An Improvement to Matrix-Based LDA

Chongyang Zhang and Jingyu Yang

Nanjing University of Science and Technology, Nanjing, China
zcy603@163.com

Abstract. The matrix-based LDA method is attracting increasing attention. Compared with classic LDA, this method can overcome the small sample size (SSS) problem. However, previous literatures neglect the fact that there are two available matrix-based LDA algorithms and usually use only one of the two algorithms to perform the experiment. By experimental analysis, this work point out the combination of the two available matrix-based LDA algorithms can obtain a better performance.

Keywords: feature extraction, LDA, biometric images.

1 Introduction

In the field of face recognition, one of the hottest branches of biometrics, Fisher discriminant analysis (FDA) has attracted much attention [1-8]. It has also been used in a variety of pattern recognition and computer vision problems. In the past years, a number of FDA methods such as Foley-Sammon linear discriminant analysis, Fisherfaces, uncorrelated linear discriminant analysis and 2DLDA have been proposed [9-17]. Besides these linear discriminant analysis methods, kernel discriminant analysis has also been developed as the nonlinear version of FDA [18-25].

Recently, the matrix-based LDA method is attracting increasing attention. Compared with classic LDA, this method can overcome the small sample size (SSS) problem. It also seems that in the matrix-based LDA method, the between-class scatter matrix and the within-class scatter matrix can be evaluated more accurately than the scatter matrices in the classic FDA owing to the lower dimension. However, previous literatures neglect the fact that there are two available matrix-based LDA algorithms and usually use only one of the two algorithms to perform the experiment.

We note that different discriminant analysis methods have different motivations. For example, Foley-Sammon linear discriminant analysis aims at generating orthogonal discriminant vectors, whereas uncorrelated linear discriminant analysis aims at obtaining uncorrelated feature components. Differing from both of Foley-Sammon linear discriminant analysis uncorrelated Foley-Sammon linear discriminant analysis, Xu et al. [9] proposed a discriminant analysis method that inherits the advantages of both uncorrelated linear discriminant analysis and Foley-Sammon linear discriminant analysis. It is also noticeable that kernel discriminant analysis has a better capability of capturing the complex features of samples.

H. Deng et al. (Eds.): AICI 2011, Part III, LNAI 7004, pp. 562–568, 2011.

Motivated by the reference [26], we formally present two available matrix-based LDA algorithms and propose to combine them for face recognition. The experimental results show that the combination algorithm can obtain a better performance.

The other parts of this paper are organized as follows: In Section 2 we describe two available matrix-based LDA algorithms. In Section 3 we present our scheme that combines the two available matrix-based LDA algorithms for face recognition. In Section 4 we show the experimental results and in Section 5 we provide a short conclusion.

2 Two Available Matrix-Based LDA Algorithms

In this section we will formally describe two available matrix-based LDA algorithms, respectively.

2.1 The First Matrix-Based LDA Algorithm

Suppose that there are L classes. Let A_j^i represent the j th training sample of the i th class and N_i denote the number of the training samples of the i th class. Let $\overline{A^i}$ denote the mean of the i th class and $\overline{A}$ denote the mean of all the training samples. For simplicity, we assume that each sample is a m by n matrix. The first matrix-based LDA algorithm defines S_b and S_t as follows:

$$S_b = \frac{1}{L}\sum_{i=1}^{L} N_i (A^i - \overline{A})(A^i - \overline{A})^T \tag{1}$$

$$S_t = \frac{1}{L}\sum_{i=1}^{L}\sum_{j=1}^{N_i} (A_j^i - A^i)(A_j^i - A^i)^T \,. \tag{2}$$

S_b and S_t are the so-called between-class and within-class scatter matrices, respectively. A_j^i stands for the j th training sample of the i th class. The first matrix-based LDA algorithm is based on the following eigen-equation:

$$S_b w = \lambda S_t w \,, \tag{3}$$

Suppose that the eigenvalues of Equation (3) is $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_m$ and the corresponding eigenvectors are $w_1, w_2, \ldots, w_m$, respectively. The first matrix-based LDA algorithm takes the first d eigenvectors, $w_1, w_2, \ldots, w_d$ as discriminant vectors and exploits the following equation to transform a sample A into a d by n matrix.

$$B = W^T A \,, \tag{4}$$

where $W = [w_1 ... w_d]$. B is referred to as the feature extraction result, of A, obtained using the first matrix-based LDA algorithm.

