (11).
6. Obtain the final feature scaling and extraction matrix by (18), (20) and (21) with a predefined number of features d.

At the recognition stage:

1. Transform each n-D face image vector X into l-D feature pattern vector ζ_{ij} using the kernel function K and (8).
2. Transform each l-D feature pattern vector ζ_{ij} into d-D feature vector F by (22) using the feature regularization and extraction matrix $\mathbf{U}_{\Phi}$ obtained in the training stage.
3. Apply a classifier trained on the gallery set to recognize the probe feature vectors. (A simple first nearest neighborhood classifier is applied in the experiments)

3.4 Computational complexity

The difference in the computational complexity among various kernel based approaches largely depends on the training procedure and database structure. In general, the computational complexity of the training procedure of the proposed CDEFE approach is $\mathcal{O}(l^3)$, where l is the number of training samples. This is same as other kernel based subspace methods such as KFDA, KDDA, NKDA, CKFD and K1PRFD. However, some approaches require three eigen-decompositions and some need only two. The eigen-decomposition is the most time consuming part in the training. The proposed CDEFE approach requires only two eigen-decompositions. Different from KFDA, KDDA and NKDA methods, however, CDEFE, CKFD, K1PRFD and the dual space approach [32] need to compute all eigenvectors of the within-class scatter matrix. In this respect, CDEFE, CKFD, K1PRFD and dual-space approach in general may require additional computation comparing to KFDA, KDDA and NKDA. Note that K1PRFD uses an iterative conjugate gradient method to process the eigenvalues, which is time consuming.

After the processing of the within-class scatter matrix, CKFD and dual-space approach require two eigen-decompositions in two complementary subspaces while CDEFE and K1PRFD compute one eigen-decomposition in the whole space. As the number of subjects p is at most half of the total number of the training samples l (single sample per subject cannot contribute to the within-class scatter matrix), snapshot method [7] can be used. Using this method, CDEFE only requires one eigen-decomposition of the between-class scatter matrix of dimensionality $p \times p$, which is smaller than that of at least one of the two between-class scatter matrices of CKFD and dual-space approach. In this respect, the proposed CDEFE may require less computation than CKFD and dual-space approach, depending on the database structure.

Our experiments on several face databases (the descriptions of the face databases are given in the experimental section) show that KDDA training is the fastest, followed by NKDA and the proposed CDEFE. K1PRFD takes longer time of training than CDEFE as it uses an iterative conjugate gradient method to process the eigenvalues. KFDA and CKFD require the longest training time because they apply three and four eigen- decompositions, respectively.

The testing/recognition time of KFDA, KDDA, NKDA, K1PRFD and CDEFE are the same for the same number of features because the recognition procedure is computing a distance between the probe image and all the gallery images. Recognition time of CKFD is slightly more than the above five methods because it uses summed normalized distance between the probe and gallery images [33]. In practical face recognition systems, training is usually an off-line process and recognition is usually an online process. Thus, the recognition time is usually more critical than the training time. Although, the recognition time of the CDEFE approach is same as other approaches for the same number of features, it can be faster than other approaches for the same recognition rate because the proposed CDEFE approach, as we will see in the experiments, achieves a given recognition rate with fewer features than other approaches.

4 Experiments and discussions

In all experiments reported in this work, images are aligned and normalized following the CSU Face Identification Evaluation System [2]. Four databases: ORL, GT and two from FERET are used for testing. Each database is partitioned into training and testing sets. For FERET databases, there is no overlap in subject between the training and testing sets. As ORL and GT databases have only a small number of subjects, both training and testing sets contain all subjects. However, there is no overlap in the sample image between the training and testg sets. In our experiments, polynomial cosine kernel function is chosen, $K(X_{ij}, X_{st}) = \frac{\tilde{K}(X_{ij},X_{st})}{\sqrt{\tilde{K}(X_{ij},X_{ij})\tilde{K}(X_{st},X_{st})}}$, where $\tilde{K}(X_{ij}, X_{st}) = \langle \Phi(X_{ij}), \Phi(X_{st}) \rangle = (a\langle X_{ij} \cdot X_{st}\rangle + b)^c$, since cosine kernel gave good performances in the experiments of [15,17] and better performances than the original polynomial kernels [11,13,15,21]. The kernel parameters are set same as that in [15–17]. The recognition error rate given in this work is the percentage of the incorrect top 1 match on the testing set. The proposed CDEFE method is tested and compared with KFDA [22], KDDA [18], NKDA [17], CKFD [33] and K1PRFD [5] approaches. The parameters of CKFD are applied that are mentioned in the experiments of [33].

