



Multi-class Fukunaga Koontz discriminant analysis for enhanced face recognition



Felix Juefei-Xu*, Marios Savvides

CyLab Biometrics Center, Electrical and Computer Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA

ARTICLE INFO

Article history:

Received 9 September 2013

Received in revised form

21 July 2015

Accepted 9 October 2015

Available online 17 October 2015

Keywords:

Linear discriminant analysis

Unsupervised discriminant projection

Locality preserving projections

Fukunaga Koontz transform

Face recognition

ABSTRACT

Linear subspace learning methods such as Fisher's Linear Discriminant Analysis (LDA), Unsupervised Discriminant Projection (UDP), and Locality Preserving Projections (LPP) have been widely used in face recognition applications as a tool to capture low dimensional discriminant information. However, when these methods are applied in the context of face recognition, they often encounter the small-sample-size problem. In order to overcome this problem, a separate Principal Component Analysis (PCA) step is usually adopted to reduce the dimensionality of the data. However, such a step may discard dimensions that contain important discriminative information that can aid classification performance. In this work, we propose a new idea which we named Multi-class Fukunaga Koontz Discriminant Analysis (FKDA) by incorporating the Fukunaga Koontz Transform within the optimization for maximizing class separation criteria in LDA, UDP, and LPP. In contrast to traditional LDA, UDP, and LPP, our approach can work with very high dimensional data as input, without requiring a separate dimensionality reduction step to make the scatter matrices full rank. In addition, the FKDA formulation seeks optimal projection direction vectors that are orthogonal which the existing methods cannot guarantee, and it has the capability of finding the exact solutions to the "trace ratio" objective in discriminant analysis problems while traditional methods can only deal with a relaxed and inexact "ratio trace" objective. We have shown using six face database, in the context of large scale unconstrained face recognition, face recognition with occlusions, and illumination invariant face recognition, under "closed set", "semi-open set", and "open set" recognition scenarios, that our proposed FKDA significantly outperforms traditional linear discriminant subspace learning methods as well as five other competing algorithms.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

In the past two decades, researchers have been focused on finding meaningful subspaces such that the low dimensional representation of high dimensional data can facilitate better classification performance. For instance, principal component analysis (PCA) [1] aims to find a subspace that preserves well the second-order statistics and captures maximal variability of the data, but does not take into account class separation for the purpose of classification. Discriminant analysis methods such as Fisher's linear discriminant analysis (LDA) [2] and unsupervised discriminant projection (UDP) [3], on the other hand, seek to obtain subspaces where similarity criteria are enhanced, especially when the Gaussianity assumption (as in the LDA) does not hold for the training and testing data. Manifold approaches such as Isomap [4], locally linear embedding (LLE) [5], and Laplacian eigenmap (LE) [6]

conduct non-linear dimensionality reduction, with the assumption that the high dimensional data lies on a low dimensional manifold embedded within the ambient space. Locality preserving projections (LPP) [7] method is a direct linear approximation of Laplacian eigenmap and shares many of the data representation properties of nonlinear techniques such as Isomap, LLE, and LE. LPP finds linear projective subspaces that optimally preserve the neighborhood proximity structure of the data. Some recent advances in manifold-based face recognition can be found in [8–11].

The aforementioned linear discriminant subspace learning methods such as LDA, UDP and LPP all suffer from the small-sample-size problem [12], whenever the number of samples is smaller than the sample dimensionality. In this case, the sample scatter matrices can become singular in these methods, resulting in computational difficulty, due to the inversion of a singular matrix. To tackle this, a separate PCA step is adopted [2,3,7] to project images from the original image space into a face-subspace, where dimensionality is reduced to make certain scatter matrices

* Corresponding author.

non-singular and invertible. The downside of this is that the PCA step may discard dimensions that contain important discriminative information that can aid classification performance.

In this work, we propose new idea which we named multi-class Fukunaga Koontz discriminant analysis (FKDA) by incorporating the Fukunaga Koontz transform [13] within the optimization for maximizing class separation criteria in LDA, UDP, and LPP. In contrast to these traditional methods, our approach can work with very high dimensional data as input, without requiring a separate dimensionality reduction step to make the scatter matrices invertible. Moreover, our proposed FKDA formulation seeks optimal projection direction vectors that are orthogonal while traditional methods may not guarantee, and it has the capability of finding the exact solutions to the “trace ratio” objective in discriminant analysis problems while traditional methods can only deal with a relaxed and inexact “ratio trace” objective. We have performed face recognition experiments on the FRGC ver 2.0 database and YaleB database where we report significantly better results using our proposed FKDA method compared to LDA, UDP, and LPP.

The rest of the paper is organized as follows: Section 3 reviews the LDA, UDP, and LPP approach. Section 4 details the proposed FKDA method. Section 5 briefly describes the database to be used in the experiments as well as the preprocessing scheme. Section 6 reports the experimental results. Finally, we conclude my work in Section 7.

The major contributions of this work are: (1) we have proposed an alternative way of maximizing the class separation criteria in LDA, UDP, and LPP without deriving the generalized Rayleigh quotient. (2) We show that the optimization in LDA, UDP, and LPP in the ratio form can be equivalently replaced by the proposed fixed-sum form. (3) Such fixed-sum form in the proposed FKDA framework does not require any scatter matrices to be non-singular since no matrix inversion is required. (4) The optimal projection direction vectors obtained under FKDA formulation are orthogonal to each other which will aid the discriminability and classification performance. (5) The FKDA finds the exact solution to the “trace ratio” objective in the discriminant analysis problems while traditional methods in LDA, UDP, and LPP only solves for the relaxed and inexact “trace ratio” problem in the objective function. (6) We show that better face recognition performance can be achieved using the FKDA framework because our approach does not require a separate dimensionality reduction step using PCA which may discard important discriminant dimensions as in the traditional formulation of LDA, UDP, and LPP approaches, also the orthogonality of the projection vectors and the capability of finding exact solution to the “trace ratio” objective all add merits to the superiority of the FKDA.

2. Related work

The general problem we address in this paper has already been considered in the pattern recognition literature.

2.1. Small-sample-size problem

There are some related work on solving the small-sample-size problem raised in the execution of discriminant analysis. Since LDA is probably one of the most widely used and best known discriminant methods, the following related work will be limited to LDA method. However, similar limitations are also found in UDP and LPP formulations. To be discussed below are some well known approaches to solving the small-sample-size problem in LDA.

One of the most common ways to deal with the small-sample-size problems is to apply an intermediate dimension reduction

step using PCA to reduce the dimension of the original data before traditional LDA is carried out. The algorithm is known as PCA+LDA [2,14]. In this two-stage PCA+LDA algorithm, the discriminant stage is preceded by a dimension reduction stage using PCA. The dimension of the subspace transformed by PCA is chosen such as the “reduced” total scatter matrix \mathbf{S}_T or within-class scatter matrix \mathbf{S}_W in the subspace is nonsingular, so that classical LDA can be applied. A limitation of this approach is that the optimal value of the reduced dimension for PCA is difficult to determine. Moreover, the PCA stage may discard some useful dimensions that may contain discriminative information.

It has been suggested that the null space of the \mathbf{S}_W scatter matrix is important for discrimination. The claim is that applying PCA in Fisher's LDA may discard discriminative information since the null space of \mathbf{S}_W contains the most discriminative information. Upon this idea, direct LDA (DLDA) [15] method has been proposed, making use of the nullspace of the \mathbf{S}_W . DLDA derives eigenvectors after simultaneous diagonalization [12]. Unlike previous approaches, DLDA simultaneously diagonalizes the between-class scatter matrix \mathbf{S}_B first and then diagonalizes \mathbf{S}_W which can be expressed as $\mathbf{W}^\top \mathbf{S}_B \mathbf{W} = \mathbf{I}$ and $\mathbf{W}^\top \mathbf{S}_W \mathbf{W} = \Lambda$. The eigenvectors with very small (close to zero) eigenvalues in the \mathbf{S}_B can be discarded since they contain no discriminative power, while simultaneously keeping the eigenvectors with small eigenvalues in the \mathbf{S}_W , especially those in the null-space.

Another way to deal with the singularity of \mathbf{S}_T is to apply regularization, by adding some constant values to the diagonal elements of \mathbf{S}_T , as $\mathbf{S}_T + \mu \mathbf{I}_m$, for some $\mu > 0$, where \mathbf{I}_m is an identity matrix. It is easy to verify that $\mathbf{S}_T + \mu \mathbf{I}_m$ is positive definite, hence nonsingular. This approach is called regularized LDA (RLDA) [16]. It is evident that when $\mu \rightarrow \infty$, we lose the information on \mathbf{S}_T , while very small values of μ may not be sufficiently effective. Cross-validation is commonly applied for estimating the optimal μ . But in practice, it is always not an easy task to determine the optimal μ value because of the ad-hoc property of the parameter, especially when the training and testing data are not from the exact same distribution. For more studies on RLDA, readers can refer to [17,18].

One generalization of regularized LDA is called the penalized LDA (PLDA) [19]. Instead of regularizing the total scatter matrix, the PLDA penalizes, or regularize, the within-class scatter matrix as $\mathbf{S}_W + \Gamma$, for some penalty matrix Γ . Γ is symmetric and positive semidefinite which is more general than a diagonal matrix. The penalties can produce smoothness in the discriminant functions, hence bypass the singularity problem of the scatter matrix.

Pseudo-inverse is designed to tackle the matrix singularity problems by approximating the inversion solution in a least-squares sense. The pseudo Fisher linear discriminant analysis (PFLDA) [12,20] is based on the pseudo-inverse of the singular scatter matrices. The generalization error of PFLDA was studied in [21], when the size and dimension of the training data vary. Pseudo-inverses of the scatter matrices were also studied in [18]. Again, this circumvents the singularity problem in LDA formulation, but at the cost of replacing the original objective function with its approximate form.

By utilizing the generalized singular value decomposition (GSVD) [22], the LDA/GSVD algorithm [23,24] is developed. The criterion J_0 used in [24] is $J_0(\mathbf{W}) = \text{tr}((\mathbf{S}_B^L)^+ \mathbf{S}_W^L)$, where $(\mathbf{S}_B^L)^+$ denotes the pseudo-inverse of the between-class scatter matrix. LDA/GSVD aims to obtain the optimal projection \mathbf{W} that minimizes $J_0(\mathbf{W})$, subject to the constraint that $\text{rank}(\mathbf{W}^\top \mathbf{H}_B) = q$, where q is the rank of \mathbf{S}_B , and \mathbf{H}_B will be mentioned in the following section. The above constraint is enforced to preserve the dimension of the spaces spanned by the centroids in the original and transformed spaces. The optimal solution can be reached by applying the GSVD.

One downside of this method is the high computational cost of GSVD, especially for large-scale and high-dimensional data.

Huang et al. [25] make use of the null space of \mathbf{S}_W for solving the small-sample-size problem in LDA. The core idea is to project all the samples onto the null space of \mathbf{S}_W , where the within-class scatter is zero, and then the optimal discriminant vectors of LDA are those vectors that can maximize the between-class scatter \mathbf{S}_B . The authors propose a new method to tackle this problem in a more computationally efficient way.

Kyperountas et al. [26] utilize weighted piecewise discriminant hyperplanes in order to provide a more accurate discriminant decision than the one produced by the traditional LDA approach, while also avoiding the small-sample-size problem. To be more specific, the dimensionality of the samples are broken down into subsets of feature vectors of smaller dimensions, and LDA is applied on each subset. The resulting discriminant weight sets are weighted under a normalization criterion so that the piecewise discriminant functions can be continuous in order to provide the overall discriminant solution.

2.2. Trace ratio problem

Among many linear subspace modeling and embedding methods, their objective is to maximize the “trace ratio” i.e. $\text{maximize}_{\mathbf{W}} \text{tr}(f_1(\mathbf{W}))/\text{tr}(f_2(\mathbf{W})) = \text{maximize}_{\mathbf{W}} \text{tr}(\mathbf{W}^\top \mathbf{S}_i \mathbf{W})/\text{tr}(\mathbf{W}^\top \mathbf{Z}_i \mathbf{W})$ where \mathbf{S}_i and \mathbf{Z}_i are the scatter matrices for some discriminant analysis, to be discussed in the following sections. However, this optimization is non-convex hence solving it directly is hard. Traditionally, this is evaded by relaxing it to solving the “ratio trace” problem i.e. $\text{maximize}_{\mathbf{W}} \text{tr}(f_1(\mathbf{W})/f_2(\mathbf{W})) = \text{maximize}_{\mathbf{W}} \text{tr}(f_2(\mathbf{W})^{-1} f_1(\mathbf{W})) = \text{maximize}_{\mathbf{W}} \text{tr}((\mathbf{W}^\top \mathbf{Z}_i \mathbf{W})^{-1} (\mathbf{W}^\top \mathbf{S}_i \mathbf{W}))$. Simpler solution can therefore be found. But the result is no longer optimal for the original optimization.

One of the earliest attempts to solve the “trace ratio” problem directly can be traced back to Guo et al. [27] where they propose the generalized Foley–Sammon transform (GFST) on the basis of generalized Fisher discriminant criterion. Following their derivation, an iterative algorithm is proposed to converge to the precise “trace ratio” solution. Shen et al. [28] reformulate the original non-convex “trace ratio” problem such that it can be solved by a sequence of semi-definite feasibility problems efficiently, and the solution is globally optimal via semi-definite programming. One by-product of their method is that the projection matrix is naturally orthonormal. Jia et al. [29] provide a theoretical overview of the globally optimal solution to the “trace ratio” problem through the equivalent trace difference problem where they first introduce the eigenvalue perturbation theory to derive an efficient algorithm that is based upon the Newton–Raphson method.

After extensive literature search, we find that there has not been many studies on the exact recovery of the non-convex solution, and this work, FKDA, will join them to provide an alternative for finding the exact solution to the “trace ratio” problem.

3. Linear subspace learning methods review

In this section, the three most widely used linear discriminant subspace learning methods are reviewed, i.e. LDA, UDP, and LPP, with emphases on traditional ways of solving the optimization problems using generalized Rayleigh quotient.

3.1. Linear discriminant analysis

Linear discriminant analysis (LDA) [2] aims to find the projection direction such that the ratio of the between-class scatter and the within-class scatter is maximized, known as the Fisher's

criterion. The between-class scatter matrix \mathbf{S}_B and the within-class scatter matrix \mathbf{S}_W are defined as

$$\mathbf{S}_B = \sum_{i=1}^C N_i (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^\top \quad (1)$$

$$\mathbf{S}_W = \sum_{i=1}^C \sum_{\mathbf{x}_k \in C_i} (\mathbf{x}_k - \boldsymbol{\mu}_i)(\mathbf{x}_k - \boldsymbol{\mu}_i)^\top \quad (2)$$

where $\boldsymbol{\mu}_i$ is the mean image of class C_i , $\boldsymbol{\mu}$ is the mean image of all the images, C is the total class number, and N_i is the number of images in class C_i . The data matrix $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_C)$, where $\mathbf{X}_i \in \mathbb{R}^{d \times N_i}$.

The within-class and between-class scatter matrices can also be written as matrix form as follows: $\mathbf{S}_B = \mathbf{H}_B \mathbf{H}_B^\top$ and $\mathbf{S}_W = \mathbf{H}_W \mathbf{H}_W^\top$, where

$$\mathbf{H}_B = ((\boldsymbol{\mu}_1 - \boldsymbol{\mu})\mathbf{e}_1^\top, (\boldsymbol{\mu}_2 - \boldsymbol{\mu})\mathbf{e}_2^\top, \dots, (\boldsymbol{\mu}_C - \boldsymbol{\mu})\mathbf{e}_C^\top)$$

$$\mathbf{H}_W = (\mathbf{X}_1 - \boldsymbol{\mu}_1 \mathbf{e}_1^\top, \mathbf{X}_2 - \boldsymbol{\mu}_2 \mathbf{e}_2^\top, \dots, \mathbf{X}_C - \boldsymbol{\mu}_C \mathbf{e}_C^\top)$$

and $\mathbf{e}_i^\top = (1, \dots, 1)^\top \in \mathbb{R}^{N_i \times 1}$.

The optimal projection \mathbf{w} is chosen such that the Fisher criterion $J(\mathbf{w})$ is satisfied, i.e., the ratio of between-class scatter and within-class scatter is maximized:

$$\mathbf{w}^* = \arg \max_{\mathbf{w}} J(\mathbf{w}) = \arg \max_{\mathbf{w}} \frac{\text{tr}(\mathbf{w}^\top \mathbf{S}_B \mathbf{w})}{\text{tr}(\mathbf{w}^\top \mathbf{S}_W \mathbf{w})} \quad (3)$$

where $\{\mathbf{w}_i | i = 1, 2, \dots, m\}$ is a set of generalized eigenvectors of \mathbf{S}_B and \mathbf{S}_W , corresponding to the m largest generalized eigenvalues $\{\lambda_i | i = 1, 2, \dots, m\}$. The upper bound of m is $C - 1$, where again C is the total number of classes. This generalized eigenvalue problem can be shown as:

$$\mathbf{S}_B \mathbf{w}_i = \lambda_i \mathbf{S}_W \mathbf{w}_i \implies \mathbf{S}_W^{-1} \mathbf{S}_B \mathbf{w}_i = \lambda_i \mathbf{w}_i \quad (4)$$

This only holds when \mathbf{S}_W is invertible. However, in the face recognition problem, \mathbf{S}_W is very often singular due to the fact that its rank is at most $(N - C)$, where N is the total number of images in the training set, which is much smaller than the image dimension.

