A New Regularized Orthogonal Local Fisher Discriminant Analysis for Image Feature Extraction

Zhongfeng WANG[1], Zhan WANG[2]

dept. name of organization, name of organization, City, Country
[1.] *School of Management Engineering*, Henan Institute of Engineering, Zhengzhou 451191, P.R. China
[2.] *Institute of Information Science*, Beijing Jiaotong University, Beijing 100044, P.R. China

***Abstract* — Local Fisher Discriminant Analysis (LFDA) is a feature extraction method which combines the ideas of Fisher discriminant analysis (FDA) and locality preserving projection (LPP). It works well for multimodal problems. But LFDA suffers from the under-sampled problem of the linear discriminant analysis (LDA). To deal with this problem, we propose a regularized orthogonal local Fisher discriminant analysis (ROLFDA) to improve the performance of LFDA. ROLFDA finds the optimal projection vectors in the range space of local mixture scatter matrix. The new local within-class scatter matrix is approximated by regularization method in order to regulate the eigenvalues of the newly local within-class scatter matrix. The projection vectors are orthogonal by utilizing the trace ratio criterion. Experimental results on the Yale, ORL and CMU PIE face databases demonstrate that the ROLFDA is an effective algorithm.**

Keywords - Information extraction and pattern analysis; Fisher discrimant analysis; image feature extraction, undersampled problem; face databases

I. INTRODUCTION

Principal component analysis (PCA) [1] and linear discriminant analysis (LDA) [2] are two famous dimensionality reduction algorithms and have been widely used in the areas of pattern recognition and computer vision. PCA aims to find a set of mutually orthogonal basis vectors by maximizing the variance of the projected data. Rather than PCA, LDA is a supervised technique. It is to find an optimal projection matrix by maximizing the ratio of the between-class scatter to the within-class scatter. PCA and LDA can only discover the global Euclidean structure of the data. They fail to find the nonlinear data structure.

Recent studies show that the face images may reside on a nonlinear sub-manifold [3]. Manifold-based learning algorithms have been proposed to discover the intrinsic low-dimensional structure. LPP [3] was proposed for face recognition straightly. It finds an embedding subspace that preserving local information and discovering the essential face manifold structure. For face recognition or classification, discriminant information is very important. LDA-based methods have been proposed to tackle the problem of tradition LDA, such as the undersampled problem. LPP also can be extended to supervised and discriminant models. Recently, Sugiyama proposed local fisher discriminant analysis (LFDA) [4], which combines the ideas of FDA and LPP. It reformulates LDA and utilizes the neighborhood information indirect to redefine the local between-class scatter matrix and the local within-class scatter matrix. Can work well when within-class multimodality or outliers exist.

LFDA also suffers from the undersampled problem as well as LDA. To overcome this drawback, we propose a regularized orthogonal local fisher discriminant analysis (ROLFDA) method to improve the performance of LFDA. We firstly find the projection matrix in the range space of local mixture scatter matrix. The newly local within-class scatter matrix can be approximated well by the regularization method [5,6] in order to alleviate the bad impacts of small and zero eigenvalues caused by noise and limited number of training samples. The trace ratio criterion was applied to obtain the orthogonal projection vectors. Since the orthogonal projection is of desirable property and often has good performance empirically. The proposed method can improve the performance of LFDA significantly. Experiment results on some face databases verify the effectiveness of the proposed algorithm.

The rest of this paper is organized as follows. In section 2, we review the idea of local fisher discriminant analysis (LFDA). In section 3, we propose a regularized orthogonal local fisher discriminant analysis (ROLFDA) in detail. In section 4, experiments with face images databases are implemented to demonstrate the effectiveness of ROLFDA algorithm. Conclusions are given in section 5.

II. OVERVIEW OF LFDA

LFDA is also a discriminant analysis method. Given N Samples $X=\{x_1,x_2,\cdots x_N\}$ in n -dimensional space .The class label of sample x_i is denoted as $y_i=\{1,2,\cdots c\}$. N_c is the number of classes. We denote n_c as the number of the samples in the c th class.