4.1 Results on FERET database 1

In FERET database, the face image variations include facial expression and other details (like glasses or no glasses), illumination, pose, and aging [26]. We select 2,388 images comprising of 1,194 subjects (two images per subject) from this

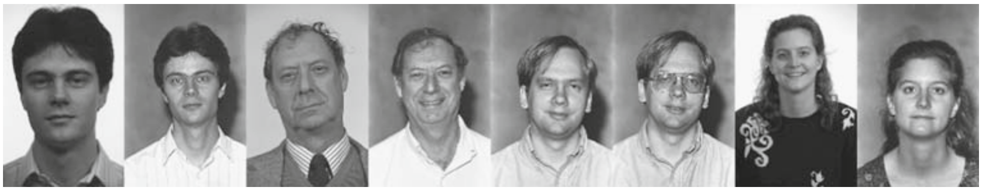

Fig. 3 Sample images from the FERET database for four subjects with two sample images per subject

database. Some of the sample images are displayed in Fig. 3. For this database, images are cropped into the size of 38 × 33.

Before comparing the proposed CDEFE method with other state-of-the-art kernel based approaches we first test the three contributions of this work: the subspace decomposition point m, the eigenspectrum regularization and the discriminant evaluation in the whole kernel space. Five hundred images of 250 subjects are randomly selected for training and the remaining 1,888 images of 944 subjects are used for testing. Figure 4 shows the recognition error rate on the testing set against the number of features d used in the matching. In Fig. 4, CDEFE-95% represents a variant of our approach where the subspace decomposition point m is determined by keeping 95% eigenvalue energy in the reliable subspace instead of using (14); CDEFE-Av represents another variant of our approach by replacing λ^w_{const} in (16) with the average eigenvalue over the unreliable subspace as used in [24] and in [32]; and CDEFE-Dual represents the third variant of our approach where features are extracted separately from the principal and unreliable subspaces as the idea of the dual-space approach in [32].

Figure 4 shows that our subspace decomposition point is much better than the empirical setting of keeping 95% eigenvalue energy. It also illustrates the advantage of the upper bound eigenvalue instead of the average eigenvalue for the eigenspectrum regularization. Furthermore, Fig. 4 clearly demonstrates that the global optimization in the whole space outperforms the two local optimization processes in two subspaces. In fact, the recent proposed CKFD approach [33] also extracts features separately from two subspaces, whose recognition performance is shown in all of the following experiments.

Figure 5 compares the proposed CDEFE method with other state-of-the-art kernel based approaches. Both KFDA and KDDA perform badly because they extract features only from the principal subspace. K1PRFD achieves slightly higher accuracy than NKDA and CKFD as it extracts features from the whole space. Our proposed CDEFE approach consistently outperforms all other approaches for all number of features and the accuracy gain is significant for smaller number of features.

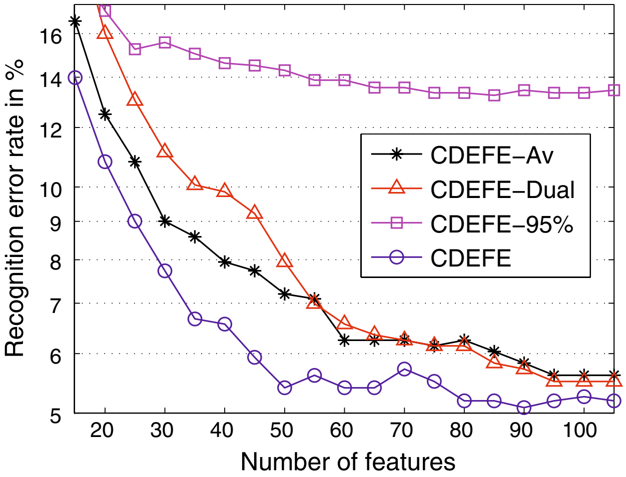

Fig. 4 Recognition error rate against the number of features used in the matching on the FERET database of 500 training images (250 subjects) and 1,888 testing images (944 subjects)