To overcome this problem and make LDA work, a separate PCA step is applied to reduce the dimensionality of the images to $(N - C)$ and then standard LDA is executed to reduce the dimension to $(C - 1)$.

3.2. Unsupervised discriminant projection

Instead of using the within-class and between-class scatter matrices as in LDA, unsupervised discriminant projection (UDP) [3,30] tries to find an optimal projection that minimizes the local scatter matrix \mathbf{S}_L and maximizes the non-local scatter matrix \mathbf{S}_N simultaneously. Contrary to LDA, UDP is an unsupervised classification method where the training labels are not taken into account. Given a training set in C classes with N un-labeled items $\mathbf{x}_1, \dots, \mathbf{x}_N$ where $\mathbf{x}_i \in \mathbb{R}^d$, the optimal projection \mathbf{w} for UDP is obtained by maximizing the objective function $J(\mathbf{w})$:

$$\mathbf{w}^* = \arg \max_{\mathbf{w}} J(\mathbf{w}) = \arg \max_{\mathbf{w}} \frac{\text{tr}(\mathbf{w}^\top \mathbf{S}_N \mathbf{w})}{\text{tr}(\mathbf{w}^\top \mathbf{S}_L \mathbf{w})} \quad (5)$$

where \mathbf{S}_L and \mathbf{S}_N are the local and non-local scatter:

$$\mathbf{S}_L = \frac{1}{2} \sum_i \sum_j \mathbf{A}_{ij} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top \quad (6)$$

¹ In this paper, we use bold lower-case letter to represent a column vector (or column-wise vectorized image), and we use bold upper-case letter to represent a matrix.

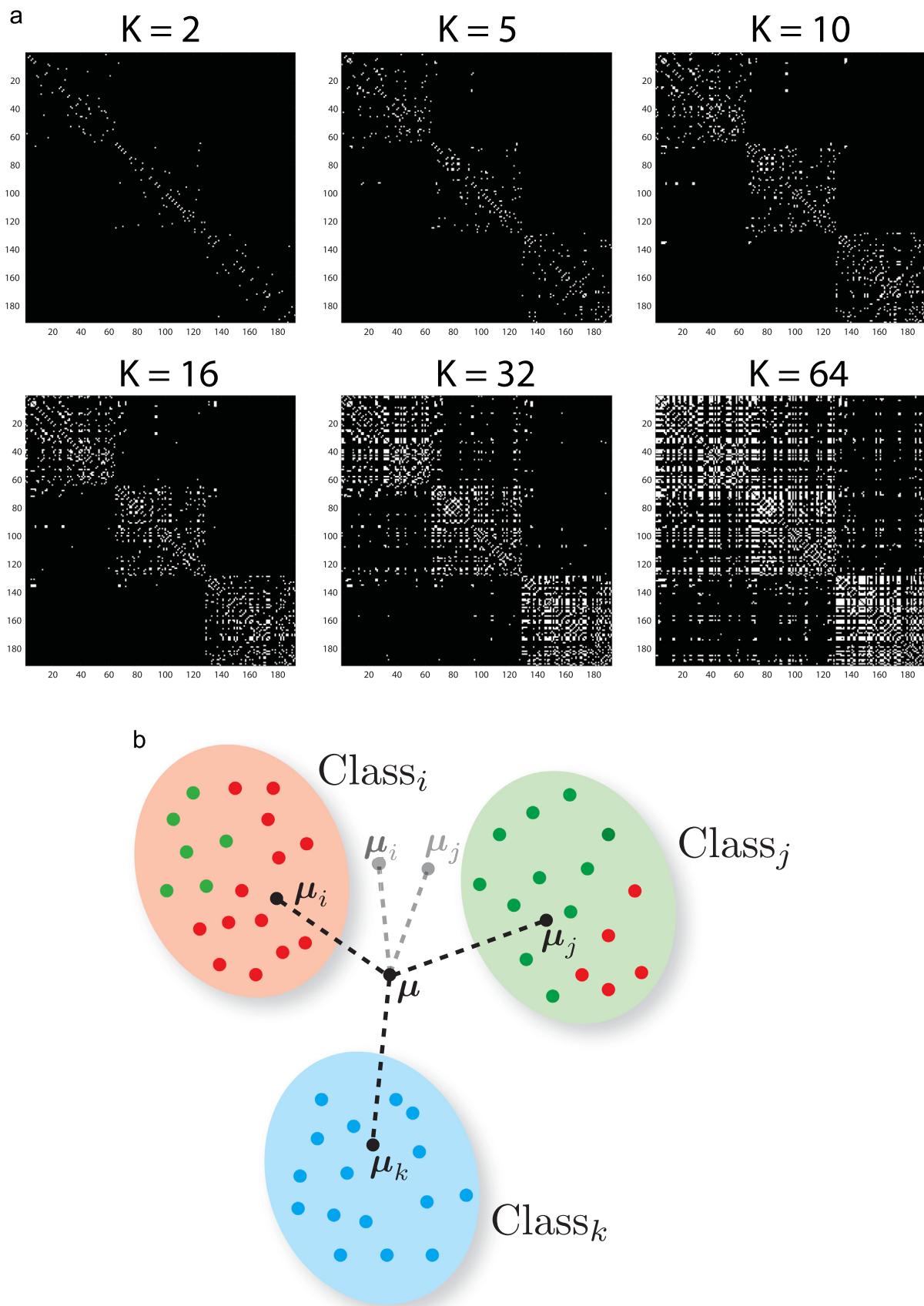


Fig. 1. (a) The adjacency map in UDP for a 3-class problem given various K in the KNN. (b) The advantage of UDP over LDA when handling outliers.

$$\mathbf{S}_N = \frac{1}{2} \sum_i \sum_j (1 - \mathbf{A}_{ij}) (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)^T \quad (7)$$

\mathbf{A} is an adjacency matrix that is produced by K -nearest neighbors (KNN) method:

$$\mathbf{A}_{ij} = \begin{cases} 1, & \text{if } \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are mutually KNN} \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

The local and non-local scatter matrices can also be written in the matrix form as follows:

$$\begin{aligned} \mathbf{S}_L &= \frac{1}{2} \left(\sum_{i=1}^N \sum_{j=1}^N \mathbf{A}_{ij} \mathbf{x}_i \mathbf{x}_i^T + \sum_{i=1}^N \sum_{j=1}^N \mathbf{A}_{ij} \mathbf{x}_j \mathbf{x}_j^T \right) - \frac{1}{2} \left(2 \sum_{i=1}^N \sum_{j=1}^N \mathbf{A}_{ij} \mathbf{x}_i \mathbf{x}_j^T \right) \\ &= \mathbf{X} \mathbf{D}_L \mathbf{X}^T - \mathbf{X} \mathbf{A} \mathbf{X}^T = \mathbf{X} \mathbf{L}_L \mathbf{X}^T \end{aligned} \quad (9)$$

where \mathbf{X} is the data matrix, and \mathbf{D} is a diagonal matrix whose elements on the diagonal are the column/row sum of adjacency matrix \mathbf{A} . $\mathbf{L}_L = \mathbf{D}_L - \mathbf{A}$ is also known as the Laplacian matrix for the local scatter. Similarly, the non-local scatter matrix can be written as

$$\mathbf{S}_N = \mathbf{X} \mathbf{L}_N \mathbf{X}^T \quad (10)$$

where $\mathbf{L}_N = \mathbf{D}_N - \mathbf{A}_N$. Here, $\mathbf{A}_N = \mathbf{1} - \mathbf{A}$, and \mathbf{D}_N is a diagonal matrix whose elements on the diagonal are column/row sum of \mathbf{A}_N . Eq. (5) can be solved by the generalized eigenvalue problem: $\mathbf{S}_N \mathbf{w} = \lambda \mathbf{S}_L \mathbf{w}$. Again, it requires that \mathbf{S}_L is non-singular. Similar to LDA method, a separate PCA step is applied to reduce the dimensionality of the data before UDP is carried out.

The choice of K is empirical. Higher K will lead to denser adjacency map as shown in Fig. 1(a). In this work, we choose $K=5$ through cross validation.

3.2.1. UDP and LDA

Although supervised method such as the LDA (which assumes Gaussianity in the training and testing samples) exploits class label information, and should lead to better class discrimination. However, in real-world application, such supervised method could be inferior to some unsupervised algorithms.

Compared to LDA, UDP handles outliers better due to its unsupervised characteristics. In real-world application, instances of the same subject may not appear to be within a perfect cluster. This may due to mis-labeling or simply noise embedded in the system. As shown in Fig. 1(b), there are 3 data clusters belonging to 3 classes. But, there are several green dots in the red cluster, and several red dots in the green cluster. In this case, LDA will hurt because the class-means μ_i and μ_j are shifted dramatically, thus leading to erroneous subspace modeling. While UDP will remain correct because it relies on unsupervised clustering.

3.3. Locality preserving projections

Locality preserving projections (LPP) [7] seeks to preserve the local structure and intrinsic geometry of the data, unlike PCA and LDA whose views are the global structure. The objective function of LPP is defined as follows:

$$\begin{aligned} J(\mathbf{w}) &= \frac{1}{2} \sum_{ij} (\mathbf{w}^T \mathbf{x}_i - \mathbf{w}^T \mathbf{x}_j)^2 \mathbf{A}_{ij} = \sum_{ij} \mathbf{w}^T \mathbf{x}_i \mathbf{A}_{ij} \mathbf{x}_i^T \mathbf{w} - \sum_{ij} \mathbf{w}^T \mathbf{x}_i \mathbf{A}_{ij} \mathbf{x}_j^T \mathbf{w} \\ &= \sum_i \mathbf{w}^T \mathbf{x}_i \mathbf{D}_{ii} \mathbf{x}_i^T \mathbf{w} - \mathbf{w}^T \mathbf{X} \mathbf{A} \mathbf{X}^T \mathbf{w} = \mathbf{w}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{w} - \mathbf{w}^T \mathbf{X} \mathbf{A} \mathbf{X}^T \mathbf{w} \\ &= \mathbf{w}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{w} \end{aligned} \quad (11)$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ is the data matrix, \mathbf{D} is a diagonal matrix whose entries on the diagonal are the column/row sum of \mathbf{A} , $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ji}$. $\mathbf{L} = \mathbf{D} - \mathbf{A}$ is the Laplacian matrix [31]. \mathbf{A} is an adjacency

matrix defined as

$$\mathbf{A}_{ij} = \begin{cases} e^{-\|\mathbf{x}_i - \mathbf{x}_j\|^2/t}, & \text{if } \mathbf{x}_i, \mathbf{x}_j \text{ are mutually KNN} \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

The optimal projection direction is obtained by minimizing the objective function $J(\mathbf{w})$, with a constraint imposed: $\mathbf{w}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{w} = 1$. Therefore, the final optimization is shown as follows:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}, \mathbf{w}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{w} = 1} \mathbf{w}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{w} \quad (13)$$

which is equivalent to the following generalized Rayleigh quotient form as used in LDA and UDP:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{w}}{\mathbf{w}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{w}} = \arg \max_{\mathbf{w}} \frac{\text{tr}(\mathbf{w}^T \mathbf{S}_D \mathbf{w})}{\text{tr}(\mathbf{w}^T \mathbf{S}_L \mathbf{w})} \quad (14)$$

where $\mathbf{S}_D = \mathbf{X} \mathbf{D} \mathbf{X}^T$ and $\mathbf{S}_L = \mathbf{X} \mathbf{L} \mathbf{X}^T$. The optimal projection direction \mathbf{w} and the linear subspace spanned by \mathbf{w} is obtained by the maximum eigenvalue solution to the generalized eigenvalue problem:

$$\mathbf{S}_D \mathbf{w} = \lambda \mathbf{S}_L \mathbf{w} \quad (15)$$

When LPP is applied to high dimensional data, a separate PCA step is also required to reduce the dimensionality of the data so that \mathbf{S}_L becomes non-singular and invertible, just like in LDA and UDP formulation. LPP is a linear approximation of the Laplacian eigenmap [6], where the Laplacian matrix for finite graph in LPP is analogous to the Laplace Beltrami operator on compact Riemannian manifolds. More details of the linear approximation is in [32].

3.3.1. UDP and LPP

Deng et al. [33] have pointed out that UDP is a simplified version of LPP under the assumption that the local density is uniform. In this paper, we still show experimental results for both algorithms for the following reasons. First of all, the assumption of the local density being uniform is hardly achieved in the context of face recognition. Therefore, the UDP and LPP formulation hardly returns the same optimal projection subspaces. Empirically and experimentally, in the FRGC database, UDP achieves better results, because naturally lit face images are closer to uniformly distributed in the high dimensional space, while in the YaleB database, LPP is sometimes better, because faces with illumination variations are not uniformly distributed. UDP and LPP has its own advantage, depending on the database, therefore, we include both in the evaluation. Second, UDP is an unsupervised method while LPP can be applied in a supervised way. We have shown the eigenfaces for both methods in Fig. 2, and they are visually different from each other. These are the major reasons we keep both methods in the FKDA formulation and study how FKDA can improve upon both UDP and LPP methods.

4. Proposed method

In this section, the Fukunaga Koontz transform (FKT) method is first reviewed and then the proposed multi-class Fukunaga Koontz discriminant analysis (FKDA) is described.

The traditional FKT is similar to PCA in terms of dimensionality reduction but does a better job in discriminating two classes. The FKT incorporates data from both positive and negative classes and uses eigen decomposition on the joint correlation matrix in order to find the optimal basis vectors that very well represent one class while having the least representation power on the other class.

Let $\mathbf{X} \in \mathbb{R}^{d \times m}$ be the data set containing the images from Class 1, with each column a vectorized image with dimension d . Let $\mathbf{Y} \in \mathbb{R}^{d \times n}$ be the data set containing all the images from Class 2. Both \mathbf{X}

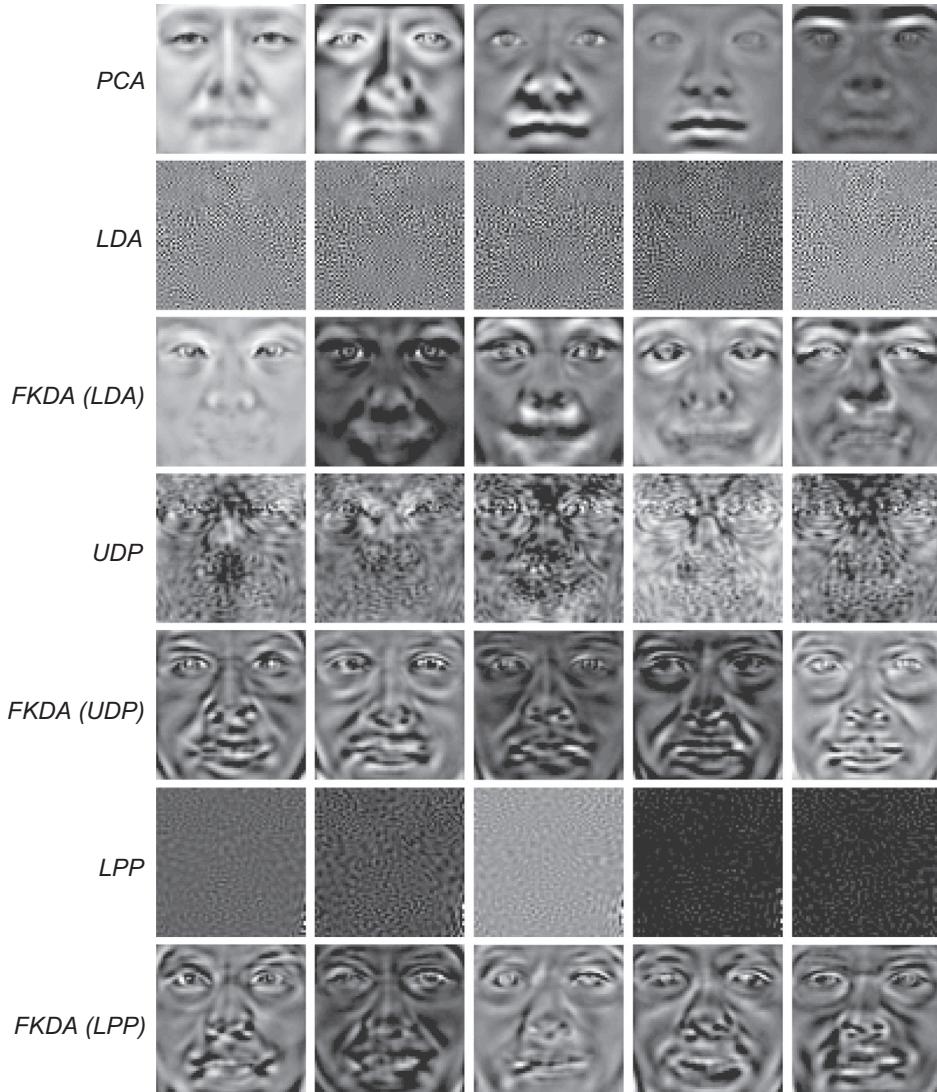


Fig. 2. Visualization of the top 5 projection direction vectors $\mathbf{w}_1, \dots, \mathbf{w}_5$ obtained by traditional linear subspace learning methods as well as the proposed FKDA framework.

and \mathbf{Y} are mean removed. The total correlation matrix Σ of both the Class 1 and Class 2 images are the summation of the correlation for each set $\Sigma_{\mathbf{X}}$ and $\Sigma_{\mathbf{Y}}$. The total correlation matrix Σ is symmetric and can be diagonalized using eigen-decomposition as

$$\Sigma = \Sigma_{\mathbf{X}} + \Sigma_{\mathbf{Y}} = \mathbf{X}\mathbf{X}^T + \mathbf{Y}\mathbf{Y}^T = \Phi\Lambda\Phi^T \quad (16)$$

where Φ contains the entire span of eigenvectors of Σ and Λ houses the corresponding eigenvalues of Σ on its diagonal. Next, a linear transformation step is applied in the FKT. Both the Class 1 and Class 2 data are transformed by a linear transformation matrix $\mathbf{P} = \Phi\Lambda^{-1/2}$. So the transformed data $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$ becomes $\hat{\mathbf{X}} = \mathbf{P}^T \mathbf{X}$ and $\hat{\mathbf{Y}} = \mathbf{P}^T \mathbf{Y}$.

