

FDLDA: An Fast Direct LDA algorithm For Face Recognition

GUO Zhibo, LIN Kejun
School of Information Engineer
Yangzhou University
Yangzhou, China
zhibo_guo@163.com

YAN Yunyang
Faculty of Computer Engineering
Huaiyin Institute of Technology
Huaian, China
areyyyke@163.com

Abstract—Feature extraction is one of the hot topics in face recognition. However, many face extraction methods will suffer from the “small sample size” problem, such as Linear Discriminant Analysis (LDA). Direct Linear Discriminant Analysis (DLDA) is an effective method to address this problem. But conventional DLDA algorithm is often computationally expensive and not scalable. In this paper, DLDA is analyzed from a new viewpoint via SVD and an fast and robust method named FDLDA algorithm is proposed. The proposed algorithm achieves high efficiency by introducing the SVD on a small-size matrix, while keeping competitive classification accuracy. Experimental results on ORL face database demonstrate the effectiveness of the proposed method.

Keywords—feature extraction; fast direct linear discriminant analysis; face recognition

I. INTRODUCTION

Face Recognition (FR) has a wide range of applications, such as military, commercial and law enforcement et al. Within the past two decades, numerous FR [1-3],[6-20] algorithms have been proposed. Among these FR methods, the most popular methods are appearance-based approaches.

Of the appearance-based FR methods, those utilizing Linear Discriminant Analysis (LDA) [4-20] techniques have shown promising results. Conventional LDA [4,5] algorithm aims to find an optimal transformation by minimizing the within-class scatter matrix and maximizing the between class scatter matrix simultaneously. The optimal transformation is readily computed by applying the Eigen-decomposition to the scatter matrices. But an intrinsic limitation of conventional LDA is that its objective function requires the within-class scatter matrix nonsingular. For many real applications, the within-class scatter matrix is often singular since the dimension of sample exceeds the number of sample and conventional LDA based methods suffer from the so-called “small sample size” [6-8] problem.

In the last decades, numerous methods have been proposed to solve this problem. Tian et al [9] used the pseudo inverse method by replacing inverse of within-class scatter matrix with

This work was sponsored by the Prospective Joint Research Project of Jiangsu Province (BY201506-01), Domestic Research Collaboration Innovation Project of Yangzhou (2014-8).

its pseudo inverse. The perturbation method is used in [10], where a small perturbation matrix is added to within-class scatter matrix in order to make it nonsingular. Cheng et al [11] proposed the Rank Decomposition method based on successive eigen-decomposition of the total scatter matrix and the between-class scatter matrix. However, the above methods are typically computationally expensive since the scatter matrices are very large. Swets and Weng [12] proposed a two stages PCA+LDA method, also known as the Fisherface [6] method, in which PCA is first used for dimension reduction so as to make the within-class scatter matrix nonsingular before the application of LDA. By far, the PCA+LDA method is popular used. However, algorithms based on this solution may discard effective features in PCA step. To prevent this from happening, many extended LDA algorithms with null space conception were proposed. Chen et al [13] and Yang et al [14] developed DLDA algorithm for face recognition, which can effectively solve “small sample size” problem and extract optimal classification features from original samples. But conventional DLDA [13-15] algorithm is often computationally expensive and not scalable.

In this paper, first of all, we briefly recall the DLDA algorithm. Then we perform an in-depth analysis on DLDA algorithm and proposed a Fast DLDA algorithm. The utilization of the singular value decomposition (SVD)[19-20] on the small-size matrix is one of key steps. The proposed method is efficient and robust. Moreover, the theoretical foundation of the proposed method is revealed..

II. OUTLINE OF DIRECT LDA

Throughout the paper, C denotes the number of classes, m is the dimension, N is the number of samples in each class, μ_i is the centroid of the i th class, and μ is the holistic centroid of the whole data set. S_b , S_w and S_t represent between-class scatter matrix, within-class scatter matrix and total class scatter matrix, respectively.