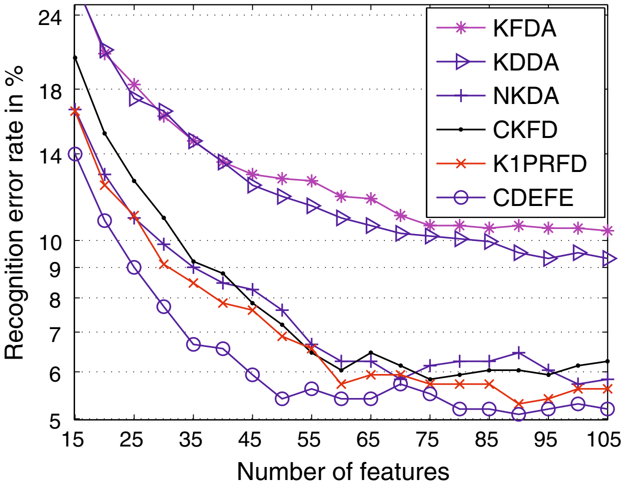

Fig. 5 Recognition error rate against the number of features used in the matching on the FERET database of 500 training images (250 subjects) and 1,888 testing images (944 subjects)

Since this database is larger than others, we conduct a second experiment with larger number of training samples where 497 subjects are randomly selected for training and the remaining images of 697 subjects are used for testing. Figure 6 shows the results. CKFD that uses information from both the subspaces performs better than KFDA, KDDA, NKDA and K1PRFD approaches. Although K1PRFD uses the whole space, it does not outperform CKFD. Possible reasons for this are, as stated previously, that adding a constant

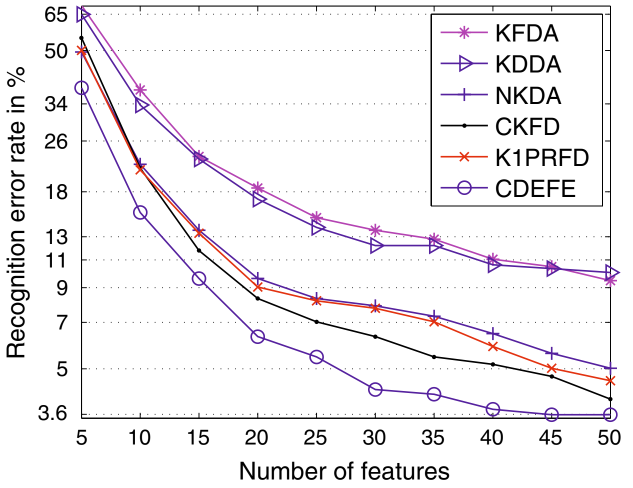

Fig. 6 Recognition error rate against the number of features used in the matching on the FERET database of 994 training images (497 subjects) and 1,394 testing images (697 subjects)

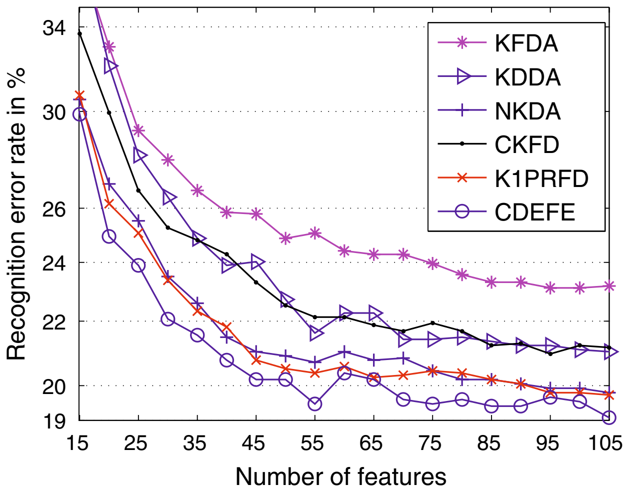

Fig. 7 Average recognition error rate against the number of features used in the matching on the FERET database of 512 training images (128 subjects) and 512 testing images (128 subjects)

to the eigenspectrum may bias back the rapidly changing eigenvalues too much that introduces additional error source, and bias back the flat eigenvalues too little at the same time. The proposed CDEFE approach again consistently outperforms other approaches.