Therefore, the correlation matrices of the transformed Class 1 data $\hat{\mathbf{X}}$ and Class 2 data $\hat{\mathbf{Y}}$ become

$$\Sigma_{\hat{\mathbf{X}}} = \hat{\mathbf{X}}\hat{\mathbf{X}}^T = \mathbf{P}^T \mathbf{X}\mathbf{X}^T \mathbf{P} = \mathbf{P}^T \Sigma_{\mathbf{X}} \mathbf{P} \quad (17)$$

$$\Sigma_{\hat{\mathbf{Y}}} = \hat{\mathbf{Y}}\hat{\mathbf{Y}}^T = \mathbf{P}^T \mathbf{Y}\mathbf{Y}^T \mathbf{P} = \mathbf{P}^T \Sigma_{\mathbf{Y}} \mathbf{P} \quad (18)$$

The transformed total correlation matrix for both Class 1 and Class 2 data becomes

$$\hat{\Sigma} = \Sigma_{\hat{\mathbf{X}}} + \Sigma_{\hat{\mathbf{Y}}} = \mathbf{P}^T \Sigma_{\mathbf{X}} \mathbf{P} + \mathbf{P}^T \Sigma_{\mathbf{Y}} \mathbf{P} \quad (19)$$

$$= \mathbf{P}^T (\Sigma_{\mathbf{X}} + \Sigma_{\mathbf{Y}}) \mathbf{P} = \mathbf{P}^T \Sigma \mathbf{P} = \mathbf{I} \quad (20)$$

So, the new total correlation matrix ($\mathbf{P}^T \Sigma \mathbf{P}$), which is essentially the old correlation matrix Σ linearly transformed using \mathbf{P} , is actually an identity matrix. This is because we have performed a global linear transformation instead of a class-specific linear transformation to sphere the data.

Here, we again perform an eigen-decomposition on the correlation matrix of Class 1 $\Sigma_{\hat{\mathbf{X}}}$, which yields $\Sigma_{\hat{\mathbf{X}}} \mathbf{v} = \lambda \mathbf{v}$.

From Eq. (20) we can obtain the following by multiplying \mathbf{v} on both sides:

$$\Sigma_{\hat{\mathbf{X}}} \mathbf{v} + \Sigma_{\hat{\mathbf{Y}}} \mathbf{v} = \mathbf{v} \quad (21)$$

With Eqs. (20) and (21), we have

$$\Sigma_{\hat{\mathbf{Y}}} \mathbf{v} = \mathbf{v} - \Sigma_{\hat{\mathbf{X}}} \mathbf{v} = \mathbf{v} - \lambda \mathbf{v} = (1 - \lambda) \mathbf{v} \quad (22)$$

This effectively means that the correlation matrices from two classes share the same eigenvectors \mathbf{v} and the eigenvalue of one class is exactly the complement of the eigenvalue of the other class. Taking advantage of the complementary property of the eigenvalues, the eigenvectors that are the most dominant in one class, are the least dominant in the other class. Therefore, in the traditional FKT method as applied to any two-class problem, a discriminative subspace is created by selecting a few of the most

dominant eigenvectors for one class and the least dominant ones for the other class. By ignoring the eigenvectors in the middle range, the subspace obtained contains basis that are highly discriminative and will yield discriminative feature selection after projection.

The limit of traditional FKT is that it, by design, is a tool to discriminate two classes. However, real-world face recognition problem is usually multi-class, so in order to extend FKT to accommodate multiple classes, we replace the correlation Σ_X and Σ_Y in the traditional FKT with \mathbf{S}_B , \mathbf{S}_W from LDA, \mathbf{S}_N , \mathbf{S}_L from UDP, and \mathbf{S}_D , $\mathbf{S}_{L'}$ from LPP as follows:

$$\mathbf{S}_{lda} = \mathbf{S}_B + \mathbf{S}_W = \Phi_{lda} \Lambda_{lda} \Phi_{lda}^\top \quad (23)$$

$$\mathbf{S}_{udp} = \mathbf{S}_N + \mathbf{S}_L = \Phi_{udp} \Lambda_{udp} \Phi_{udp}^\top \quad (24)$$

$$\mathbf{S}_{lpp} = \mathbf{S}_D + \mathbf{S}_{L'} = \Phi_{lpp} \Lambda_{lpp} \Phi_{lpp}^\top \quad (25)$$

A global transformation matrix is first obtained through diagonalizing the sum of the scatter matrices \mathbf{S}_{lda} , \mathbf{S}_{udp} , \mathbf{S}_{lpp} , and then the corresponding scatter matrices are transformed using the corresponding linear transformation matrix $\mathbf{P}_{lda} = \Phi_{lda} \Lambda_{lda}^{-1/2}$, $\mathbf{P}_{udp} = \Phi_{udp} \Lambda_{udp}^{-1/2}$, and $\mathbf{P}_{lpp} = \Phi_{lpp} \Lambda_{lpp}^{-1/2}$ as follows:

$$\widehat{\mathbf{S}}_B = \mathbf{P}_{lda}^\top \mathbf{S}_B \mathbf{P}_{lda} \quad \text{and} \quad \widehat{\mathbf{S}}_W = \mathbf{P}_{lda}^\top \mathbf{S}_W \mathbf{P}_{lda} \quad (26)$$

$$\widehat{\mathbf{S}}_N = \mathbf{P}_{udp}^\top \mathbf{S}_N \mathbf{P}_{udp} \quad \text{and} \quad \widehat{\mathbf{S}}_L = \mathbf{P}_{udp}^\top \mathbf{S}_L \mathbf{P}_{udp} \quad (27)$$

$$\widehat{\mathbf{S}}_D = \mathbf{P}_{lpp}^\top \mathbf{S}_D \mathbf{P}_{lpp} \quad \text{and} \quad \widehat{\mathbf{S}}_{L'} = \mathbf{P}_{lpp}^\top \mathbf{S}_{L'} \mathbf{P}_{lpp} \quad (28)$$

The sum of the two transformed scatter matrices now becomes the identity matrix, and by applying eigen-analysis on the transformed scatter matrix which is supposed to be maximized ($\widehat{\mathbf{S}}_B$ in LDA, $\widehat{\mathbf{S}}_N$ in UDP, and $\widehat{\mathbf{S}}_D$ in LPP), we can obtain a set of eigenvectors that achieve maximal class separation as in the traditional Rayleigh quotient formulation:

$$\widehat{\mathbf{S}}_B \mathbf{w}_{lda} = \lambda_{lda} \mathbf{w}_{lda}, \quad \widehat{\mathbf{S}}_W \mathbf{w}_{lda} = (1 - \lambda_{lda}) \mathbf{w}_{lda} \quad (29)$$

$$\widehat{\mathbf{S}}_N \mathbf{w}_{udp} = \lambda_{udp} \mathbf{w}_{udp}, \quad \widehat{\mathbf{S}}_L \mathbf{w}_{udp} = (1 - \lambda_{udp}) \mathbf{w}_{udp} \quad (30)$$

$$\widehat{\mathbf{S}}_D \mathbf{w}_{lpp} = \lambda_{lpp} \mathbf{w}_{lpp}, \quad \widehat{\mathbf{S}}_{L'} \mathbf{w}_{lpp} = (1 - \lambda_{lpp}) \mathbf{w}_{lpp} \quad (31)$$

This property allows us to find the projections that simultaneously maximize the $\mathbf{w}^\top \mathbf{S}_B \mathbf{w}$ in LDA ($\mathbf{w}^\top \mathbf{S}_N \mathbf{w}$ in UDP and $\mathbf{w}^\top \mathbf{S}_D \mathbf{w}$ in LPP) and minimize the $\mathbf{w}^\top \mathbf{S}_W \mathbf{w}$ in LDA ($\mathbf{w}^\top \mathbf{S}_L \mathbf{w}$ in UDP and $\mathbf{w}^\top \mathbf{S}_{L'} \mathbf{w}$ in LPP), which is equivalent to achieving the optimal projections in LDA, UDP, and LPP without deriving the Rayleigh quotient and solving the generalized eigenvalue problem, thus avoiding inverting a singular matrix.

Visualization of the top 5 projection direction vectors $\mathbf{w}_1, \dots, \mathbf{w}_5$ obtained by traditional linear subspace learning methods as well as the proposed FKDA framework are shown in Fig. 2.

4.1. A fresh look to solving generalized Rayleigh quotient using Fukunaga Koontz transform

The objective functions for finding one single optimal projection direction vector in the aforementioned subspace learning methods are in the form of a generalized Rayleigh quotient as follows:

$$J(\mathbf{w}) = \frac{\mathbf{w}^\top \mathbf{S}_i \mathbf{w}}{\mathbf{w}^\top \mathbf{Z}_i \mathbf{w}} \quad (32)$$

where $(\mathbf{S}_i, \mathbf{Z}_i)$ are positive semi-definite matrices and can take value $(\mathbf{S}_B, \mathbf{S}_W)$ from LDA, $(\mathbf{S}_N, \mathbf{S}_L)$ from UDP, and $(\mathbf{S}_D, \mathbf{S}_{L'})$ from LPP. The traditional way to solve this problem is first taking the

derivative with respect to \mathbf{w} and setting it to 0:

$$\begin{aligned} \frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} &= \frac{2\mathbf{w}^\top \mathbf{Z}_i \mathbf{S}_i \mathbf{w} - 2\mathbf{w}^\top \mathbf{S}_i \mathbf{w} \mathbf{Z}_i \mathbf{w}}{(\mathbf{w}^\top \mathbf{Z}_i \mathbf{w})^2} = 0 \\ &\Rightarrow \frac{(\mathbf{w}^\top \mathbf{Z}_i \mathbf{w}) \mathbf{S}_i \mathbf{w} - (\mathbf{w}^\top \mathbf{S}_i \mathbf{w}) \mathbf{Z}_i \mathbf{w}}{\mathbf{w}^\top \mathbf{Z}_i \mathbf{w}} = 0 \\ &\Rightarrow \mathbf{S}_i \mathbf{w} - J(\mathbf{w}) \mathbf{Z}_i \mathbf{w} = 0 \end{aligned} \quad (33)$$

Therefore, the projection direction vector \mathbf{w} that maximizes $J(\mathbf{w})$ is the eigenvector corresponding to the largest eigenvalue from eigen-decomposition of the matrix $\mathbf{Z}_i^{-1} \mathbf{S}_i$. Note that the \mathbf{w} 's obtained may not be orthogonal because $\mathbf{Z}_i^{-1} \mathbf{S}_i$ may not be positive semidefinite even though both \mathbf{Z}_i^{-1} and \mathbf{S}_i are semidefinite matrices.

Under the FKT framework, we replace the correlation matrices Σ_X and Σ_Y with linearly transformed $\widehat{\mathbf{S}}_i = \mathbf{P}^\top \mathbf{S}_i \mathbf{P}$ and $\widehat{\mathbf{Z}}_i = \mathbf{P}^\top \mathbf{Z}_i \mathbf{P}$ which is traditionally solved in the Rayleigh quotient from Eq. (32), and following the same FKT derivation, a linear subspace whose dominant projection vectors $\widehat{\mathbf{w}}$ maximize the $\widehat{\mathbf{S}}_i$ in Eq. (32) can be obtained as

$$\widehat{\mathbf{S}}_i \widehat{\mathbf{w}} = \widehat{\lambda} \widehat{\mathbf{w}} \quad \text{and} \quad \widehat{\mathbf{Z}}_i \widehat{\mathbf{w}} = (1 - \widehat{\lambda}) \widehat{\mathbf{w}} \quad (34)$$

Pre-multiplied by $\widehat{\mathbf{w}}^\top$, Eq. (34) becomes

$$\widehat{\mathbf{w}}^\top \widehat{\mathbf{S}}_i \widehat{\mathbf{w}} = \widehat{\lambda} \quad \text{and} \quad \widehat{\mathbf{w}}^\top \widehat{\mathbf{Z}}_i \widehat{\mathbf{w}} = 1 - \widehat{\lambda} \quad (35)$$

$$\Rightarrow \frac{\widehat{\mathbf{w}}^\top \widehat{\mathbf{S}}_i \widehat{\mathbf{w}}}{\widehat{\mathbf{w}}^\top \widehat{\mathbf{Z}}_i \widehat{\mathbf{w}}} = \frac{\widehat{\lambda}}{1 - \widehat{\lambda}} \quad (36)$$

Written in matrix form, Eq. (34) becomes

$$\widehat{\mathbf{W}}^\top \widehat{\mathbf{S}}_i \widehat{\mathbf{W}} = \widehat{\Lambda} \quad \text{and} \quad \widehat{\mathbf{W}}^\top \widehat{\mathbf{Z}}_i \widehat{\mathbf{W}} = \mathbf{I} - \widehat{\Lambda}$$

$$\Rightarrow \frac{\text{tr}(\widehat{\mathbf{W}}^\top \widehat{\mathbf{S}}_i \widehat{\mathbf{W}})}{\text{tr}(\widehat{\mathbf{W}}^\top \widehat{\mathbf{Z}}_i \widehat{\mathbf{W}})} = \frac{\text{tr}(\widehat{\Lambda})}{\text{tr}(\mathbf{I} - \widehat{\Lambda})} = \frac{\text{tr}(\widehat{\Lambda})}{\text{tr}(\mathbf{I}) - \text{tr}(\widehat{\Lambda})} = J(\widehat{\mathbf{W}}) \quad (37)$$

where $\widehat{\mathbf{W}}$ is a matrix housing all the $\widehat{\mathbf{w}}_i$. When $\widehat{\Lambda}$ takes the maximum value $\widehat{\lambda}^*$, $\mathbf{I} - \widehat{\Lambda}$ takes the minimum $1 - \widehat{\lambda}^*$, the ratio of $\widehat{\lambda}^*/(1 - \widehat{\lambda}^*)$ is maximized. Moreover, this ratio is monotonically increasing in $(0, 1)$ as depicted in Fig. 3, so the objective function $J(\widehat{\mathbf{w}})$ is maximally achieved by taking $\widehat{\mathbf{w}}$'s corresponding to $\widehat{\lambda}$'s near 1.

It is worth mentioning that the ratio of $\widehat{\lambda}^*/(1 - \widehat{\lambda}^*)$ cannot be infinite. This equivalently means that $\widehat{\lambda}$'s cannot all take value 1. The reason is as follows. According to Eq. (37), when the trace ratio is infinite, $\widehat{\Lambda}$ becomes an identity matrix \mathbf{I} . This implies that the scatter matrix $\widehat{\mathbf{S}}_i$ has all its eigenvalues equal to 1, which further implies that the scatter matrix itself is an identity matrix. However, this is never the case for the aforementioned discriminant methods such as the LDA, UDP, and LPP.

Moreover, the $\widehat{\mathbf{w}}$'s obtained via FKDA formulation are orthogonal, while is not guaranteed through traditional formulation of discriminant analysis.