DLDA [13-15] algorithm was proposed by Chen and Yang, which attempts to avoid the shortcomings existing in

conventional solution to the “small sample size” problem. The basic idea behind the DLDA algorithm is that the null space of S_w may contain effective discriminant information if the projection of S_b is not zero in that direction, and that no effective information will be lost if the null space of S_b is discarded. For example, assuming that N_b and N_w represent the null space of S_b and S_w , respectively, the complement spaces of N_b and N_w can be written as $N'_b = R^n - N_b$ and $N'_w = R^n - N_w$. Therefore, the optimal discriminant subspace extracted by the DLDA algorithm is the intersection space $N'_b \cap N'_w$.

The difference between Chen's method and Yang's method is that Yang's method first diagonalizes S_b to find N'_b , while Chen's method first diagonalizes S_w to find N'_w . Although there is no significant difference between the two approaches, it may be intractable to calculate N'_w when the size of S_w is large, which is the case in most FR application. Therefore, we adopted Chen's method as the derivation of the proposed algorithm.

III. FDLDA ALGORITHM FOR DIMENSION REDUCTION AND FEATURE EXTRACTION

In this section, we will present the FDLDA algorithm. Based on the analysis of Chen's DLDA algorithm.

We analysis the Chen's algorithm about DLDA. They perform the Eigen value decomposition of S_w as $S_w = V\Lambda V^T$ and let $V = [v_1, v_2, \dots, v_r, v_{r+1}, \dots, v_n]$, then they let $Q_\perp = [v_1, v_2, \dots, v_r]$ and $Q = [v_{r+1}, \dots, v_n]$ and the null space of S_w can be spanned by v_{r+1}, \dots, v_n , where r is rank of S_w and n is the dimensionality of the feature vectors.

They Compute \tilde{S}_b according to $\tilde{S}_b = QQ^T S_b QQ^T$ and calculate the eigenvectors u_1, u_2, \dots, u_d corresponding to the set of the largest eigenvalues of \tilde{S}_b and use them to form the most discriminant vector set for LDA.

Due to $r \ll n$, so the calculation of the S_w , Q and \tilde{S}_b will greatly improve the computational complexity. our thoughts are calculating the final differential vector according to Q_\perp . Q_\perp can be obtained by solving the eigen problem on S_w similar to PCA algorithm based SVD.

Because between-class scatter matrix S_b and within-class scatter matrix S_w can be represented by following:

$$S_b = H_b H_b^T, S_w = H_w H_w^T \quad (1)$$

H_b , H_w as following:

$$H_b = \frac{1}{\sqrt{c}} (\bar{x}_1 - \bar{x}, \bar{x}_2 - \bar{x}, \dots, \bar{x}_c - \bar{x}) \quad (2)$$

$$H_w = \frac{1}{\sqrt{N}} (x_1^{(1)} - \bar{x}_1, \dots, x_1^{(n_1)} - \bar{x}_1, \dots, x_c^{(1)} - \bar{x}_c, \dots, x_c^{(n_c)} - \bar{x}_c) \quad (3)$$

Assuming that $G_w = H_w^T H_w$, and making the eigenvalue decomposition of G_w . We can calculate the eigenvectors z_1, z_2, \dots, z_d corresponding to the nonzero eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_r$ of G_w . According eq. 4, we can get the feature vector set of $Q_\perp = (v_1, \dots, v_r)$ corresponding to nonzero eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_r$.

$$v_i = \frac{1}{\sqrt{\lambda_i}} H_w z_i, \quad i = 1, 2, \dots, r \quad (4)$$

Q is the null space matrix of S_w . we can obtain:

$$\tilde{S}_b = QQ^T H_b (QQ^T H_b)^T = \tilde{H}_b \tilde{H}_b^T \quad (5)$$

We can construct matrix $\tilde{G}_b = \tilde{H}_b^T \tilde{H}_b$, and calculate the eigenvectors z_1, z_2, \dots, z_d corresponding to the set of the largest eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_d$ of \tilde{G}_b , therefore the eigenvectors u_1, u_2, \dots, u_d corresponding to the set of the largest eigenvalues of \tilde{S}_b can be calculated by the following equation :

$$u_i = \frac{1}{\sqrt{\lambda_i}} \tilde{H}_b z_i, \quad i = 1, \dots, d \quad (6)$$