LFDA redefines the between-class metric matrix $\overline{S_b}$ and the with-class metric matrix $\overline{S_w}$ as follows:

$$\overline{S_b} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij}^{(b)}(x_i - x_j)^T(x_i - x_j) \tag{1}$$

$$\overline{S_w} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij}^{(w)}(x_i - x_j)^T(x_i - x_j) \tag{2}$$

$W^{(b)}$ and $W^{(w)}$ is the weight matrices:

$$W_{ij}^{(b)} = \begin{cases} A_{ij}(1/n - 1/n_{l_i}) & \text{if } y_i = y_j \\ 1/n & \text{if } y_i \neq y_j \end{cases} \tag{3}$$

$$W_{ij}^{(w)} = \begin{cases} A_{ij}/n_{l_i} & \text{if } y_i = y_j \\ 0 & \text{if } y_i \neq y_j \end{cases} \tag{4}$$

A is the affinity matrix, A reflects the relationship between x_i and x_j. $A_{ij} = \exp(-\|x_i - x_j\|^2 / \sigma_i\sigma_j)$. $\sigma_i = \|x_i - x_i^{(k)}\|$. $x_i^{(k)}$ is the k-th nearest neighbor of x_i. If the entire element of A are one, the LFDA method can be transformed to standard LDA. The objection function of LFDA is:

$$V_{opt} = \arg\max_V \frac{V^T\overline{S_b}V}{V^T\overline{S_w}V} = \arg\max_V \frac{tr(V^T\overline{S_b}V)}{tr(V^T\overline{S_w}V)} \tag{5}$$

The above optimal problem is transformed to generalized eigenvalue problem as follows:

$$\overline{S_b}v_i = \lambda\overline{S_w}v_i, i = 1,2,\cdots,d \tag{6}$$

$V = [v_1,\cdots,v_d]$ are the eigenvectors corresponding to the d largest eigenvalues. Because of the choice of A, the LFDA can reveal some local information and overcome some drawbacks of tradition LDA.

Let $\overline{S_t} \triangleq \overline{S_b} + \overline{S_w}$ be the local mixture scatter matrix, then the optimal problem (5) is equivalent to:

$$V_{opt} = \arg\max_V \frac{V^T\overline{S_b}V}{V^T(\overline{S_w}+\overline{S_b})V} = \arg\max_V \frac{V^T\overline{S_b}V}{V^T\overline{S_t}V} \tag{7}$$

III. REGULARIZED ORTHOGONALLOCAL FISHER DISCRIMINANT ANALYSIS

A. *Discriminant feature analysis*

We decompose the whole space into the following four sunspaces according to the eigenvalues of $\overline{S_b}$ and $\overline{S_w}$ [7].

Subspace 1: span ($\overline{S_w}$)$\cap$span($\overline{S_b}$). In this subspace, the eigenvalues of $\overline{S_w}$ and $\overline{S_b}$ are greater than zero. The discriminant information exists in this subspace.

Subspace 2: span ($\overline{S_w}$)$\cap$null($\overline{S_b}$). In this subspace, the eigenvalues of $\overline{S_w}$ are greater than zero, the eigenvalues of $\overline{S_b}$ are equal zero. The discriminant information exists in this subspace.

Subspace 3: null ($\overline{S_w}$)$\cap$span($\overline{S_b}$). In this subspace, the eigenvalues of $\overline{S_w}$ are equal zero, the eigenvalues of $\overline{S_b}$ are greater than zero. The discriminant information exists in this subspace.

Subspace 4: null ($\overline{S_w}$)$\cap$null($\overline{S_b}$). In this subspace, the eigenvalues of $\overline{S_w}$ and $\overline{S_b}$ are equal zero. The objective function (5) is meaningless. This subspace does not contain discriminant information.