To summarize, the FKT provides an alternative way of finding a discriminant subspace whose projection direction vectors $\widehat{\mathbf{w}}$ maximize the objective function $J(\widehat{\mathbf{w}})$. Instead of solving a simultaneous maximization and minimization problem in the form of maximizing the ratio of the two, FKDA formulates a *fixed-sum*

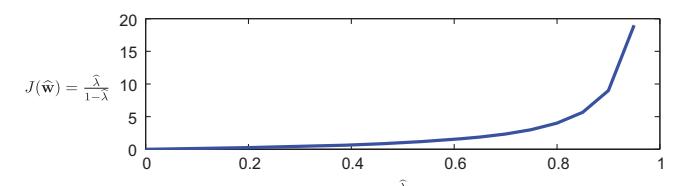


Fig. 3. Plot of objective function $J(\widehat{\mathbf{w}}) = \widehat{\lambda}/(1 - \widehat{\lambda})$ which is monotonically increasing in $(0, 1)$ under FKDA framework.

situation, where the sum of two terms is fixed, and when one term is maximized, the other is simultaneously minimized. Another advantage of FKT is that, no separate PCA step is needed to reduce the dimensionality of the input data. The FKT directly takes \mathbf{S}_i and \mathbf{Z}_i and obtains the optimal $\widehat{\mathbf{W}}$'s without any matrix inversion involved. Reducing dimensionality usually hurts the classification performance because it often discards some dimensions that are discriminative.

4.2. FKDA finds exact solution to discriminant analysis problems

When multi-class discriminant analysis is carried out, the objective function $J(\mathbf{W})$ takes the ratio of the scalar representation of the scatter matrices using trace or determinant. The matrix \mathbf{W} houses all the optimal projection direction vectors \mathbf{w}_i 's at its columns. The following are two possible objective functions:

$$\text{Trace Ratio : } J_1(\mathbf{W}) = \frac{\text{tr}(\mathbf{W}^\top \mathbf{S}_i \mathbf{W})}{\text{tr}(\mathbf{W}^\top \mathbf{Z}_i \mathbf{W})} \quad (38)$$

$$\text{Determinant Ratio : } J_3(\mathbf{W}) = \frac{\det(\mathbf{W}^\top \mathbf{S}_i \mathbf{W})}{\det(\mathbf{W}^\top \mathbf{Z}_i \mathbf{W})} \quad (39)$$

The objective $J_3(\mathbf{W})$ in terms of “determinant ratio” is equivalent to a “ratio determinant” objective $J_4(\mathbf{W})$ as follows:

$$J_3(\mathbf{W}) = \frac{\det(\mathbf{W}^\top \mathbf{S}_i \mathbf{W})}{\det(\mathbf{W}^\top \mathbf{Z}_i \mathbf{W})} = \frac{1}{\det(\mathbf{W}^\top \mathbf{Z}_i \mathbf{W})} \det(\mathbf{W}^\top \mathbf{S}_i \mathbf{W}) \quad (40)$$

$$J_3(\mathbf{W}) = \det[(\mathbf{W}^\top \mathbf{Z}_i \mathbf{W})^{-1}] \det(\mathbf{W}^\top \mathbf{S}_i \mathbf{W}) \quad (41)$$

$$J_3(\mathbf{W}) = \det[(\mathbf{W}^\top \mathbf{Z}_i \mathbf{W})^{-1} (\mathbf{W}^\top \mathbf{S}_i \mathbf{W})] = J_4(\mathbf{W}) \text{ Ratio Determinant} \quad (42)$$

Both $J_3(\mathbf{W})$ and $J_4(\mathbf{W})$ can be solved using a generalized eigenvalue problem formulation. Note that \mathbf{W} is an orthogonal matrix whose determinant is ± 1 .

$$J_3(\mathbf{W}) = J_4(\mathbf{W}) = \det(\mathbf{W}^\top \mathbf{Z}_i^{-1} \mathbf{W} \mathbf{W}^\top \mathbf{S}_i \mathbf{W}) = \det(\mathbf{Z}_i^{-1} \mathbf{S}_i) = \det(\Lambda) \quad (43)$$

The entries λ_i of Λ correspond to the eigenvalues from $\mathbf{Z}_i^{-1} \mathbf{S}_i \mathbf{W}_i = \lambda_i \mathbf{W}_i$.

However, the objective $J_1(\mathbf{W})$ in terms of “trace ratio” does not have a closed form solution. Generally, the “trace ratio” problem is simplified to a more tractable yet inexact one, $J_2(\mathbf{W})$, called the “ratio trace” problem [12,34].

$$J_2(\mathbf{W}) = \text{tr}[(\mathbf{W}^\top \mathbf{Z}_i \mathbf{W})^{-1} (\mathbf{W}^\top \mathbf{S}_i \mathbf{W})] = \text{tr}(\mathbf{Z}_i^{-1} \mathbf{S}_i) = \text{tr}(\Lambda) \quad (44)$$

The \mathbf{W} is also obtainable through the same manner as in the aforementioned $J_3(\mathbf{W})$ and $J_4(\mathbf{W})$.

Among the four objectives, $J_1(\mathbf{W})$ is the ideal objective of discriminant analysis algorithms, and $J_2(\mathbf{W})$, $J_3(\mathbf{W})$, $J_4(\mathbf{W})$ are the inexact approximation that has closed form solutions [12,34]. In the traditional formulation of LDA, UDP, and LPP, the inexact objectives are considered, leading towards the generalized eigenvalue problem for obtaining the optimal projection directions. However, in the proposed FKDA formulation, exact objective $J_1(\widehat{\mathbf{W}})$ can be solved directly with a closed form solution as follows:

$$J_1(\widehat{\mathbf{W}}) = \frac{\text{tr}(\widehat{\mathbf{W}}^\top \widehat{\mathbf{S}}_i \widehat{\mathbf{W}})}{\text{tr}(\widehat{\mathbf{W}}^\top \widehat{\mathbf{Z}}_i \widehat{\mathbf{W}})} = \frac{\text{tr}(\widehat{\mathbf{W}}^\top \mathbf{P}^\top \mathbf{S}_i \mathbf{P} \widehat{\mathbf{W}})}{\text{tr}(\widehat{\mathbf{W}}^\top \mathbf{P}^\top \mathbf{Z}_i \mathbf{P} \widehat{\mathbf{W}})} = \frac{\text{tr}(\widehat{\Lambda})}{\text{tr}(\mathbf{I}) - \text{tr}(\widehat{\Lambda})} \quad (45)$$

Here, when $\mathbf{W} = \widehat{\mathbf{P}} \widehat{\mathbf{W}}$, $J_1(\widehat{\mathbf{W}})$ is identical to $J_1(\mathbf{W})$, while the latter cannot be solved in a closed form.

4.3. Connection to Li's work [13]

Li and Savvides [13] have proposed a FKT subspaces for enhanced face recognition. The traditional FKT is suitable for 2-class problems, and in their work, the FKT is extended to multi-class by applying a class-adaptive sub-eigenspace projection and thresholding. For each class, a one-against-the-rest framework is adopted under FKT to obtain a class-specific subspace, and the normalized distance between reconstruction error and the optimal threshold for each of the FKT subspaces is computed. The classification of the test data is based on 1-nearest-neighbor method with this normalized distance metric.

While in this work, we harness the scatter matrices from discriminant analysis (LDA, UDP, and LPP) directly using the FKT formulation. By doing that, we obtain an optimal subspace for multi-class problems, with many advantages over traditional discriminant analysis. The subspace found by FKDA is orthogonal, while traditional formulation does not guarantee this. Also, the FKDA finds the exact solution to the objectives of discriminant analysis problems by directly solving the “trace ratio” problem, while traditional methods cannot find closed form solution and need to relax to the “ratio trace” problem. Lastly, the FKDA does not require the scatter matrices to be non-singular because no matrix inversion is needed in FKDA. This is also a merit property because discarding dimensions using a separate PCA step may lose important discriminative information.

5. Databases and preprocessing

In this section, we introduce the 6 databases to be used in our extensive experiments. The first one is a large-scale unconstrained face database: FRGC ver 2.0 database, and the second is a database best known for illumination invariant face recognition: YaleB database. These two face databases span the most important spectrum of face recognition tasks. The FRGC ver 2.0 is the largest-scale unconstrained face recognition database publicly available. Images exhibit various facial variations such as expression variations, lighting variations, blur, resolution variations, pose variations, etc. YaleB database, on the other hand, shows the most harsh and extreme lighting conditions for illumination related face recognition tasks. Therefore, benchmarking the proposed FKDA method on these two databases for face recognition is necessary, but may not yet be sufficient. In light of this, we will test our algorithm on four additional databases for more comprehensive understanding of the propose method.

The third database is a face occlusion database: AR Face database. The fourth one is a constrained illumination database: CMU MPIE database. The fifth one is a large-scale mugshot-type database: PCSO database. The last one is an unconstrained celebrity-type face verification database: LFW. The preprocessing and normalization steps are briefly discussed towards the end of this section.

5.1. FRGC ver 2.0 database

The first database we use is a large-scale one. NIST's Face Recognition Grand Challenge (FRGC) ver 2.0 database [35] is collected at the University of Notre Dame. Each subject session consists of controlled and uncontrolled still images. The controlled full frontal facial images were taken under two lighting conditions under studio setting with two facial expressions. The uncontrolled images were taken in various locations such as hallways, atria, or outdoors under varying illumination settings also with two expressions, smiling and neutral.

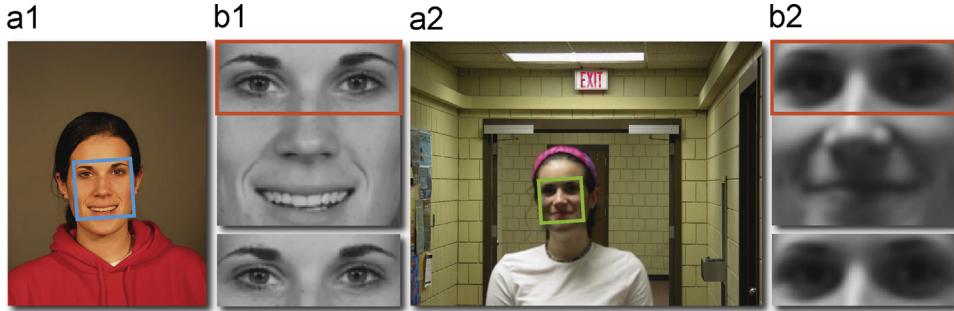


Fig. 4. Examples from the FRGC database: (a1, a2) controlled and uncontrolled still, (b1, b2) cropped full face and periocular region.



Fig. 5. Examples from the YaleB database: facial images under various illumination conditions from one subject.

The FRGC ver 2.0 database has three components: first, the generic *training* set is typically used in the training phase to extract features. It contains both controlled and uncontrolled images of 222 subjects, and a total of 12,776 images. Second, the *target* set represents the people that we want to find. It has 466 different subjects, and a total of 16,028 images. Last, the *probe* set represents the unknown images that we need to match against the *target* set. It contains the same 466 subjects as in target set, with half as many images for each person as in the target set, bringing the total number of probe images to 8014. All the *probe* subjects that we are trying to identify are present in the *target* set. Images from the three sets are mutually exclusive.

The FRGC Experiment 4 is the hardest experiment in the FRGC protocol where face images captured in *controlled* indoor setting are matched against *uncontrolled* outdoor conditions, with harsh lighting conditions that considerably alter the appearance of the face image, as in Fig. 4.

5.2. YaleB and extended YaleB database

The second database we use is the YaleB and Extended YaleB database. This database contains the most harsh illumination conditions and is well known for illumination invariant face recognition. We want to test the performance of the proposed FKDA compared to other subspace learning methods under the scenario of recognizing faces under various lighting changes. The YaleB [36] contains 5760 single light source images of 10 subjects each seen under 576 viewing conditions (9 poses \times 64 illumination conditions). For every subject in a particular pose, an image with ambient (background) illumination was also captured. Hence, the total number of images is in fact $5760 + 90 = 5850$. The extended YaleB [37] contains 16,128 images of 28 human subjects under 9 poses and 64 illumination conditions. Examples are shown in Fig. 5.

We combine both (YaleB and Extended YaleB) databases, with $10 + 28 = 38$ unique subjects in total. We only choose the frontal image with all the illumination variations. So there are 64 images for each subjects. So the entire database on which we conduct the experiments contains $38 \times 64 = 2432$ images.

5.3. AR Face database

The third database we use is the AR Face database [38], which is one of the most widely used face databases with facial occlusions. It contains 3288 color images from 135 subjects (76 male subjects + 59 female subjects). Typical occlusions include scarves

and sunglasses. The database also captures expression variations and lighting changes. This database will be used for our experiment where we deal with faces with occlusions and study whether the proposed FKDA method is advantageous over its competitors under such scenarios. Fig. 6 shows some sample images with occlusions from the AR Face database as well as facial expression variations and lighting changes.

5.4. CMU Multi-PIE database

The fourth database we use is the CMU Multi-PIE [39] face database, which contains more than 750,000 images of 337 people recorded in up to four sessions over the span of five months. Subjects were imaged under 15 view points and 19 illumination conditions (with 1 neutral as well) while displaying a range of facial expressions. In our experiment, we will use frontal images from session 1 only. There are a total of 249 subjects in session 1. Each subject is captured under 20 illumination changes and two facial expressions: neutral and smiling. Therefore the total number of frontal face images in session 1 is 9960.

5.5. Large-scale PCSO Mugshot database

The fifth database we use in our experiments is the Pinellas County Sheriff's Office (PCSO) mugshots database [40] which has over 1.4 million mugshot photographs. The mugshots were acquired in the public domain through Florida's "Sunshine" laws. In this work, we will utilize 100,000 randomly selected faces from the PCSO database. Examples are shown in Fig. 7 with the periocular region masked.

5.6. Labeled Faces in the Wild database

The sixth and last database we use is the Labeled Faces in the Wild (LFW) database [41] which contains a total of 13,233 face images of 5749 different subjects. The whole database is split into 2 views: View 1 is used for model selection and algorithm development which has 2200 pairs for training and 1000 pairs for testing; View 2 consists of 10-fold cross-validation set of 6000 pairs, on which we report the performance only.

5.7. Region of interest normalization

Here we discuss the face alignment details and the cropped image dimensions on various databases. For the FRGC database, rotation and eye coordinate normalization are performed to horizontally align left and right eyes with fixed eye coordinates for all full face frontal images from the target set, probe set and the generic training set. We only use two eye locations for facial alignment because the FRGC images are mostly frontal or near frontal. For the evaluations on the FRGC database, the full face is cropped to the size of 128×128 and the periocular region [42–47,63–67] is cropped to be 50×128 .

For the YaleB database, square crops are already available, and they are of size 64×64 . We further crop the corresponding periocular region to be of size 25×64 .



Fig. 6. Examples from the AR face database: each column shows facial images under various expression, illumination, and occlusion conditions from one subject. A total of 10 subjects are shown.

For the AR Face database, MPIE database, PCSO database, and LFW database, facial images are cropped in the same way as the FRGC database, the cropped full face is of size 128×128 , and 50×128 for the periocular region.

6. Experiments

6.1. Experimental setup overview

The normalized cosine distance (NCD) is adopted to compute the similarity matrix between target and probe images:

$$d(\mathbf{x}, \mathbf{y}) = \frac{-\mathbf{x}^\top \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|}. \quad (46)$$

Compared to other commonly used distance measurement such as ℓ_1 -norm, ℓ_2 -norm, NCD exhibits the best result [48–55,68,69]. The result of each algorithm is a similarity matrix whose entry SimM_{ij} is the NCD between the feature vector of probe image i and target image j . The performance is analyzed using verification rate (VR) at **0.001** (0.1%) false accept rate (FAR), equal error rate (EER), rank-1 identification rate, and the receiver operating characteristic (ROC) curves. For the AR Face evaluation, VRs at **0.01** (1%) FAR are reported instead, because for this challenging occlusion database, VRs at **0.001** (0.1%) FAR would be too low to draw any informative conclusions. See details in the following subsection on the AR Face database evaluation. Experiments that involve matching all the images to themselves do not report the rank-1 identification rate because it will always be 100%. For the LFW database, area under the ROC curve (AUC) value is reported, as set forth by the LFW protocol [56].

6.1.1. Competing algorithms

In the following experiments, we want to investigate whether the proposed FKDA on LDA, UDP and LPP can outperform the traditional Fisher discriminant method [2] on LDA, UDP and LPP. In addition, we will also include several competing algorithms, discussed in Section 2. They are Direct LDA (UDP and LPP) [15], Regularized LDA (UDP and LPP) [16], Penalized LDA (UDP and LPP) [19], Pseudo Fisher LDA (UDP and LPP) [20], and GSVD LDA (UDP and LPP) [24]. They all have various means to avoid the small-sample-size problem. We also use standard PCA as a baseline. Our proposed FKDA (LDA, UDP and LPP) will be benchmarked using these competing algorithms on various databases.

6.1.2. Closed set, open set, and semi-open set recognition

Most pattern recognition tasks in computer vision have taken the form of “closed set” recognition where all testing classes are known during training stage. A more realistic scenario for face recognition applications is “open set” recognition, where incomplete knowledge of the world is present at training time, and unknown classes can be submitted to an algorithm during testing [57]. In this work, we further make distinctions between “open set” and “semi-open set” recognition. In “open set” recognition, none of the testing classes has been seen during training stage. For “semi-open set” recognition, some of the testing classes have been seen during training stage, with additional unknown classes involved in the testing stage. Compared to “closed set” recognition, “open set” and “semi-open set” recognition are of course more challenging because they require the algorithm to be able to generalize well for unseen classes. We have not yet able to model the entire humanity with over 7 billion classes, and therefore, “open set” capability should be desired when developing an algorithm.