According to specialty of orthal vectors, we can obtain:

$$QQ^T = I - Q_\perp Q_\perp^T \quad (7)$$

According to eq.7, we have:

$$\tilde{H}_b = QQ^T H_b = H_b - Q_\perp Q_\perp^T H_b \quad (8)$$

Therefore, we can get:

$$\begin{aligned} \tilde{G}_b &= (H_b - Q_\perp Q_\perp^T H_b)^T (H_b - Q_\perp Q_\perp^T H_b) \\ &= H_b^T H_b - H_b^T Q_\perp Q_\perp^T H_b \\ &= H_b^T H_b - (Q_\perp^T H_b)^T Q_\perp^T H_b \end{aligned} \quad (9)$$

According to eq.9, \tilde{G}_b can be calculated in a low-dimensional space, which greatly saves storage space and reduces the amount of computation. For example, the original

sample dimension $M=10304$, the number of categories $c=40$ in the ORL face database. if we choose 200 training samples, the rank r of S_w is 160. the vector number of Q_\perp is 160, but the vector number of Q_\perp is up to 10304. The memory space of Q is dozens of times of Q_\perp .

Based on the above discussion, the proposed FDLDA algorithm is described as follows:

Step 1: Obtain H_w , H_b and calculate Q_\perp based on SVD. Let $\tilde{G}_b = H_b^T H_b - (Q_\perp^T H_b)^T Q_\perp^T H_b$.

Step 2: Calculate the eigenvector matrix $Z=[z_1, z_2, \dots, z_d]$ and the eigenvectors z_1, z_2, \dots, z_d corresponding to the set of the largest eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_d$ of \tilde{G}_b .

Step 3: According to $u_i = \frac{1}{\sqrt{\lambda_i}}(H_b - Q_\perp Q_\perp^T H_b)z_i, i=1, \dots, d$ Calculate the eigenvector matrix $U=[u_1, u_2, \dots, u_d]$ and the discriminant matrix $A=U - Q_\perp Q_\perp^T U$.

Step 4: Using the discriminant matrix A extract efficient features and classify.

IV. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of our method, experiments were done on the ORL face database. Fig.1 depicts some images from the ORL face database. ORL database includes 200 samples (40 individuals, each 10 images).



Fig. 1. Some samples from the ORL face database

For comparison of FDLDA and Chen's DLDA algorithm, the experiment uses three resolutions ($46 \times 56, 35 \times 42, 23 \times 28$). the sample image is obtained by bilinear interpolation. In our experiments, the training and testing set are selected randomly from each subject. In each round, the training samples are

selected randomly from the gallery and the remaining samples are used for testing. This procedure was repeated 10 times by randomly choosing different training and testing sets. After feature extraction, a nearest neighbor classifier with different distance metrics is employed for classification.

Experimental environment: MATLAB2010, Core i5, SRAM 2G and so on.

Table 1. The experimental Comparison using 45×56 resolution

	FDLDA	Chen's DLDA
Recognition rate(%)	95.0	95.0
Training time (s)	2.3	552

Table 2. The experimental Comparison using 35×42 resolution

	FDLDA	Chen's DLDA
Recognition rate(%)	95.25	95.25
Training time (s)	1.3	88

Table 3. The experimental Comparison using 23×28 resolution

	FDLDA	Chen's DLDA
Recognition rate(%)	95.5	95.5
Training time (s)	0.6	5.4

The experimental results of FDLDA algorithm and Chen's DLDA algorithm is shown on the Tab.1~Tab3. The accuracy of two algorithms is same using the same kind of resolution, but the FDLDA algorithm is much faster than Chen's DLDA on the training time. Especially with the resolution increases, the gap between training time is also growing. For different resolutions, FLDA's training time is similar and the difference is not obvious. Experimental results show that the proposed algorithm effectively solves an important problem about DLDA algorithm.