4.2 Results on FERET database 2

This database is constructed by choosing 256 subjects with at least four images per subject. We use the same number of images (four) per subject for all subjects. Five hundred and twelve images of the first 128 subjects are used for training and the remaining 512 images serve as testing images. The size of the normalized image is 150 × 130, same as that in [20]. The ith images of all testing subjects are chosen to form gallery set and the remaining three images per subject serve as the probe images to be identified from the gallery set. Therefore, for each run there is only one image per subject in the gallery data set and three images per subject in the testing data set. Figure 7 shows the average recognition error rates over the four probe sets, each of which has a distinct gallery set ($i = 1, 2, 3, 4$).

Comparing to Fig. 6, the recognition error rates of all methods in Fig. 7 increase due to larger variation of the testing images. Unlike the previous two experiments, KDDA performs much better than KFDA because of the availability of more number of samples per subject during training. K1PRFD which uses the full kernel eigenspace performs better than CKFD, KDDA and KFDA methods. However, K1PRFD does not outperform NKDA consistently. Similar to the first two experiments, the proposed CDEFE approach achieves consistently lowest recognition error rate for all number of features in Fig. 7.

4.3 Results on ORL database

In the experiments on ORL database [27], images are cropped into the size of 57 × 50. The ORL database contains 400 images of 40 subjects (ten images per subject). Some images were captured at different times and have different variations including expression (open or closed eyes, smiling or non-smiling) and facial details (glasses or no glasses). The images were taken with a tolerance for some tilting and rotation of the faces up to 20°. Some of the samples images are displayed in Fig. 8.

In the first experiment, we test various approaches using the first five samples per subject (200 images) for training and the remaining five samples per subject (200 images) for testing. Figure 9 shows the recognition error rate on the testing set against the number of features. Similar to the previous experiments, our method consistently outperforms others.

To obtain more reliable results on ORL database, we conduct another experiment with leave-one-out training and testing strategy. In each of the 400 runs of the training and testing, one sample is picked out for testing and the remaining 399 samples are included in the training set. The testing results are numerically recorded in Table 1. KDDA outperforms both KFDA and NKDA but not consistently. Similar to the results of FERET database 2, KDDA performs better when more number of samples per subject are present in the training database. CKFD, which uses a summed normalized distance measures from the two subspaces outperforms KDDA and K1PRFD. However, CDEFE which evaluates the discriminative information in the whole eigenspace consistently outperforms CKFD, K1PRFD and all others. It shows that, for this training task, the complementary subspace is still useful but not well handled by the CKFD and K1PRFD approaches.

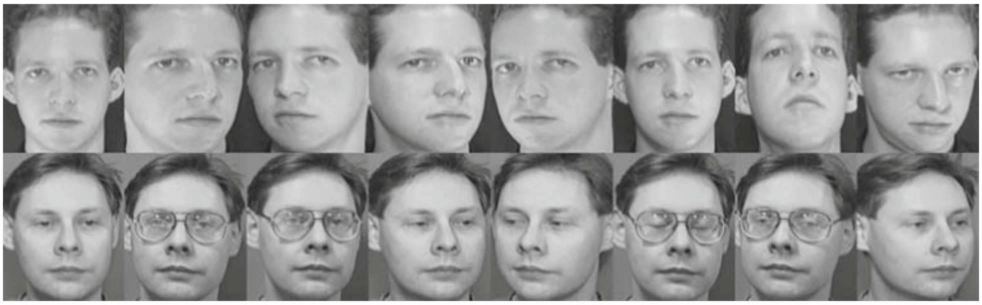

Fig. 8 Sample images from the ORL database for two subjects with eight sample images per subject