Since the null space of the local mixture scatter matrix is the intersection of the null space of $\overline{S_b}$ and the null space of $\overline{S_w}$. We can remove the null space of $\overline{S_t}$ without discarding discriminant information. Suppose that the singular value decomposition (SVD) of $\overline{S_t}$ is

$$\overline{S_t} = U\Lambda Q^T \tag{8}$$

Since the matrix $\overline{S_t}$ is symmetric and positive semidefinite, $U=Q$. If the rank of $\overline{S_t}$ is r, U_r is the matrix consisting of the first r columns of matrix U. So the cost function (5) is equivalent to the following objection function by $V = U_rP$:

$$P_{opt} = \arg\max_P \frac{P^TU_r^T\overline{S_b}U_rP}{P^TU_r^T\overline{S_w}U_rP} = \arg\max_P \frac{P^T\overline{\overline{S_b}}P}{P^T\overline{\overline{S_w}}P} \tag{9}$$

where $\overline{\overline{S_b}} = U_r^T\overline{S_b}U_r$, $\overline{\overline{S_w}} = U_r^T\overline{S_w}U_r$. The orthogonal constraint is added to avoid the trivial solution. So the objection function (9) is changed to:

$$P_{opt} = \arg\max_{P^TP=I} \frac{P^T\overline{\overline{S_b}}P}{P^T\overline{\overline{S_w}}P} \tag{10}$$

B. *Regularized orthogonal local fisher discriminant analysis*

In formula (10), $\overline{\overline{S_w}}$ may be singular. For alleviating the impacts of small and zero eigenvalues, we utilized regularization techniques to regulate the eigenvalues of $\overline{\overline{S_w}}$. $\overline{\overline{S_w}}$ is symmetric and positive semidefinite , So the singular value decomposition of matrix $\overline{\overline{S_w}}$ is represented as follows:

$$\overline{\overline{S_w}} = \sum_{i=1}^{m}\overline{\sigma}_i\overline{u}_i\overline{u}_i^T = \sum_{i=1}^{t}\overline{\sigma}_i\overline{u}_i\overline{u}_i^T \tag{11}$$

where $\overline{\sigma}_1 \geq \overline{\sigma}_2 \geq \cdots \geq \overline{\sigma}_t > \overline{\sigma}_{t+1} = \cdots = \overline{\sigma}_m = 0$, t is the rank of matrix $\overline{\overline{S_w}}$. As pointed in [5], the $\overline{\overline{S_w}}$ also can be decomposed into three subspaces: the face subspace, noise subspace, null subspace. So the $\overline{\overline{S_w}}$ can be expressed as follows:

$$\overline{\overline{S_w}} = \underbrace{\sum_{i=1}^{k} \bar{\sigma}_i \bar{u}_i \bar{u}_i^T}_{\text{face subspace}} + \underbrace{\sum_{i=k+1}^{t} \bar{\sigma}_i \bar{u}_i \bar{u}_i^T}_{\text{noise subspace}} + \underbrace{\sum_{i=t+1}^{m} \bar{\sigma}_i \bar{u}_i \bar{u}_i^T}_{\text{null subspace}} \quad (12)$$

In LPP, PCA was used for dimensionality reduction, keep 98% information and get the principal components. So the value of k is calculated by the following rule:

$$\frac{\sum_{i=1}^{k} \bar{\sigma}_i}{\sum_{i} \bar{\sigma}_i} \approx 98\% \quad (13)$$

The function form $\frac{1}{f}$ [5] was used to model the spectrum of face space:

$$\overline{\overline{S_w}}^* = \sum_{i=1}^{k} \bar{\sigma}_i \bar{u}_i \bar{u}_i^T + \sum_{i=k+1}^{t} \frac{\alpha}{i+\beta} \bar{u}_i \bar{u}_i^T + \sum_{i=t+1}^{m} \xi \bar{u}_i \bar{u}_i^T$$

(14)