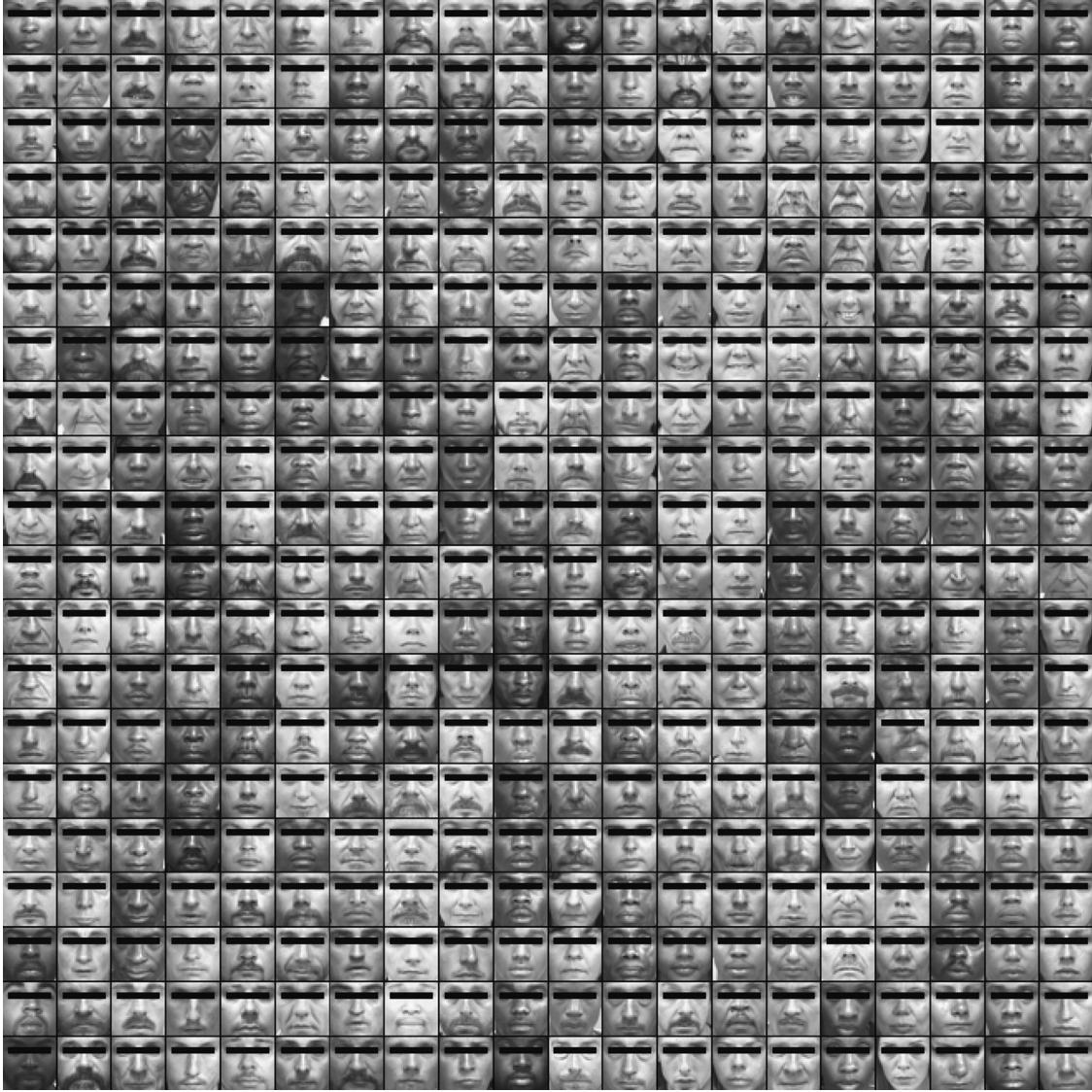


Fig. 7. Example Mugshots from the PCSO database.

6.1.3. Experiments on the FRGC database

We strictly follow NIST's FRGC Experiment 4 protocol which involves 1-to-1 matching of 8014 uncontrolled probe images to 16,028 controlled target images (~ 128 million pair-wise face match comparisons). The similarity matrix is of size $8014 \times 16,028$. Both full face and periocular region are considered. The training set (222 subjects) and the testing set (466 subjects) share 153 common subjects. Therefore, this is a “semi-open set” recognition problem.

6.1.4. Experiments on the YaleB database

We randomly select 32 images per subject for training purpose, and the remaining 32 images for testing in a 1-to-1 matching fashion. Therefore, in the testing stage, there are totally 1216 images from 38 classes. The similarity matrix is of size 1216×1216 . Both full face and periocular region are also considered. This is a “closed set” recognition problem.

6.1.5. Experiments on the AR Face database

We use 2 images per subject for training. The training images are image No. 1 and No. 2 for each subject, with neutral and smiling expressions, and with no occlusions as shown in Fig. 6. The point is to train the classifiers using mugshot-type of clean images and see how

robust it can perform recognition tasks on occluded images, mimicking real-world law-enforcement face matching. Therefore, the number of training images is 270, and it leaves 3018 for testing. The entire testing set is matched against itself, resulting in a similarity matrix of size 3018×3018 , ideally resulting in a block-diagonal matrix. This is also a “closed set” recognition problem.

6.1.6. Experiments on the MPIE database

We carry out an “open set” recognition on the MPIE database where we train the linear subspaces using FRGC training set and test on the MPIE images. The entire MPIE frontal faces from session 1 are matched against themselves, which result in a similarity matrix of size 9960×9960 . The point is to see how robustly a general-purpose face subspace (as compared to leaning an illumination-aware subspace like in the YaleB evaluation) can perform illumination invariant face recognition tasks.

6.1.7. Experiments on the PCSO database

We use 100,000 faces from the PCSO database for our experiments. Directly carrying out experiments on the PCSO database is hard because most of the mugshot subjects have only one image, which makes face verification experiments unsuitable. To account for that, and also fully utilize the large-scale face image database,

we will use PCSO to augment the FRGC database and perform a large-scale face verification experiment. We will conduct FRGC Experiment 1, with augmented faces. The gallery set will be the FRGC target set which contains 16,028 images, plus the 100,000 PCSO images, a total of 116,028 faces. The probe set will just be the FRGC target set. Therefore, the resulting similarity matrix will be of size $16,028 \times 116,028$.

6.1.8. Experiments on the LFW database

We will conduct unsupervised LFW face verification according to its latest protocols [41,56]. The LFW unsupervised paradigm specifies that an algorithm should conform with the following requirements to be considered valid under the unsupervised protocol. First of all, the algorithm should return a scalar-valued function $f(I, J) = d$ of two images I and J , such that increasing d implies a greater distance between the images I and J . Any threshold θ can produce a binary classification (same vs. different). This threshold can then be varied to produce the ROC curve. The protocol further requires that the function $f(\cdot)$ should have no parameters that are set using any information about the LFW class labels of “same” and “different”. Also, $f(\cdot)$ cannot be trained using images labeled with the individual’s names or any unique identifiers because it would be tantamount to training the classifier on “same” and “different” pairs. In our work, the subspace learning algorithms do not capitalize on the identity information. For each available subject in the training portion, of each fold, we pick only one single image and adopt the 3DGEM method [58] for reconstructing multiple faces with slight pose variations to create multiple training images for one subject in order to account for the slight pose variations in the LFW database. The pose range is -10° to $+10^\circ$ in yaw with step of 2° , and -10° to $+10^\circ$ in pitch also with step of 2° . Therefore, the total images for one subject is $11 \times 11 = 121$. We then extract high dimensional LBP features [59–62] from the full face as well as only from the periocular region. The experiments are carried out following the 10-fold cross validation paradigm according to the unsupervised protocol.

6.2. Experimental results

6.2.1. Experiments on the FRGC database

Tables 1 and 2 show the VR at 0.1% FAR, EER, and Rank-1 ID rate for both the full face and periocular region. Figs. 8 and 9 show the ROC curves accordingly. From the results, the proposed FKDA method significantly outperforms traditional Fisher LDA, UDP, and LPP for both cases of full face recognition and periocular region recognition. Also, the proposed method is still superior to other competing methods that try to solve the SSS problem in their own ways. To re-formulate the traditional LDA, UDP, LPP problems in the proposed FKDA framework, high dimensional data can be directly utilized to achieve optimal projection directions that satisfy the corresponding criteria. As can be seen, FKDA on LDA scatter matrices outperforms traditional LDA. Improvements can also be found in UDP and LPP. The main reasons for this are: (1) in the FKDA framework, no separate PCA step is needed to reduce the dimensionality of the data. Dimensionality reduction may lose important discriminative information that can aid the face recognition performance. This stems from the fact that PCA bases are not designed to retain the discriminability of the data. (2) Using FKT framework to solve the generalized Rayleigh quotient problem will achieve projection direction vectors \mathbf{w} that are truly maximizing the objectives in the discriminant analysis, while traditional method may not because the Rayleigh quotient representation is non-convex and non-concave.

Table 1

Performance on the FRGC Experiment 4 evaluation using full face.

Methods	VR at 0.001 FAR	EER	Rank1 ID rate
PCA baseline	0.164	0.209	0.733
Fisher (LDA) [2]	0.168	0.153	0.544
Direct (LDA) [15]	0.277	0.110	0.714
Regularized (LDA) [16]	0.187	0.129	0.612
Penalized (LDA) [19]	0.240	0.115	0.676
Pseudo Fisher (LDA) [20]	0.280	0.115	0.705
GSVD (LDA) [24]	0.275	0.106	0.707
FKDA (LDA)	0.489	0.051	0.954
Fisher (UDP) [2]	0.201	0.139	0.610
Direct (UDP) [15]	0.287	0.111	0.740
Regularized (UDP) [16]	0.304	0.103	0.747
Penalized (UDP) [19]	0.275	0.105	0.733
Pseudo Fisher (UDP) [20]	0.287	0.106	0.743
GSVD (UDP) [24]	0.319	0.089	0.806
FKDA (UDP)	0.322	0.114	0.793
Fisher (LPP) [2]	0.162	0.150	0.532
Direct (LPP) [15]	0.268	0.109	0.715
Regularized (LPP) [16]	0.243	0.114	0.676
Penalized (LPP) [19]	0.258	0.115	0.721
Pseudo Fisher (LPP) [20]	0.274	0.105	0.725
GSVD (LPP) [24]	0.238	0.120	0.691
FKDA (LPP)	0.520	0.061	0.926

Table 2

Performance on the FRGC Experiment 4 evaluation using periocular region.

Methods	VR at 0.001 FAR	EER	Rank1 ID rate
PCA baseline	0.169	0.235	0.661
Fisher (LDA) [2]	0.134	0.185	0.450
Direct (LDA) [15]	0.225	0.148	0.626
Regularized (LDA) [16]	0.171	0.161	0.542
Penalized (LDA) [19]	0.198	0.146	0.588
Pseudo Fisher (LDA) [20]	0.227	0.137	0.637
GSVD (LDA) [24]	0.231	0.132	0.646
FKDA (LDA)	0.578	0.057	0.921
Fisher (UDP) [2]	0.174	0.179	0.517
Direct (UDP) [15]	0.251	0.135	0.642
Regularized (UDP) [16]	0.264	0.132	0.679
Penalized (UDP) [19]	0.234	0.137	0.658
Pseudo Fisher (UDP) [20]	0.250	0.135	0.644
GSVD (UDP) [24]	0.307	0.111	0.731
FKDA (UDP)	0.338	0.110	0.817
Fisher (LPP) [2]	0.133	0.190	0.439
Direct (LPP) [15]	0.229	0.133	0.634
Regularized (LPP) [16]	0.226	0.144	0.609
Penalized (LPP) [19]	0.225	0.142	0.634
Pseudo Fisher (LPP) [20]	0.237	0.136	0.645
GSVD (LPP) [24]	0.192	0.148	0.592
FKDA (LPP)	0.568	0.061	0.928

6.2.2. Experiments on the YaleB database

Tables 3 and 4 show the VR at 0.1% FAR and EER for YaleB database evaluation on both full face and periocular region. Figs. 10 and 11 show the ROC curves for both cases. From the results, the proposed FKDA method on LDA significantly outperforms the competing algorithms. The proposed FKDA on UDP is among the top performing ones, and the FKDA on LPP beats others as well. From the ROC curves, the competing methods spread, indicating various algorithm learns illumination invariant subspaces quite differently. Very similar performance is

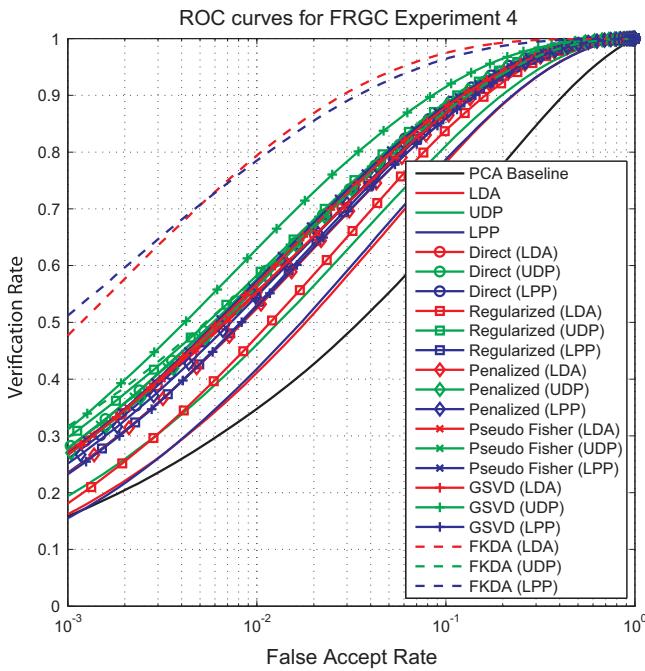


Fig. 8. ROC curves for the FRGC Experiment 4 evaluation using full face.

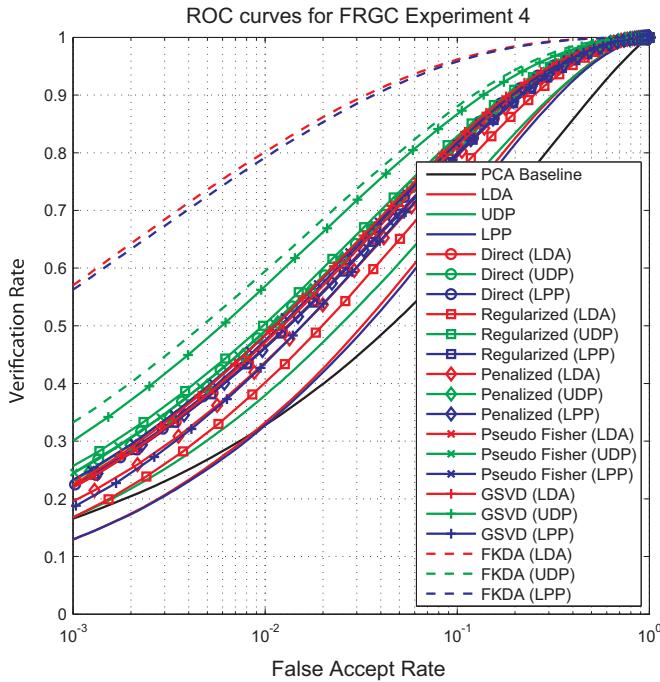


Fig. 9. ROC curves for the FRGC Experiment 4 evaluation using periocular region.

achieved by the periocular region, as compared to the full face evaluation.

6.2.3. Experiments on the AR Face database

Tables 5 and 6 show the VR at 1% FAR and EER for the AR Face database evaluation on both full face and periocular region. Figs. 12 and 13 show the ROC curves for the two cases respectively. The reason that we report VRs at 1% FAR rather than at 0.1% FAR as previously adopted is two-fold: (1) the number of images in AR Face database is more suitable for studying verification rate 1% false acceptance, and (2) we have set up the experiments to match all the images to themselves which involves matching heavily occluded faces to normal faces. This, of course, will lead to worse performance.

Table 3
Performance on the YaleB evaluation using full face.

Methods	VR at 0.001 FAR	EER
PCA baseline	0.107	0.379
Fisher (LDA) [2]	0.831	0.042
Direct (LDA) [15]	0.313	0.153
Regularized (LDA) [16]	0.167	0.216
Penalized (LDA) [19]	0.274	0.089
Pseudo Fisher (LDA) [20]	0.260	0.173
GSVD (LDA) [24]	0.283	0.101
FKDA (LDA)	0.957	0.011
Fisher (UDP) [2]	0.172	0.375
Direct (UDP) [15]	0.620	0.069
Regularized (UDP) [16]	0.169	0.218
Penalized (UDP) [19]	0.382	0.067
Pseudo Fisher (UDP) [20]	0.622	0.063
GSVD (UDP) [24]	0.475	0.059
FKDA (UDP)	0.430	0.052
Fisher (LPP) [2]	0.196	0.379
Direct (LPP) [15]	0.307	0.141
Regularized (LPP) [16]	0.201	0.199
Penalized (LPP) [19]	0.371	0.070
Pseudo Fisher (LPP) [20]	0.339	0.130
GSVD (LPP) [24]	0.379	0.065
FKDA (LPP)	0.409	0.073

Table 4
Performance on the YaleB evaluation using periocular region.