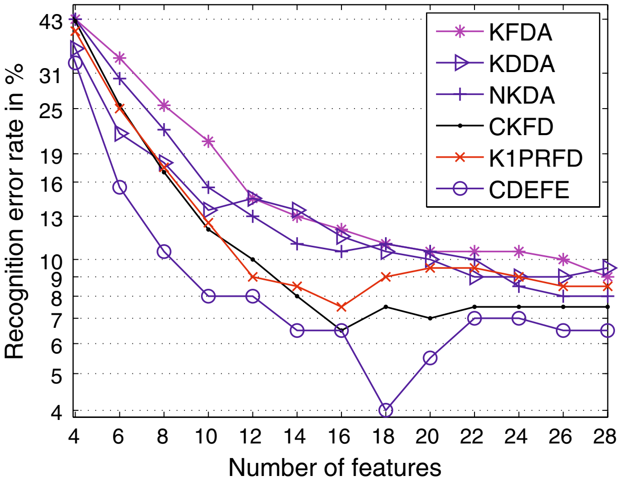

Fig. 9 Recognition error rate against the number of features used in the matching on the ORL database of 200 training images (40 subjects) and 200 testing images (40 subjects)

4.4 Results on georgia tech database

The Georgia Tech (GT) Face Database [1] consists 750 color images of 50 subjects (15 images per subject). Some of the GT sample images are shown in Fig. 10. For most of the subjects the face images were taken in two or three sessions over a period of three months, allowing for strong variation in size, facial expression, illumination, and rotation in both the image plane and perpendicular to the image plane. These images are converted to gray-scale and cropped into the size of 112 × 92. The first eight images of all the subjects are used in the training and the remaining seven images serve as testing images. The testing results are numerically recorded in Table 1. All the approaches perform relatively similar to the previous experiments. KDDA outperforms CKFD but not consistently. Both KDDA and CKFD outperform KFDA, NKDA and K1PRFD approaches but not consistently, this probably shows that KDDA and CKFD perform better when more number of samples per subject are present in the training database (similar to the previous experiments). The proposed CDEFE approach again consistently outperforms all other approaches for all number of features.

4.5 Summary of experiments

We have performed six sets of experiments with four different databases. The proposed CDEFE approach shows superior performance in the following five aspects: first, CDEFE consistently outperforms all other approaches in all six experiments, while no other approach can perform the second best consistently in all the experiments. Second, CDEFE achieves the lowest recognition error rate consistently for all number of features. In contrast to that, no other approach can perform the second best for all number of features even in a single experiment. Third, CDEFE avoids any kind of heuristic parameter setting in its implementation/application. Fourth,

Table 1 Recognition error rate of different approaches for different number of features

Database	ORL (leave-one-out training-testing)							GT (400/350 training/testing images)						
# Feature	6	8	10	20	32	36	38	6	10	14	20	28	38	46
KFDA	18.00	9.75	7.25	4.25	2.75	2.75	2.50	38.57	26.57	22.00	18.00	14.86	12.86	12.57
KDDA	15.75	8.50	6.75	2.75	2.50	3.00	2.75	33.14	22.00	15.43	12.00	9.14	9.14	9.43
NKDA	20.75	11.00	8.25	3.50	2.75	2.75	2.75	35.43	20.00	17.14	13.43	13.43	12.29	12.29
CKFD	15.50	9.00	6.00	2.25	2.25	1.75	2.00	38.00	23.43	15.71	12.86	10.86	10.29	8.86
K1PRFD	17.75	10.25	7.00	3.25	3.00	2.50	2.50	31.71	19.43	17.14	13.71	13.14	11.71	12.57
CDEFE	13.25	6.50	5.00	1.75	2.00	1.25	1.25	26.29	17.71	12.86	10.00	8.29	8.29	8.29