ξ is a constant, α, β are parameters and can be determined by letting $\bar{\sigma}_1 = \frac{\alpha}{1+\beta}, \bar{\sigma}_k = \frac{\alpha}{k+\beta}$,and get

$$\alpha = \frac{\bar{\sigma}_1 \bar{\sigma}_k (k-1)}{\bar{\sigma}_1 - \bar{\sigma}_k}, \beta = \frac{k\bar{\sigma}_k - \bar{\sigma}_1}{\bar{\sigma}_1 - \bar{\sigma}_k}$$

(15)

$\overline{\overline{S_w}}^*$ is regarded as the approximation of $\overline{\overline{S_w}}$, Let $\overline{\overline{S_w}} = \overline{\overline{S_w}}^*$, the optimal problem (10) can be changed into:

$$P_{opt} = \arg\max_{P^T P=I} \frac{P^T \overline{\overline{S_b}} P}{P^T \overline{\overline{S_w}} P} = \arg\max_{P^T P=I} \frac{tr(P^T \overline{\overline{S_b}} P)}{tr(P^T \overline{\overline{S_w}}^* P)}$$

(16)

C. *Finding the optimal solution*

The formula (16) is a trace ratio problem. The trace ratio problem does not have a closed-form solution. It is usually approximated by solving a ratio trace problem $\arg\max_{P^T P=I} tr[(P^T \overline{\overline{S_w}}^* P)^{-1}(P^T \overline{\overline{S_b}} P)]$. However, there exist different methods to solve the trace ratio problem. Wang et.al proposed an iterative method called iterative trace ratio (ITR) to solve the trace ratio problem [8]. Yang .et.al proposed a more efficient algorithm called decomposed Newton's method (DNM) to optimize the trace ratio problem [9]. In ITR algorithm, the projection matrix is initialized as an arbitrary column orthogonal matrix. DNM initializes the trace ratio value as zero, and the update rule is changed. Zhao.et.al gave the lower boundary of trace ratio value [10]. For problem (16), we can obtain that the lower boundary of trace ratio value is $tr(\overline{\overline{S_b}}) / tr(\overline{\overline{S_w}}^*)$. So we apply the DNM method to solve the trace ratio problem (16) with the fixed initialization of trace ratio value.

IV. EXPERIMENTS

To evaluate the proposed ROLFDA algorithm, we compared it with PCA, LDA, LPP, LFDA and eigenfeature regularization and extraction (ERE) on Yale, ORL and CMU PIE face databases. LDA, LPP and LFDA need PCA preprocessing to avoid the singularity problem and 98% energy was kept in PCA phase. The nearest neighborhood classifier with Euclidean distance metric was used in experiments.

A. *Experiment on Yale face database*

The Yale face database [11] contains 165 grayscale images of 15 individuals. Each individual has 11 images. These images were captained under various lighting condition (center-light, left-light, right-light), facial expressions (normal, happy, sad, surprise and wink) and facial details (with glasses or without). The original size of the images in this database is 243×320 pixels. In our experiments, all images were cropped into 32×32 by fixing the location of the two eyes.

We randomly selected $l(l = 3,4,5,6)$ images from each individual for training, while the rest images were used for testing. For each given l , we repeated experiments over 50 random splits and the average recognition rates were obtained. For LDA and ERE, the dimensionalities of feature space are at most 14. For LFDA and ROLFDA, the parameters are set to $l-1$. The parameter ξ in formula (14) is chosen as one. The best performances of different methods with corresponding dimensionality were presented in Table 1. From Table 1, ROLFDA achieves higher accuracy than the other methods. Next, we experimented with different dimensionalities of feature space. A random subset with 5 images per individual was used for training, and the remaining images were used for testing. The average recognition rates over 50 independent runs are shown in Fig .(1).