Methods	VR at 0.001 FAR	EER
PCA baseline	0.176	0.327
Fisher (LDA) [2]	0.779	0.044
Direct (LDA) [15]	0.251	0.181
Regularized (LDA) [16]	0.135	0.245
Penalized (LDA) [19]	0.220	0.097
Pseudo Fisher (LDA) [20]	0.221	0.206
GSVD (LDA) [24]	0.204	0.093
FKDA (LDA)	0.940	0.012
Fisher (UDP) [2]	0.176	0.377
Direct (UDP) [15]	0.540	0.088
Regularized (UDP) [16]	0.146	0.245
Penalized (UDP) [19]	0.325	0.087
Pseudo Fisher (UDP) [20]	0.543	0.088
GSVD (UDP) [24]	0.411	0.057
FKDA (UDP)	0.466	0.064
Fisher (LPP) [2]	0.192	0.365
Direct (LPP) [15]	0.257	0.173
Regularized (LPP) [16]	0.165	0.223
Penalized (LPP) [19]	0.275	0.061
Pseudo Fisher (LPP) [20]	0.278	0.156
GSVD (LPP) [24]	0.322	0.074
FKDA (LPP)	0.402	0.044

Through the experiments, we have found that the VRs at 0.1% FAR are far too low to be able to draw any informative conclusions about various algorithms. Therefore, we report VRs at 1% FAR only. From the results between full face and the periocular region, we have not seen an obvious gain or drop, this is because there are two types of facial occlusions in the AR Face database: sunglasses and scarves. It is apparent that periocular region would be advantageous when matching occluded faces with scarves because the eye region is not affected by this type of occlusion at all. However, the periocular region would become less advantageous when the subjects are wearing

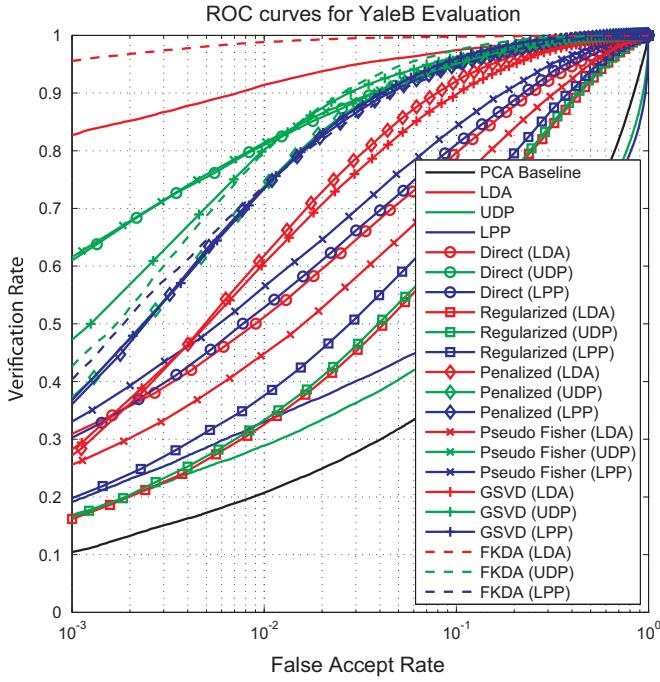


Fig. 10. ROC curves for the YaleB evaluation using full face.

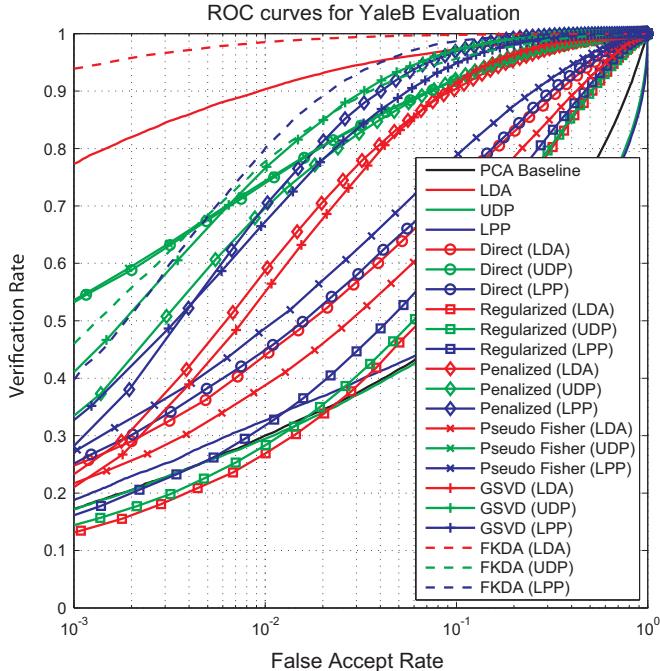


Fig. 11. ROC curves for the YaleB evaluation using periocular region.

sunglasses. Given equal number of occluded images with sunglasses and scarves, the performance of the periocular region should be quite similar to that of the full face evaluation. We have conducted an additional experiments to show the effectiveness of the periocular region when it comes to scarf-type facial occlusions. We remove all the images with sunglasses and carry out the same experiments. The results are tabulated in Tables 7 and the ROC curves are shown in Fig. 14. As can be observed, the performance of the periocular region improves quite a lot over the performance of the full face, as expected. In terms of the proposed FKDA on LDA, UDP and LPP, they have consistently outperformed other competing algorithms.

Table 5

Performance on the AR Face evaluation using full face.

Methods	VR at 0.01 FAR	EER
PCA baseline	0.058	0.446
Fisher (LDA) [2]	0.103	0.382
Direct (LDA) [15]	0.105	0.386
Regularized (LDA) [16]	0.095	0.395
Penalized (LDA) [19]	0.106	0.386
Pseudo Fisher (LDA) [20]	0.085	0.391
GSVD (LDA) [24]	0.087	0.393
FKDA (LDA)	0.174	0.328
Fisher (UDP) [2]	0.159	0.345
Direct (UDP) [15]	0.095	0.381
Regularized (UDP) [16]	0.120	0.375
Penalized (UDP) [19]	0.146	0.355
Pseudo Fisher (UDP) [20]	0.140	0.363
GSVD (UDP) [24]	0.140	0.355
FKDA (UDP)	0.292	0.306
Fisher (LPP) [2]	0.130	0.368
Direct (LPP) [15]	0.137	0.356
Regularized (LPP) [16]	0.112	0.380
Penalized (LPP) [19]	0.131	0.369
Pseudo Fisher (LPP) [20]	0.097	0.383
GSVD (LPP) [24]	0.106	0.388
FKDA (LPP)	0.204	0.315

Table 6

Performance on the AR Face evaluation using periocular region.

Methods	VR at 0.01 FAR	EER
PCA baseline	0.062	0.442
Fisher (LDA) [2]	0.108	0.391
Direct (LDA) [15]	0.103	0.401
Regularized (LDA) [16]	0.091	0.407
Penalized (LDA) [19]	0.118	0.398
Pseudo Fisher (LDA) [20]	0.094	0.403
GSVD (LDA) [24]	0.086	0.405
FKDA (LDA)	0.163	0.344
Fisher (UDP) [2]	0.148	0.363
Direct (UDP) [15]	0.092	0.393
Regularized (UDP) [16]	0.121	0.384
Penalized (UDP) [19]	0.141	0.372
Pseudo Fisher (UDP) [20]	0.134	0.376
GSVD (UDP) [24]	0.138	0.369
FKDA (UDP)	0.279	0.318
Fisher (LPP) [2]	0.127	0.384
Direct (LPP) [15]	0.132	0.375
Regularized (LPP) [16]	0.112	0.395
Penalized (LPP) [19]	0.130	0.378
Pseudo Fisher (LPP) [20]	0.093	0.397
GSVD (LPP) [24]	0.107	0.400
FKDA (LPP)	0.201	0.327

6.2.4. Experiments on the MPIE database

Tables 8 and 9 show the VR at 0.1% FAR and EER for the MPIE database evaluation on both full face and periocular region. Figs. 15 and 16 show the ROC curves for the two cases respectively. This is an “open set” evaluation, and it can be seen from the ROC curves that the competing algorithms are quite close to each other. However, our propose FKDA method on LDA, UDP, and LPP still enjoy a slight margin over competing ones. The findings are consistent from full face evaluation to the periocular one.

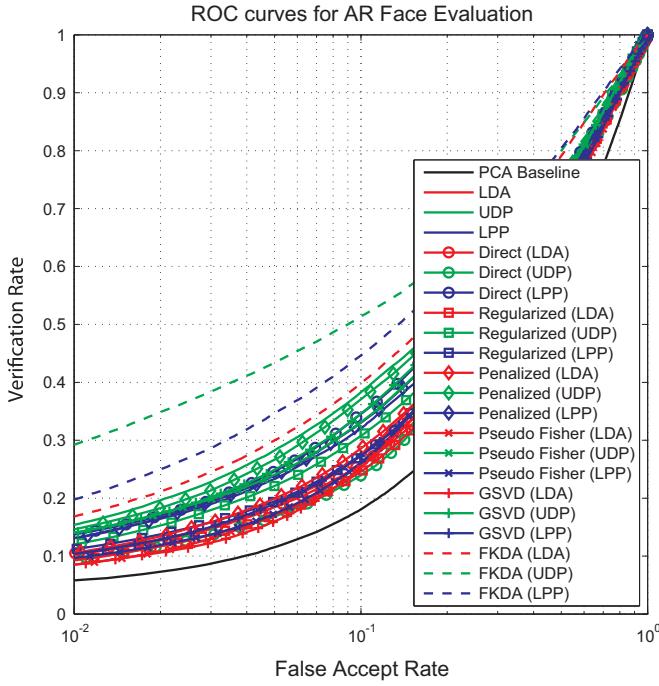


Fig. 12. ROC curves for the AR Face evaluation using full face.

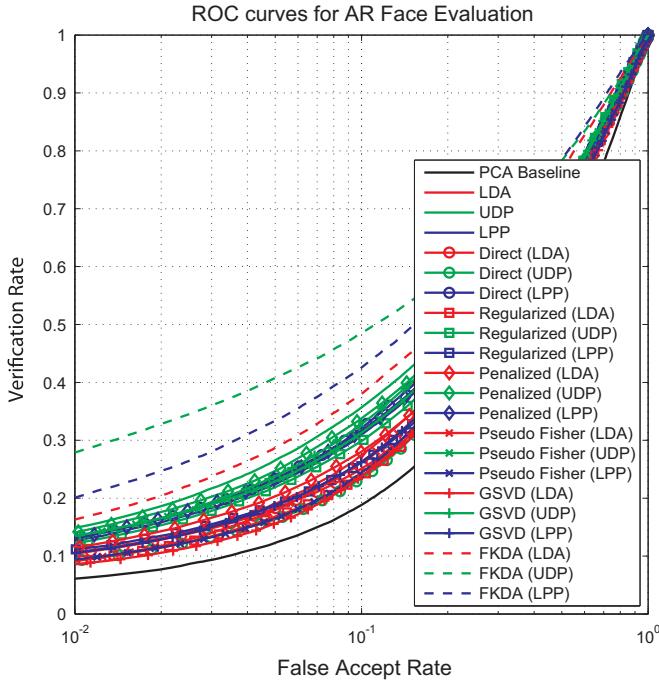


Fig. 13. ROC curves for the AR Face evaluation using periocular region.

6.2.5. Experiments on the PCSO database

This is the largest-scale face verification experiments conducted in this work which involve over 1.8 billion (16k-by-116k) face match comparisons. Tables 10 and 11 show the VR at 0.1% FAR and EER for the FRGC Experiment 1 evaluation augmented with PCSO database on both full face and periocular region. Figs. 17 and 18 show the ROC curves for the full face and the periocular cases respectively. The proposed FKDA methods are still at the top of the chart. Consistent findings are also obtained for the periocular case.

Table 7

Performance on the AR Face evaluation using periocular region on images without sunglasses.

Methods	VR at 0.01 FAR	EER
PCA baseline	0.144	0.392
Fisher (LDA) [2]	0.235	0.181
Direct (LDA) [15]	0.218	0.211
Regularized (LDA) [16]	0.202	0.208
Penalized (LDA) [19]	0.276	0.171
Pseudo Fisher (LDA) [20]	0.202	0.184
GSVD (LDA) [24]	0.170	0.191
FKDA (LDA)	0.378	0.141
Fisher (UDP) [2]	0.361	0.151
Direct (UDP) [15]	0.209	0.215
Regularized (UDP) [16]	0.302	0.170
Penalized (UDP) [19]	0.341	0.156
Pseudo Fisher (UDP) [20]	0.323	0.163
GSVD (UDP) [24]	0.292	0.161
FKDA (UDP)	0.623	0.108
Fisher (LPP) [2]	0.294	0.169
Direct (LPP) [15]	0.301	0.169
Regularized (LPP) [16]	0.257	0.192
Penalized (LPP) [19]	0.320	0.168
Pseudo Fisher (LPP) [20]	0.184	0.192
GSVD (LPP) [24]	0.238	0.192
FKDA (LPP)	0.479	0.125

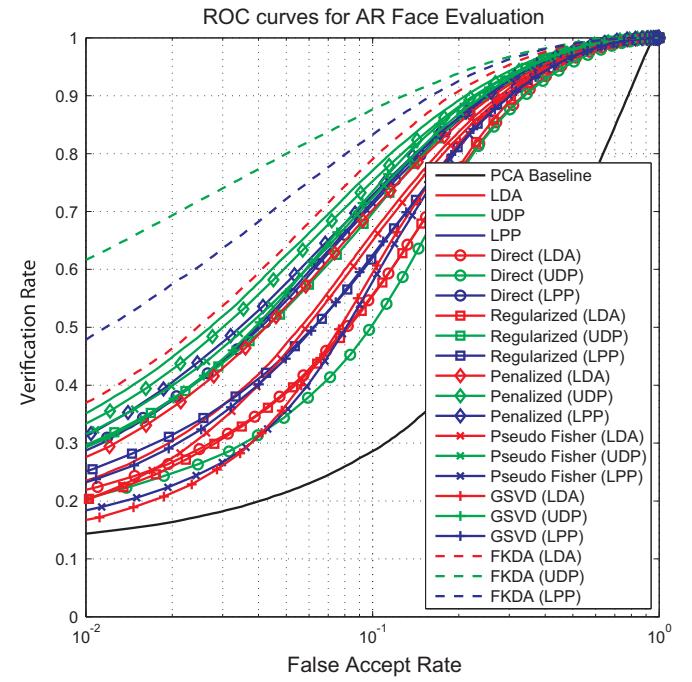


Fig. 14. ROC curves for the AR Face evaluation using periocular region on images without sunglasses.

6.2.6. Experiments on the LFW database

Finally, we follow the LFW unsupervised paradigm. Tables 12 and 13 show the AUC for the 10-fold LFW evaluation on both full face and periocular region. Figs. 19 and 20 show the ROC curves for the full face and the periocular cases respectively. Our proposed FKDA has solidly outperformed the competing algorithms.

Table 8

Performance on the MPIE evaluation using full face.