V. CONCLUSIONS

In this paper, we proposed an fast direct linear discriminant analysis algorithm, namely, FDLDA algorithm, which is highly efficient. The proposed method does not require the whole data matrix in main memory. This is desirable for large data sets. In addition, our theoretical analysis indicates that the computational complexity of the FDLDA algorithm is linear in the number of the data items in the training data set as the number of classes and the number of dimensions. It is the

calculated Q_{\perp} based on SVD that contributes to the efficiency of the FDLDA algorithm, which is not only shown by our theoretical analysis, but also strongly supported by our experimental results.

Our experiments on face database have shown that the accuracy achieved by the FDLDA algorithm is competitive with the ones achieved by Chen's DLDA. With efficiency and scalability, FDLDA algorithm is promising in real-time application involving extremely high-dimensional data.

REFERENCES

- [1] Samal A and Iyengar P A. (1992) Automatic recognition and analysis of human faces and facial expressions: a survey. *Pattern Recognition* 25(1): 65-77
- [2] W. Zhao, R. Chellappa, and J. Phillips (1999) Subspace linear discriminant analysis for face recognition. Technical Report, CS-TR4009, Univ. of Maryland
- [3] W.Zhao, R.Chellappa, P.J.Phillips, A.Rosenfeld. (2003) Face recognition: A literature survey. *ACM Computing Surveys*. 35(4): 395-458
- [4] K.Fukunaga (1990) Introduction to statistical pattern recognition. Academic Press, Boston, 2nd edition
- [5] S. R. Fisher (1936) The use of multiple measures in taxonomic problems. *Ann. Eugenics*. 7:179-188
- [6] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman (1997) Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*. 19 (7): 711-720
- [7] J. Yang, A.F. Frangi, J.Y. Yang, D. Zhang, Z. Jin (2005) KPCA plus LDA: A Complete Kernel Fisher Discriminant Framework for Feature extraction and Recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence* 27(2): 230-244
- [8] J. Yang, J.Y. Yang (2003) Why can LDA be performed in PCA transformed space?. *Pattern Recognition* 36:563-566
- [9] Q.Tian, M.Barbero, Z.Gu, S.Lee (1986) Image classification by the foley-sammon transform. *Opt. Eng.* 25(7):834-840
- [10] Z.Q.Hong, J.Y.Yang (1991) Optimal discriminant plane for a small number of samples and design method of classifier on the plane. *Pattern Recognition* 24(4):317-324
- [11] Y.Q. Cheng, Y.M.Zhuang, J.Y.Yang (1992) Optimal fisher discriminant analysis using the rank decomposition. *Pattern Recognition* 25(1):101-111.
- [12] Swets and J. Weng (1996) Using discriminant eigenfeatures for image retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 18(8):831-836.
- [13] L.F.Chen, H.Y.M.Liao, J.C.Lin, M.T.Ko, and G.J.Yu. A New LDA-based Face Recognition System Which Can Solve the Small Sample Size Problem, *Pattern Recognition*, 2000, 33(10):1713-1726.
- [14] J. Yang, H. Yu, W. Kunz (2000) An Efficient LDA Algorithm for Face Recognition. *International Conference on Automation, Robotics, and Computer Vision (ICARCV'2000)*, Singapore, December.
- [15] Yujie Zheng, Jingyu Yang, Jian Yang, Xiaojun Wu (2006) Effective classification image space which can solve small sample size problem. *Proc. Of. the 18th Int. Conf. on Pattern Recognition (ICPR'06)*, vol.3, pp.861-864.
- [16] Z. Wang, Q. Ruan, G. An (2016) Facial expression recognition using sparse local Fisher discriminant analysis, *Neurocomputing*, 2016(174): 756-766.
- [17] M. Kan, S. Shan, H. Zhang, S. Lao (2016) X Chen Multi-view Discriminant Analysis, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 38(1):808-821.
- [18] J. Yang, Z. Gu, N. Zhang, J. Xu (2014) Median-mean line based discriminant analysis, *Neurocomputing*, 123: 233-246.
- [19] Z. Q. Hong (1991) Algebraic feature extraction of image for recognition, *Pattern Recognition*, 24(3): 211-219.
- [20] W. Y Pang, N. H Yu, R. Zhang (2004) Fusion of SVD and LDA for face recognition, In *Proceedings of the International Conference on Image Processing*, Singapore, 2: 1417-1420.