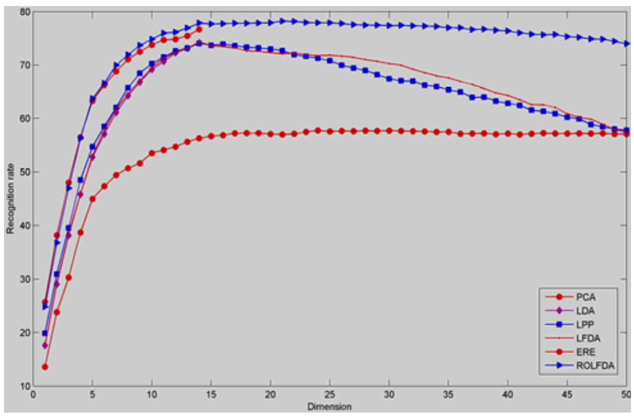

Figure 1. Recognition Rates (%) Versus the Dimension of Feature Space on the Yale Face Database

TABLE I. THE TOP RECOGNITION RATE (%) OF DIFFERENT METHODS ON THE YALE FACE DATABASE

Method	3 train	4train	5strain	6train
PCA	51.10(40)	54.86(60)	58.09(74)	60.72(89)
LDA	60.70(14)	68.59(14)	74.13(14)	77.55(14)
LPP	62.13(14)	68.97(14)	74.04(14)	77.36(14)
LFDA	62.98(14)	69.85(14)	74.31(14)	77.55(14)
ERE	64.58(14)	72.25(14)	76.69(14)	78.08(14)
ROLFDA	66.62(15)	74.29(19)	78.71(22)	82.48(17)

B. *Experiment on ORL face database*

The ORL database [12] contains 400 facial images of 40 individuals. Each individual has 10 different images. All the images were taken against a dark homogeneous background with the subjects in an upright, front position (with tolerance for some side movement). The original size of the images is 112×92 pixels. In our experiment, the images were manually cropped and resized to 32×32 pixels.

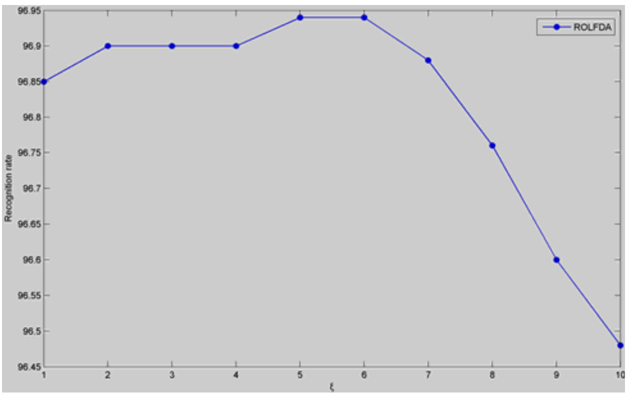

Figure 2. Recognition Rates (%) Versus the Parameter ξ on the ORL Face Database.

TABLE II. THE TOP RECOGNITION RATE (%) OF DIFFERENT METHODS ON THE ORL FACE DATABASE

Method	3 train	4train	5strain	6train
PCA	77.95(78)	83.92(100)	87.39(116)	90.30(148)
LDA	86.13(39)	91.23(39)	94.01(39)	95.25(39)
LPP	87.44(39)	91.05(39)	94.40(44)	95.72(42)
LFDA	87.66(39)	91.20(39)	94.34(39)	95.63(43)
ERE	90.04(39)	93.95(39)	96.50(39)	97.25(38)
ROLFDA	91.11(50)	94.96(42)	96.94(48)	97.56(44)

We randomly selected $l(l=3,4,5,6)$ images from each individual for training, while the rest images were used for testing. For each given l, we repeated experiments over 40 random splits and the average recognition rates were obtained. The best performances of different methods with corresponding dimensionality were presented in Table 2. From Table 2, ROLFDA is best. Next, we aim to search for optimal ξ in formula (14). Fig.(2). shows the different ξ with corresponding recognition rate.

C. *Experiment on CMU PIE face database*

The CMU PIE face database [13] contains 68 people with 41368 face images. The face images were captained under varying pose illumination and expression. We selected a subset (C29) which contains 1632 images of 68 individuals (each people has 24 images). The images were manually cropped and resized 32×32 pixels.