Methods	VR at 0.001 FAR	EER
PCA baseline	0.558	0.135
Fisher (LDA) [2]	0.616	0.104
Direct (LDA) [15]	0.657	0.089
Regularized (LDA) [16]	0.657	0.090
Penalized (LDA) [19]	0.604	0.111
Pseudo Fisher (LDA) [20]	0.614	0.107
GSVD (LDA) [24]	0.602	0.115
FKDA (LDA)	0.685	0.084
Fisher (UDP) [2]	0.655	0.094
Direct (UDP) [15]	0.651	0.097
Regularized (UDP) [16]	0.659	0.081
Penalized (UDP) [19]	0.667	0.090
Pseudo Fisher (UDP) [20]	0.655	0.090
GSVD (UDP) [24]	0.705	0.070
FKDA (UDP)	0.743	0.070
Fisher (LPP) [2]	0.608	0.099
Direct (LPP) [15]	0.683	0.083
Regularized (LPP) [16]	0.679	0.080
Penalized (LPP) [19]	0.635	0.091
Pseudo Fisher (LPP) [20]	0.627	0.087
GSVD (LPP) [24]	0.629	0.108
FKDA (LPP)	0.717	0.068

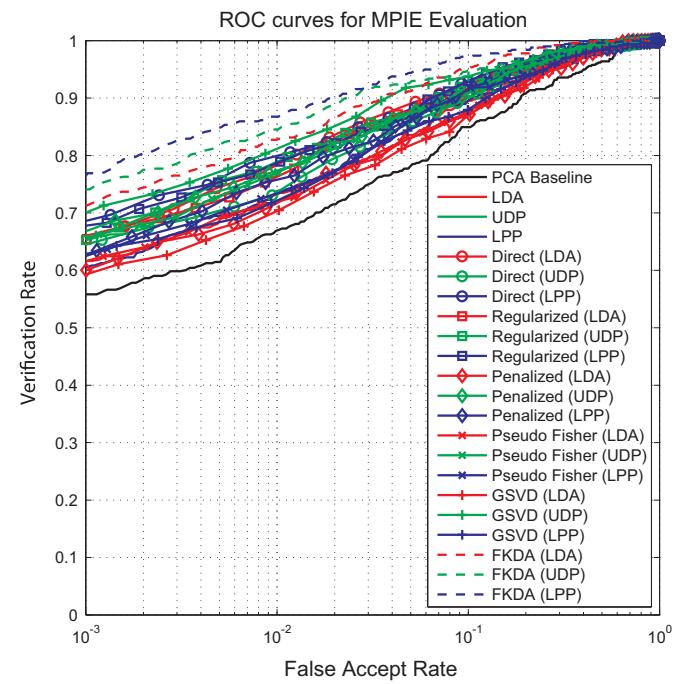
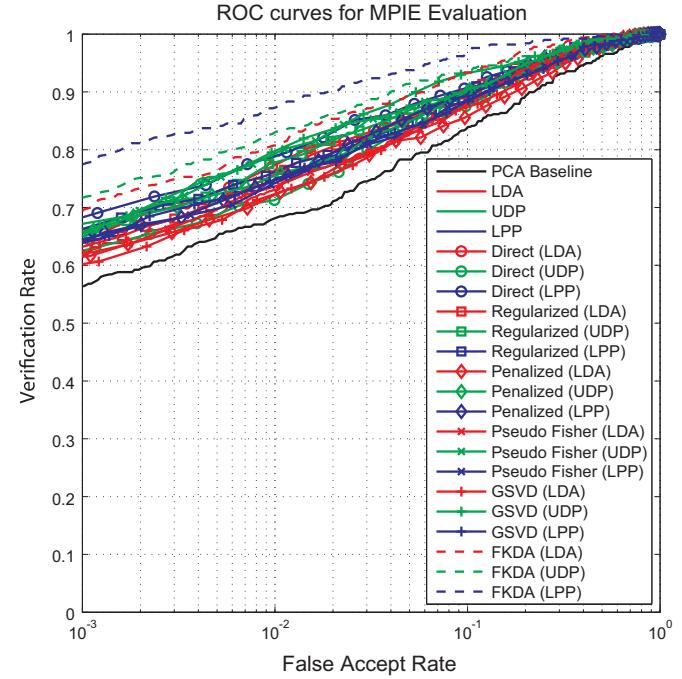
Table 9

Performance on the MPIE evaluation using periocular region.

Methods	VR at 0.001 FAR	EER
PCA baseline	0.568	0.139
Fisher (LDA) [2]	0.624	0.106
Direct (LDA) [15]	0.645	0.099
Regularized (LDA) [16]	0.635	0.097
Penalized (LDA) [19]	0.614	0.124
Pseudo Fisher (LDA) [20]	0.620	0.107
GSVD (LDA) [24]	0.600	0.113
FKDA (LDA)	0.693	0.076
Fisher (UDP) [2]	0.663	0.097
Direct (UDP) [15]	0.631	0.116
Regularized (UDP) [16]	0.655	0.103
Penalized (UDP) [19]	0.649	0.098
Pseudo Fisher (UDP) [20]	0.653	0.094
GSVD (UDP) [24]	0.675	0.081
FKDA (UDP)	0.737	0.066
Fisher (LPP) [2]	0.641	0.107
Direct (LPP) [15]	0.687	0.094
Regularized (LPP) [16]	0.661	0.095
Penalized (LPP) [19]	0.641	0.100
Pseudo Fisher (LPP) [20]	0.649	0.104
GSVD (LPP) [24]	0.629	0.108
FKDA (LPP)	0.733	0.079

6.3. Discussion

From the experimental results, the FKDA formulation on LDA, UDP, and LPP enjoys better recognition performance than traditional Fisher discriminant methods. There are multiple factors that contribute to the superiority of FKDA. First of all, in FKDA formulation, the matrix inversion of scatter matrices is no longer needed because we formulate the objective function in a “fixed-sum” fashion instead of maximizing the ratio of the two. Thanks to this, the separate dimensionality reduction step using PCA in order

**Fig. 15.** ROC curves for the MPIE evaluation using full face.**Fig. 16.** ROC curves for the MPIE evaluation using periocular region.

to bring the scatter matrices full rank is no longer needed. Such step may discard important discriminant information.

Second, the optimal projection direction vectors from FKDA are orthogonal to each other, while in the traditional methods, eigenvectors of $\mathbf{Z}_i^{-1}\mathbf{S}_i$ are not guaranteed to be orthogonal. The orthogonality gives rise to the uncorrelatedness of the subspace vectors and through our extensive experiments, we find that orthogonality helps improve the classification performance. However, such conclusion is based on observation only, and there has not been a proven linkage between the classification performance and the orthogonality of the subspace. In fact, whether orthogonality should be preferred or not is very much data-

Table 10

Performance on the PCSO evaluation using full face.

Methods	VR at 0.001 FAR	EER
PCA baseline	0.139	0.280
Fisher (LDA) [2]	0.370	0.121
Direct (LDA) [15]	0.371	0.123
Regularized (LDA) [16]	0.285	0.142
Penalized (LDA) [19]	0.333	0.138
Pseudo Fisher (LDA) [20]	0.366	0.134
GSVD (LDA) [24]	0.351	0.124
FKDA (LDA)	0.395	0.097
Fisher (UDP) [2]	0.435	0.100
Direct (UDP) [15]	0.404	0.113
Regularized (UDP) [16]	0.385	0.124
Penalized (UDP) [19]	0.380	0.119
Pseudo Fisher (UDP) [20]	0.404	0.117
GSVD (UDP) [24]	0.438	0.100
FKDA (UDP)	0.530	0.102
Fisher (LPP) [2]	0.388	0.113
Direct (LPP) [15]	0.361	0.129
Regularized (LPP) [16]	0.355	0.124
Penalized (LPP) [19]	0.365	0.119
Pseudo Fisher (LPP) [20]	0.361	0.122
GSVD (LPP) [24]	0.357	0.124
FKDA (LPP)	0.459	0.097

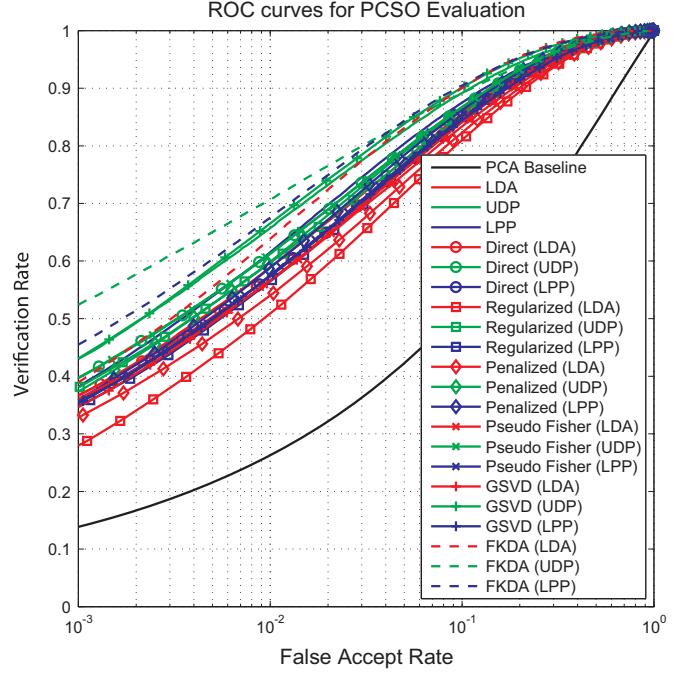
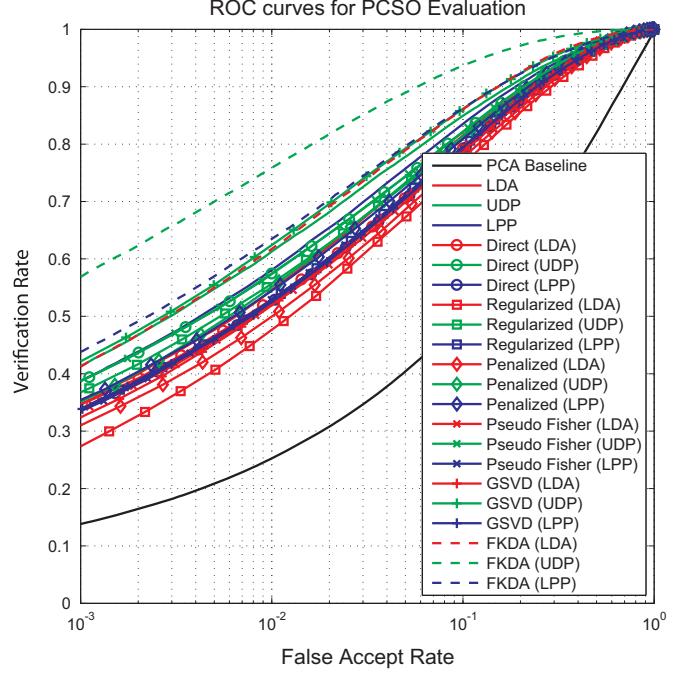
Table 11

Performance on the PCSO evaluation using periocular region.

Methods	VR at 0.001 FAR	EER
PCA baseline	0.140	0.299
Fisher (LDA) [2]	0.351	0.144
Direct (LDA) [15]	0.349	0.154
Regularized (LDA) [16]	0.280	0.164
Penalized (LDA) [19]	0.314	0.159
Pseudo Fisher (LDA) [20]	0.339	0.154
GSVD (LDA) [24]	0.330	0.148
FKDA (LDA)	0.420	0.121
Fisher (UDP) [2]	0.417	0.126
Direct (UDP) [15]	0.390	0.136
Regularized (UDP) [16]	0.370	0.145
Penalized (UDP) [19]	0.357	0.143
Pseudo Fisher (UDP) [20]	0.390	0.137
GSVD (UDP) [24]	0.428	0.117
FKDA (UDP)	0.575	0.078
Fisher (LPP) [2]	0.391	0.131
Direct (LPP) [15]	0.342	0.156
Regularized (LPP) [16]	0.342	0.148
Penalized (LPP) [19]	0.358	0.145
Pseudo Fisher (LPP) [20]	0.339	0.153
GSVD (LPP) [24]	0.342	0.151
FKDA (LPP)	0.444	0.121

dependent and the relationship should be carefully studied, which is beyond the scope of this paper. We want to make clear that the orthogonality in this context is a by-product of the proposed FKDA method and may or may not, in general, aid classification tasks.

Third, the FKDA provides an exact solution to the objective function in the form of “trace ratio”, which is ideal for the discriminant analysis problems. Traditional methods cannot find a closed form solution to this problem and resort to a simpler and inexact alternative in the form of “ratio trace”. The latter can be solved using generalized eigenvalue problem as shown in the LDA,

**Fig. 17.** ROC curves for the PCSO evaluation using full face.**Fig. 18.** ROC curves for the PCSO evaluation using periocular region.

UDP, and LPP formulation. Being able to find exact solution using FKDA also brings merits to the proposed method.

Moreover, when competing with other methods that also try to solve the SSS problem such as Direct LDA (UDP and LPP) [15], Regularized LDA (UDP and LPP) [16], Penalized LDA (UDP and LPP) [19], Pseudo Fisher LDA (UDP and LPP) [20], and GSVD LDA (UDP and LPP) [24], the proposed FKDA LDA (UDP and LPP) method still perform better through extensive experiments. This not only shows that we have succeeded in developing an alternative for solving the SSS problem in the discriminant analysis, but also demonstrates the effectiveness and superiority of the proposed method.

Table 12
Performance on the LFW evaluation using full face.

Methods	AUC
PCA baseline	0.7388
Fisher (LDA) [2]	0.7941
Direct (LDA) [15]	0.8571
Regularized (LDA) [16]	0.8441
Penalized (LDA) [19]	0.8429
Pseudo Fisher (LDA) [20]	0.8849
GSVD (LDA) [24]	0.8661
FKDA (LDA)	0.9152
Fisher (UDP) [2]	0.7970
Direct (UDP) [15]	0.8605
Regularized (UDP) [16]	0.8535
Penalized (UDP) [19]	0.8582
Pseudo Fisher (UDP) [20]	0.7792
GSVD (UDP) [24]	0.8957
FKDA (UDP)	0.9270
Fisher (LPP) [2]	0.7934
Direct (LPP) [15]	0.8614
Regularized (LPP) [16]	0.8512
Penalized (LPP) [19]	0.8538
Pseudo Fisher (LPP) [20]	0.8987
GSVD (LPP) [24]	0.8555
FKDA (LPP)	0.9171

Table 13
Performance on the LFW evaluation using periocular region.

Methods	AUC
PCA baseline	0.6558
Fisher (LDA) [2]	0.7117
Direct (LDA) [15]	0.7893
Regularized (LDA) [16]	0.7673
Penalized (LDA) [19]	0.7776
Pseudo Fisher (LDA) [20]	0.8173
GSVD (LDA) [24]	0.8097
FKDA (LDA)	0.8820
Fisher (UDP) [2]	0.7034
Direct (UDP) [15]	0.7926
Regularized (UDP) [16]	0.7917
Penalized (UDP) [19]	0.7916
Pseudo Fisher (UDP) [20]	0.7580
GSVD (UDP) [24]	0.8366
FKDA (UDP)	0.8884
Fisher (LPP) [2]	0.7083
Direct (LPP) [15]	0.7933
Regularized (LPP) [16]	0.7794
Penalized (LPP) [19]	0.7788
Pseudo Fisher (LPP) [20]	0.8402
GSVD (LPP) [24]	0.7891
FKDA (LPP)	0.8769

From all the ROC curves plotted, we can clearly see that the proposed FKDA formulation on three discriminant analysis problems does improve the traditional methods by a remarkable margin for unconstrained face recognition, large-scale face recognition, face recognition with occlusions, and illumination invariant face recognition problems, under “closed set”, “semi-open set”, and “open set” recognition scenarios. This paper focuses on the improvement we have achieved over traditional ones, even though only raw pixel (or LBP, only for the LFW experiments)

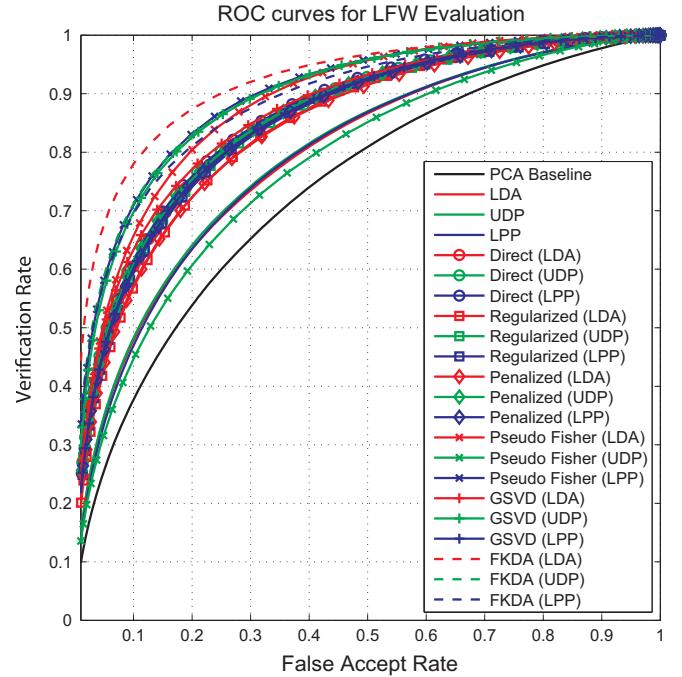


Fig. 19. ROC curves for the LFW evaluation using full face.

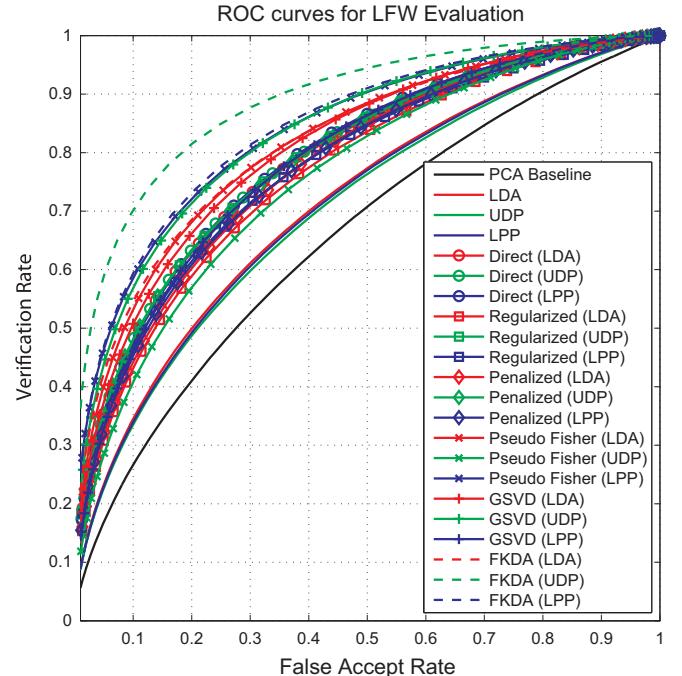


Fig. 20. ROC curves for the LFW evaluation using periocular region.

features are used. The readers are encouraged to apply this method in tandem with many more features and on many other classification problems. We hope that our work can bring value to the pattern recognition community by providing alternative and better formulation of many well-known discriminant analysis problems. In future work, we will extend the FKDA method to the kernel feature space.