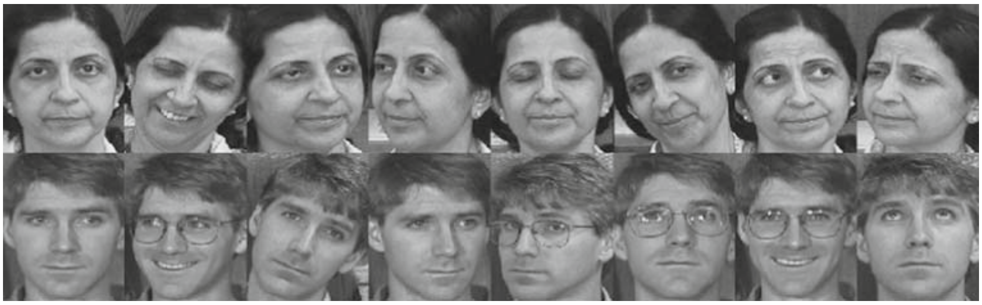

Fig. 10 Sample images from the GT database for two subjects with eight sample images per subject

although for certain number of features in certain experiments CDEFE achieves only marginal accuracy gains comparing to some other approaches, significant better performances of the CDEFE approach comparing to these approaches can be found in some other experiments. Fifth, CDEFE significantly outperforms all other approaches for small number of features. This demonstrates that the proposed CDEFE approach extracts more discriminative features than others.

4.6 Contributions of the proposed approach

There are three contributions in the proposed approach. First, CDEFE uses the minimum point of the eigenratiospectrum to decompose the kernel space into a reliable and an unreliable subspaces, which circumvents the heuristic parameter selection. Second, the proposed approach regularizes the eigenvalues in the unreliable subspace to the constant determined by the upper bound eigenvalue of the unreliable subspace. Different from other approaches that use the average eigenvalue, our method does not diminish any eigenvalue in the unreliable subspace. Third, CDEFE performs the discriminant evaluation in the whole kernel space, thereby, not lose out important discriminative information. Unlike the conventional methods which lose discriminative information either in the principal or in the null space and the recently proposed approaches which extract features separately from the two subspaces, our method extracts features from the whole kernel space which boosts its recognition performance as compared to others. This is verified and demonstrated through extensive experimentations.

5 Conclusions

In this paper, we have addressed the problems of eigenfeature scaling and its extraction from the whole kernel space. To facilitate the discriminative feature extraction from the whole kernel space, the ratios of the successive eigenvalues in the eigenspectrum are used to decompose the within-class eigenspace into a reliable and an unreliable subspaces. Eigenvalues of the unreliable subspace are regularized to the upper bound eigenvalue of this subspace. This circumvents the undue scaling of the eigenvectors corresponding to the unreliable small and zero eigenvalues and facilitates the feature extraction from the whole kernel space by the discriminant evaluation in the full kernel space. Experiments on the FERET, ORL and GT databases demonstrate that the proposed approach consistently outperforms other popular methods.

Acknowledgments This work was supported by research grant from the A*STAR Science & Engineering Research Council of Singapore (SERC grant no. 0621300056).

References

1. Georgia tech face database. http://www.anefian.com/face_reco.htm
2. Beveridge, R., Bolme, D., Teixeira, M., Draper, B.: The csu face identification evaluation system users guide: Version 5.0. Technical Report: http://www.cs.colostate.edu/evalfacerec/data/normalization.html (2003)
3. Cevikalp, H., Neamtu, M., Wilkes, M., Barkana, A.: Discriminative common vectors for face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **27**(1), 4–13 (2005)
4. Chen, L.F., Liao, H.Y.M., Ko, M.T., Lin, J.C., Yu, G.J.: A new lda-based face recognition system which can solve the small sample size problem. Pattern Recogn. **33**(10), 1713–1726 (2000)
5. Chen, W.S., Yuen, P.C., Huang, J., Dai, D.Q.: Kernel machine-based one-parameter regularized fisher discriminant method for face recognition. IEEE Trans. Syst. Man Cybern. Part B **35**(4), 659–669 (2005)
6. Friedman, J.H.: Regularized discriminant analysis. J. Am. Stat. Assoc. **84**(405), 165–175 (1989)
7. Fukunaga, K.: Introduction to Statistical Pattern Recognition. Academic, San Diego (1990)
8. Gupta, H., Agarwal, A.K., Pruti, T., Shekhar, C., Chellappa, R.: An experiment evaluation of linear and kernel-based methods for face recognition. IEEE Workshop Appl. Comput. Vis. (2002)
9. Jiang, X.D., Mandal, B., Kot, A.: Enhanced maximum likelihood face recognition. IEE Elect. Lett. **42**(19), 1089–1090 (2006)
10. Jiang, X.D., Mandal, B., Kot, A.: Eigenfeature regularization and extraction in face recognition. IEEE Trans. Pattern Anal. Machine Intell. **30**, DOI: 10.1109/TPAMI.2007.70,708 (2008)
11. Kin, K.I., Jung, K., Kim, H.J.: Face recognition using kernel principal component analysis. IEEE Signal Process. Lett. **9**(2), 40–42 (2002)