We randomly selected $l(l=4,8,12,16)$ images from each individual for training, while the rest images were used for testing. For each given l, 20 independent runs were performed and the average recognition rates were obtained. The best performances of different methods with corresponding dimensionality were presented in Table 3. From Table 3, ROLFDA outperforms other methods.

TABLE III. THE TOP RECOGNITION RATE (%) OF DIFFERENT METHODS ON THE CMU PIE FACE DATABASE

Method	4 train	8train	12train	16train
PCA	41.81(99)	63.96(148)	75.25(197)	77.22(278)
LDA	81.95(67)	89.34(67)	89.77(67)	88.69(67)
LPP	82.54(44)	89.11(67)	90.01(67)	90.69(75)
LFDA	82.87(99)	89.89(92)	90.53(94)	90.17(87)
ERE	83.20(67)	90.53(67)	91.54(67)	92.03(67)
ROLFDA	84.15(67)	90.99(77)	91.91(82)	93.01(97)

V. CONCLUSION

In this paper, we presented a regularized orthogonal local Fisher discriminant analysis (ROLFDA) for dimensionality reduction and feature extraction. ROLFDA finds the optimal solution in the range space of local mixture scatter matrix. The newly local within-class scatter matrix is approximated by regularization method in order to alleviate the unreliable impacts of small and zero eigenvalues which caused by noise and limited training samples. The trace ratio criterion is applied to obtain orthogonal projection vectors. Experimental results on Yale, ORL and CMU PIE face databases show the effectiveness of the proposed ROLFDA algorithm.

ACKNOWLEDGMENT

This work was financially supported by Doctoral Fund Project of Henan University of Engineering (D2015030).

REFERENCES

[1] [1] M.A. Turk, A.P. Pentland, Eigenfaces for recognition, J. Cognitive Neurosci. 3 (1) (1991) 71–96.

[2] [2] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenface vs. Fisherface: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.

[3] [3] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, Face recognition using Laplacianfaces, IEEE Trans. Pattern Anal. Mach. Intell. 27 (3) (2005)328–340.

[4] [4] M. Sugiyama, Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis, J. Mach. Learn. Res. 8 (2007) 1027–1061.

[5] [5] X. Jiang, B. Mandal, and A. Kot, “Eigenfeature regularization and extrac-tion in face recognition,” IEEE Trans. Pattern Anal. Mach. Intell. , vol. 30,no. 3, pp. 383–394, Mar. 2008.

[6] [6] J.Lu , Y P. Tan . Regularized locality preserving projections and its extensions for face recognition. IEEE Trans Syst Man Cybern B Cybern. 2010 Jun;40(3):958-63.

[7] [7]Z S. Zhang, T. Sim Discriminant subspace analysis: a Fukunaga–Koontz approach IEEE Trans. Pattern Anal. Mach. Intell, 29 (10) (2007), pp. 1732–1745

[8] [8] H. Wang, S. Yan, D. Xu, and X. Tang. Trace ratio vs. ratio trace for dimensionality reduction. In CVPR '07. IEEE Conference on Computer Vision and Pattern Recognition, 2007.

[9] [9] Y. Jia, F. Nie, and C. Zhang, "Trace ratio problem revisited," IEEE Trans. Neural Netw., vol. 20, no. 4, pp. 729–735, Apr. 2009.

[10] [10] M.B Zhao, Z. Zhang, W. S. Tommy: Trace ratio criterion based generalized discriminative learning for semi-supervised dimensionality reduction. Pattern Recognition 45(4): 1482-1499 (2012)

[11] [11]F. Samaria, A. Harter, Parameterisation of a stochastic model for human face identification, in: Second IEEE Workshop on Applications of Computer Vision, 1994.

[12] [12]Yale University Face Database, http://cvc.yale.edu/projects/yalefaces/ yalefaces.html , 2002.

[13] [13]T. Sim, S. Baker, and M. Bsat, The CMU pose, illumination, and ex-pression (PIE) database of human faces Robot. Instit., Carnegie-Mellon Univ., Pittsburgh, PA, 2001, Tech. Rep. CMU-RI-TR-01-02