7. Conclusion

In this work, we have proposed a new idea which we named multi-class Fukunaga Koontz discriminant analysis (FKDA) by

incorporating the Fukunaga Koontz transform within the optimization for maximizing class separation criteria in LDA, UDP, and LPP. We have proposed an alternative way of maximizing the class separation criteria in LDA, UDP, and LPP without deriving the generalized Rayleigh quotient. We show that the optimization in LDA, UDP, and LPP in ratio form can be equivalently replaced by the proposed fixed-sum form under the FKDA framework. Such fixed-sum form in the proposed FKDA framework does not require any scatter matrices to be non-singular since no matrix inversion is required. That is why, in contrast to traditional Fisher LDA, UDP, and LPP, our approach can work with very high dimensional data as input, without requiring a separate dimensionality reduction step to make the scatter matrices full rank. In addition, the FKDA formulation seeks optimal projection direction vectors that are orthogonal while traditional methods may not guarantee, and it has the capability of finding the exact solutions to the “trace ratio” objective in discriminant analysis problems while traditional methods can only deal with a relaxed and inexact “ratio trace” objective. We have reported on six face databases, in the context of large scale unconstrained face recognition, face recognition with occlusions, and illumination invariant face recognition, under “closed set”, “semi-open set”, and “open set” recognition scenarios, that our proposed FKDA remarkably outperforms traditional discriminant analysis methods as well as five other competing algorithms.

Conflict of interest

There is no conflict of interests.

References

- [1] M.A. Turk, A.P. Pentland, Face recognition using eigenfaces, in: IEEE Conference on Computer Vision and Pattern Recognition, 1991, pp. 586–591.
- [2] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, *IEEE Trans. Pattern Anal. Mach. Intell.* 19 (7) (1997) 711–720.
- [3] J. Yang, D. Zhang, Z. Jin, J.-Y. Yang, Unsupervised discriminant projection analysis for feature extraction, in: International Conference on Pattern Recognition, vol. 1, 2006, pp. 904–907.
- [4] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, *Science* 22 (2000) 2319–2323.
- [5] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, *Science* 22 (2000) 2323–2326.
- [6] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, in: Proceedings of Neural Information Processing Systems, 2001.
- [7] X. He, P. Niyogi, Locality preserving projections, in: Proceedings of Neural Information Processing Systems, 2003.
- [8] A. Hadid, M. Pietikäinen, Demographic classification from face videos using manifold learning, *Neurocomputing* 100 (2013) 197–205.
- [9] O. Arandjelović, R. Cipolla, Achieving robust face recognition from video by combining a weak photometric model and a learnt generic face invariant, *Pattern Recognit.* 46 (1) (2013) 9–23.
- [10] H.T. Ho, R. Gopalan, Model-driven domain adaptation on product manifolds for unconstrained face recognition, *Int. J. Comput. Vis.* 109 (1–2) (2014) 110–125.
- [11] O. Arandjelovic, Gradient edge map features for frontal face recognition under extreme illumination changes, in: BMVC 2012—Proceedings of the British Machine Vision Association Conference, BMVA Press, 2012, pp. 1–11.
- [12] K. Fukunaga, Introduction to Statistical Pattern Recognition, 1990.
- [13] Y. hui Li, M. Savvides, Kernel Fukunaga-Koontz transform subspaces for enhanced face recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2007.
- [14] D.L. Swets, J. Weng, Using discriminant eigenfeatures for image retrieval, *IEEE Trans. Pattern Anal. Mach. Intell.* 18 (8) (1996) 831–836.
- [15] L.F. Chen, H.Y.M. Liao, M.T. Ko, J.C. Lin, G.J. Yu, A new LDA-based face recognition system which can solve the small sample size problem, *Pattern Recognit.* 33 (2000) 1713–1726.
- [16] J.H. Friedman, Regularized discriminant analysis, *J. Am. Stat. Assoc.* 84 (405) (1989) 165–175.
- [17] D.Q. Dai, P.C. Yuen, Regularized discriminant analysis and its application to face recognition, *Pattern Recognit.* 36 (2003) 845–847.
- [18] W.J. Krzanowski, P. Jonathan, W.V. McCarthy, M.R. Thomas, Discriminant analysis with singular covariance matrices: methods and applications to spectroscopic data, *Appl. Stat.* 44 (1995) 101–115.
- [19] T. Hastie, A. Buja, R. Tibshirani, Penalized discriminant analysis, *Ann. Stat.* 23 (1995) 73–102.
- [20] S. Raudys, R.P.W. Duin, On expected classification error of the fisher linear classifier with pseudo-inverse covariance matrix, *Pattern Recognit. Lett.* 19 (5–6) (1998) 385–392.
- [21] M. Skurichina, R.P.W. Duin, Stabilizing classifiers for very small sample size, in: Proceedings of the International Conference on Pattern Recognition, 1996, pp. 891–896.
- [22] G.H. Golub, C.F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, Maryland, USA, 1996.
- [23] P. Howland, M. Jeon, H. Park, Structure preserving dimension reduction for clustered text data based on the generalized singular value decomposition, *SIAM J. Matrix Anal. Appl.* 25 (1) (2003) 165–179.
- [24] J. Ye, R. Janardan, C.H. Park, H. Park, An optimization criterion for generalized discriminant analysis on undersampled problems, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (8) (2004) 982–994.
- [25] R. Huang, Q. Liu, H. Lu, S. Ma, Solving the small sample size problem of LDA, in: 2002 Proceedings of the 16th International Conference on Pattern Recognition, vol. 3, IEEE, Quebec City, Canada, 2002, pp. 29–32.
- [26] M. Kyperountas, A. Tefas, I. Pitas, Weighted piecewise LDA for solving the small sample size problem in face verification, *IEEE Trans. Neural Netw.* 18 (2) (2007) 506–519.
- [27] Y.-F. Guo, S.-J. Li, J.-Y. Yang, T.-T. Shu, L.-D. Wu, A generalized Foley-Sammon transform based on generalized fisher discriminant criterion and its application to face recognition, *Pattern Recognit. Lett.* 24 (1) (2003) 147–158.
- [28] C. Shen, H. Li, M.J. Brooks, A convex programming approach to the trace quotient problem, in: Computer Vision—ACCV 2007, Springer, Tokyo, Japan, 2007, pp. 227–235.
- [29] Y. Jia, F. Nie, C. Zhang, Trace ratio problem revisited, *IEEE Trans. Neural Netw.* 20 (4) (2009) 729–735.
- [30] J. Yang, D. Zhang, J.-Y. Yang, B. Niu, Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (4) (2008) 650–664.
- [31] F. Chung, Spectral graph theory, in: Proceedings of the Regional Conference Series in Mathematics, No. 92, 1997.
- [32] X. He, S. Yan, Y. Hu, P. Niyogi, H.-J. Zhang, Face recognition using Laplacianfaces, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (March (3) (2005)) 328–340.
- [33] W. Deng, J. Hu, J. Guo, H. Zhang, C. Zhang, Comments on “globally maximizing, locally minimizing: unsupervised discriminant projection with application to face and palm biometrics”, *IEEE Trans. Pattern Anal. Mach. Intell.* 30 (August (8) (2008)) 1503–1504.
- [34] H. Wang, S. Yan, D. Xu, X. Tang, T. Huang, Trace ratio vs. ratio trace for dimensionality reduction, in: 2007 IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
- [35] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Morey, Overview of the face recognition grand challenge, in: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2005, pp. 947–954.
- [36] A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (6) (2001) 643–660.
- [37] K.-C. Lee, J. Ho, D.J. Kriegman, Acquiring linear subspaces for face recognition under variable lighting, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (5) (2005) 684–698.
- [38] A.M. Martinez, R. Benavente, The AR face database, Technical Report 24, CVC Technical Report, June 1998.
- [39] R. Gross, I. Matthews, J.F. Cohn, T. Kanade, S. Baker, Multi-PIE, in: FG, 2008.
- [40] Pinellas County Sheriff's Office (<http://www.pcsoweb.com/>).
- [41] G.B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, Technical Report 07-49, University of Massachusetts, Amherst, October 2007.
- [42] F. Juefei-Xu, M. Savvides, An augmented linear discriminant analysis approach for identifying identical twins with the aid of facial asymmetry features, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2013, pp. 56–63.
- [43] F. Juefei-Xu, K. Luu, M. Savvides, T. Bui, C. Suen, Investigating age invariant face recognition based on periocular biometrics, in: 2011 International Joint Conference on Biometrics (IJCB), 2011, pp. 1–7.
- [44] F. Juefei-Xu, M. Savvides, Unconstrained periocular biometric acquisition and recognition using COTS PTZ camera for uncooperative and non-cooperative subjects, in: 2012 IEEE Workshop on Applications of Computer Vision (WACV), 2012, pp. 201–208.
- [45] F. Juefei-Xu, M. Savvides, Can your eyebrows tell me who you are?, in: 2011 5th International Conference on Signal Processing and Communication Systems (ICSPCS), 2011, pp. 1–8.
- [46] F. Juefei-Xu, M. Cha, M. Savvides, S. Bedros, J. Trojanova, Robust periocular biometric recognition using multi-level fusion of various local feature extraction techniques, in: IEEE 17th International Conference on Digital Signal Processing (DSP), 2011.
- [47] F. Juefei-Xu, M. Cha, J.L. Heyman, S. Venugopalan, R. Abiantun, M. Savvides, Robust local binary pattern feature sets for periocular biometric identification, in: 4th IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS), 2010, pp. 1–8.

- [48] F. Juefei-Xu, M. Savvides, Pokerface: partial order keeping and energy repressing method for extreme face illumination normalization, in: 2015 IEEE Seventh International Conference on Biometrics: Theory, Applications and Systems (BTAS), 2015, pp. 1–8.
- [49] F. Juefei-Xu, M. Savvides, Single face image super-resolution via solo dictionary learning, in: IEEE International Conference on Image Processing (ICIP), 2015.
- [50] F. Juefei-Xu, M. Savvides, Pareto-optimal discriminant analysis, in: IEEE International Conference on Image Processing (ICIP), 2015.
- [51] F. Juefei-Xu, M. Savvides, Encoding and decoding local binary patterns for harsh face illumination normalization, in: IEEE International Conference on Image Processing (ICIP), 2015.
- [52] F. Juefei-Xu, D.K. Pal, M. Savvides, NIR-VIS heterogeneous face recognition via cross-spectral joint dictionary learning and reconstruction, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2015.
- [53] F. Juefei-Xu, D.K. Pal, K. Singh, M. Savvides, A preliminary investigation on the sensitivity of COTS face recognition systems to forensic analyst-style face processing for occlusions, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2015.
- [54] K. Seshadri, F. Juefei-Xu, D.K. Pal, M. Savvides, Driver cell phone usage detection on strategic highway research program (SHRP2) face view videos, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2015.
- [55] F. Juefei-Xu, D.K. Pal, M. Savvides, Hallucinating the full face from the periocular region via dimensionally weighted K-SVD, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2014.
- [56] G.B. Huang, E. Learned-Miller, Labeled faces in the wild: updates and new reporting procedures, in: UMass Amherst Technical Report UM-CS-2014-003, 2014.
- [57] W.J. Scheirer, A. Rocha, A. Sapkota, T.E. Boult, Towards open set recognition, *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)* 36 (2013).
- [58] U. Prabhu, J. Heo, M. Savvides, Unconstrained pose-invariant face recognition using 3d generic elastic models, *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (10) (2011) 1952–1961.
- [59] D. Chen, X. Cao, F. Wen, J. Sun, Blessing of dimensionality: high-dimensional feature and its efficient compression for face verification, in: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Portland, Oregon, USA, 2013, pp. 3025–3032.
- [60] M. Savvides, F. Juefei-Xu, Image matching using subspace-based discrete transform encoded local binary patterns, US Patent US 2014/0212044 A1, September 2013.
- [61] F. Juefei-Xu, M. Savvides, Weight-optimal local binary patterns, in: Computer Vision—ECCV 2014 Workshops, Lecture Notes in Computer Science, vol. 8926, Springer International Publishing, Zurich, Switzerland, 2015, pp. 148–159.
- [62] F. Juefei-Xu, M. Savvides, Subspace based discrete transform encoded local binary patterns representations for robust periocular matching on NIST's face recognition grand challenge, *IEEE Trans. Image Process.* 23 (8) (2014) 3490–3505.
- [63] Felix Juefei-Xu, Khoa Luu, Marios Savvides, Spartans: single-sample periocular-based alignment-robust recognition technique applied to non-frontal scenarios, *IEEE Trans. Image Process.* 24 (12) (2015) 4780–4795.
- [64] Felix Juefei-Xu, Dipan K. Pal, Marios Savvides, Methods and software for hallucinating facial features by prioritizing reconstruction errors, U.S. Provisional Patent Application Serial No. 61/998,043, June 17, 2014.
- [65] Niv Zehngut, Felix Juefei-Xu, Rishabh Bhardia, Dipan K. Pal, Chandrasekhar Bhagavatula, and Marios Savvides, Investigating the feasibility of image-based nose biometrics, in: Proceedings of the IEEE International Conference on Image Processing (ICIP), 2015.
- [66] Felix Juefei-Xu, Marios Savvides, Facial ethnic appearance synthesis, in: Proceedings of Soft Biometrics Workshop, European Conference on Computer Vision (ECCV), 2014.
- [67] Felix Juefei-Xu, Marios Savvides, An image statistics approach towards efficient and robust refinement for landmarks on facial boundary, in: Proceedings of IEEE 6th International Conference on Biometrics: Theory, Applications and Systems (BTAS), 2013.
- [68] Shreyas Venugopalan, Felix Juefei-Xu, Benjamin Cowley, Marios Savvides, Electromyograph and keystroke dynamics for spoof-resistant biometric authentication, in: Biometrics Workshop, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
- [69] Felix Juefei-Xu, Chandrasekhar Bhagavatula, Aaron Jaech, Unni Prasad, Marios Savvides, Gait-ID on the move: pace independent human identification using cell phone accelerometer dynamics, in: Proceedings of IEEE 5th International Conference on Biometrics: Theory, Applications and Systems (BTAS), 2012.

Felix Juefei-Xu received the B.S. degree in electronic engineering from Shanghai Jiao Tong University (SJTU), Shanghai, China, and the M.S. degree in electrical and computer engineering from Carnegie Mellon University (CMU), Pittsburgh, PA, USA, where he is currently pursuing the Ph.D. degree in electrical and computer engineering. He is currently with the CMU CyLab Biometrics Center, in a research group specializing in pattern recognition, machine learning, computer vision, and image processing, especially as applied to the field of biometrics. His research is focused on unconstrained periocular face recognition, especially under extreme resolution, illumination, and occlusion (xRIO) conditions. He also has broader interests in pattern recognition, computer vision, machine learning, optimization, statistics, compressive sensing, and image processing. He is the recipient of the Best Poster Paper Award of the IEEE/IAPR Joint International Conference on Biometrics (IJCB) in 2011, and the Best Paper Award of the IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS) in 2015.

Marios Savvides is the Founder and Director of the Biometrics Center and currently a Research Professor of the Electrical and Computer Engineering Department with Carnegie Mellon University (CMU). He is also one of the tapped researchers to form the Office of the Director of National Intelligence 1st Center of Academic Excellence in Science and Technology. His research is mainly focused on developing algorithms for robust face and iris biometrics and pattern recognition, machine vision, and computer image understanding for enhancing biometric systems performance. He has authored or co-authored over 170 journal and conference publications, including several book chapters in the area of Biometrics and served as the Area Editor of the Springer's Encyclopedia of Biometrics. His achievements include leading the research and development in CMU's past participation at NIST's Face Recognition Grand Challenge 2005 (CMU ranked one in academia and industry at the hardest experiment four open challenge) and also in NIST's Iris Challenge Evaluation (CMU ranked one in academia and two against iris vendors). He has filed over 20 patent applications in the area of biometrics and was a recipient of the CMU's 2009 Carnegie Institute of Technology Outstanding Faculty Research Award. He is on the Program Committee on several Biometric conferences, such as the IEEE BTAS, ICPR, SPIE Biometric Identification, the IEEE AutoID, and others, and has been the Organizer and the Co-Chair of the Robust Biometrics Understanding the Science and Technology Conference.