



# Multi-class Fukunaga Koontz discriminant analysis for enhanced face recognition

Felix Juefei-Xu<sup>\*</sup>, 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

<sup>\*</sup> 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  $\mathbf{W}^T \mathbf{S}_B \mathbf{W} = \mathbf{I}$  and  $\mathbf{W}^T \mathbf{S}_W \mathbf{W} = \mathbf{\Lambda}$ . 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  $\mu$  may not be sufficiently effective. Cross-validation is commonly applied for estimating the optimal  $\mu$ . But in practice, it is always not an easy task to determine the optimal  $\mu$  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  $\Gamma$ .  $\Gamma$  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  $\text{rank}(\mathbf{W}^T \mathbf{H}_B) = q$ , 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.

Download English Version:

<https://daneshyari.com/en/article/533211>

Download Persian Version:

<https://daneshyari.com/article/533211>

[Daneshyari.com](https://daneshyari.com)