12. Liu, K., Cheng, Y.Q., Yang, J.Y., Liu, X.: An efficient algorithm for foley-sammon optical set of discriminant vectors by algebraic method. Int. J. Pattern Recogn. Artif. Intell. **6**, 817–829 (1992)
13. Liu, Q.S., Huang, R., Lu, H.Q., Ma, S.D.: Face recognition using kernel based fisher discriminant analysis. In: Proceedings of 5th IEEE International Conference on Automatic Face and Gesture Recognition, pp. 197–201 (2002)
14. Liu, Q.S., Lu, H., Ma, S.: Improving kernel fisher discriminant analysis for face recognition. IEEE Trans. Circuits Syst. Video Technol. **14**(1), 42–49 (2004)
15. Liu, Q.S., Lu, H.Q., Ma, S.D.: Improving kernel fisher discriminant analysis for face recognition. IEEE Trans. Circuits Syst. Video Technol. **14**(1), 42–49 (2004)
16. Liu, Q.S., Tang, X., Lu, H., Ma, S.D.: Face recognition using kernel scatter-difference-based discriminant analysis. IEEE Trans. Neural Netw. **17**(4), 1081–1085 (2006)
17. Liu, W., Wang, Y.W., Li, S.Z., Tan, T.N.: Null space-based kernel fisher discriminant analysis for face recognition. In: Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, pp. 369–374 (2004)
18. Lu, J., Plataniotis, K.N., Venetsanopoulos, A.N.: Face recognition using kernel direct discriminant analysis algorithms. IEEE Trans. Neural Netw. **14**(1), 117–126 (2003)
19. Lu, J., Plataniotis, K.N., Venetsanopoulos, A.N.: Face recognition using lda-based algorithms. IEEE Trans. Neural Netw. **14**(1), 195–200 (2003)
20. Lu, J., Plataniotis, K.N., Venetsanopoulos, A.N., Li, S.Z.: Ensemble-based discriminant learning with boosting for face recognition. IEEE Trans. Neural Netw. **17**(1), 166–178 (2006)
21. Yang, M., Ahuja, N., Kriegman, D.: Face recognition using kernel eigenfaces. IEEE Conf. Image Process. pp. 37–40 (2000)
22. Mika, S., Ratsch, G., Weston, J., Scholkopf, B., Muller, K.R.: Fisher discriminant analysis with kernels. IEEE Workshop Neural Netw. Signal Process. IX, pp. 41–48 (1999)
23. Mika, S., Ratsch, G., Weston, J., Scholkopf, B., Smola, A.J., Muller, K.R., Solla, S.A., Leen, T.K.: Invariant feature extraction and classification in kernel spaces. Adv. Neural Inform. Process. Syst. MIT Press, Cambridge, vol. 12, pp. 526–532 (2000)
24. Moghaddam, B., Pentland, A.: Probabilistic visual learning for object representation. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 696–710 (1997)
25. Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the face recognition grand challenge. IEEE Conf. Comp. Vis. Pattern Recogn. pp. 947–954 (2005)
26. Phillips, P.J., Moon, H., Rizvi, S., Rauss, P.: The feret evaluation methodology for face recognition algorithms. IEEE Trans. Pattern Anal. Mach. Intell. **22**(10), 1090–1104 (2000)
27. Samaria, F., Harter, A.: Parameterization of a stochastic model for human face identification. In: Proceedings of 2nd IEEE Workshop on Applications of Computer Vision, pp. 138–142. Sarasota, FL, USA (1994)
28. Scholkopf, B., Mika, S., Burges, C.J., Knirsch, P., Muller, K.R., Ratsch, G., Smola, A.: Input space versus feature space in kernel-based methods. IEEE Trans. Neural Netw. **10**, 1000–1017 (1999)
29. Scholkopf, B., Smola, A., Muller, K.R.: Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput. **10**(5), 1299–1319 (1998)
30. Shakhnarovich, G., Moghaddam, B.: Face Recognition in Subspaces. Springer, Heidelberg (2004)
31. Vapnik, V.N.: Statistical Learning Theory. Wiley, New York (1998)
32. Wang, X., Tang, X.: Dual-space linear discriminant analysis for face recognition. IEEE Conf. Comput. Vis. Pattern Recogn. **2**, pp. 564–569 (2004)
33. Yang, J., Frangi, A.F., Yang, J.Y., Zhang, D., Jin, Z.: Kpca plus lda: a complete kernel fisher discriminant framework for feature extraction and recognition. IEEE Trans. Pattern Anal. Mach. Intell. **27**(2), 230–244 (2005)
34. Yang, M.H.: Kernel eigenfaces vs. kernel fisherfaces: face recognition using kernel methods. In: Proceedings 5th IEEE International Conference on Automatic Face and Gesture Recognition, pp. 215–220 (2002)
35. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: a literature survey. ACM Comput. Surv. **35**(4), 399–458 (2003)

Author Biographies

X. Jiang received the B.Eng. and M.Eng. degrees from the University of Electronic Science and Technology of China (UESTC), in 1983 and 1986, respectively, and the Ph.D. degree from the University of German Federal Armed Forces, Hamburg, Germany, in 1997, all in electrical and electronic engineering. From 1986 to 1993, he was a lecturer at UESTC, where he received two Science and Technology Awards from the Ministry for Electronic Industry of China. From 1993 to 1997, he was with the University of German Federal Armed Forces as a scientific assistant. From 1998 to 2002, he was with Nanyang Technological University (NTU), Singapore, as a Senior Research Fellow, where he developed a fingerprint verification algorithm that achieved the most efficient and the second most accurate fingerprint verification in the International Fingerprint Verification Competition (FVC2000). From 2002 to 2004, he was a Lead Scientist and Head of the Biometrics Laboratory at the Institute for Infocomm Research, Singapore. He joined NTU as a faculty member in 2004. Currently, he serves as the Director of the Centre for Information Security, the School of Electrical and Electronic Engineering, NTU, Singapore. His research interest includes pattern recognition, signal and image processing, computer vision, and biometrics.

B. Mandal received the B.Tech degree in Electrical Engineering in 2003 from the Indian Institute of Technology, Roorkee, India. In 2001, he received the summer undergraduate research award (SURA) from the same Institute. He has over 1 year (2003–2004) of work experience in software engineering. Currently, from 2004 onwards, he is a Ph.D. candidate in the Department of Electrical and Electronic Engineering at Nanyang Technological University, Singapore. His research interests are in the fields of biometrics, pattern recognition, computer vision and machine learning. He is working on feature extraction using subspace methods for face identification and verification.

A. Kot received his BSEE and MBA degrees at University of Rochester, NY and a Ph.D. degree at University of Rhode Island, RI, USA. He was with AT&T briefly. Since 1991, he has been with Nanyang Technological University, Singapore. He headed the Division of Information Engineering at the School of Electrical and Electronic Engineering for eight years and is currently a Professor and Vice Dean (Research) for the School of Electrical and Electronic Engineering. He has published extensively in the areas of signal processing for communication, biometrics, data-hiding and authentication. Dr. Kot has served as Associate Editor for IEEE Transactions on Signal Processing, IEEE Transactions on Circuits and Systems II and IEEE Transactions on CSVT. He has served as Guest Editor for IEEE and EURASIP journals. Currently, he is an Associate Editor for IEEE Transactions on Circuits and Systems I and EURASIP JASP. He has served on numerous conference committees and technical committees, including co-chairing the IEEE International Conference on Image Processing (ICIP) in 2004 and an IEEE Distinguished Lecturer (2005–2006). He is a Fellow of IEEE.