2.2 The Second Matrix-Based LDA Algorithm

The second matrix-based LDA algorithm defines the between-class and within-class scatter matrices as follows:

$$S_b^{'} = \frac{1}{L} \sum_{i=1}^{L} N_i (A^i - \overline{A})^T (A^i - \overline{A}) \tag{5}$$

$$S_t^{'} = \frac{1}{L} \sum_{i=1}^{L} \sum_{j=1}^{N_i} (A_j^i - A^i)^T (A_j^i - A^i). \tag{6}$$

The eigen-equation of this algorithm is as follows:

$$S_b^{'} w^{'} = \lambda S_t^{'} w^{'}, \tag{7}$$

Suppose that the eigenvalues of Equation (7) is $\lambda_1^{'} \geq \lambda_2^{'} ... \geq \lambda_n^{'}$ and the corresponding eigenvectors are $w_1^{'}, w_2^{'}, ..., w_m^{'}$. The second matrix-based LDA algorithm takes the first d eigenvectors, $w_1^{'}, w_2^{'}, ..., w_d^{'}$ as discriminant vectors and exploits the following equation to transform a sample A into an m by d matrix.

$$B^{'} = AW', \tag{8}$$

where $W^{'} = [w_1^{'} ... w_d^{'}]$. $B^{'}$ is referred to as the feature extraction result, of A, obtained using the second matrix-based LDA algorithm.

3 The Scheme to Combine the First and Second Matrix-Based LDA Algorithms

The scheme that combines the two available matrix-based LDA algorithms for face recognition is based a matching score level fusion strategy and works as follows: it first divides all of the samples into two parts, the training set and test set. Then it exploits the samples in the training set to produce the eigen-equation and to compute the discriminant vectors and perform feature extraction, which will be implemented for the first and second matrix-based LDA algorithms, respectively. Suppose that for a test sample A, the feature extraction results obtained using the first and second matrix-based LDA algorithms are B and $B^{'}$ respectively. Then our scheme calculates the distances between B and the feature extraction results, of all the training samples, obtained using the first matrix-based LDA algorithm and denotes these distances by $dist1_j^i$. Specifically, $dist1_j^i$ stands for the distance between B and the feature extraction result of the j th training sample of the i th class. Our scheme denotes the distances between $B^{'}$ and the feature extraction results, of all the training samples, obtained using the second matrix-based LDA algorithm by $dist2_j^i$.

Our scheme then calculates $dist_j^i = u_1 * dist1_j^i + u_2 * dist2_j^i$. u_1, u_2 are the weighting coefficients. It is clear that our scheme considers $dist_j^i$ as the distance between the test sample and the j th training sample of the i th class. Our scheme identifies the training sample that has the minimum distance with the test sample and assumes that the test sample is from the same class as the identified training sample. It is clear that our scheme uses a matching score level fusion strategy and treats the distance as the matching score. As a larger distance means a low similarity, our scheme indeed classifies the test sample into the class of the training sample that has the maximum matching score. This is different from the conventional matching score level fusion strategy.

4 Experiments

The ORL face database contains a set of face images taken between April 1992 and April 1994 at the lab. The database was used in the context of a face recognition project carried out in collaboration with the Speech, Vision and Robotics Group of the Cambridge University Engineering Department [27]. There are ten different images of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement) [27]. The files are in PGM format, and can conveniently be viewed on UNIX (TM) systems using the 'xv' program. The size of each image is 92x112 pixels, with 256 grey levels per pixel. The images are organised in 40 directories (one for each subject), which have names of the form sX, where X indicates the subject number (between 1 and 40). In each of these directories, there are ten different images of that subject [27]. Figure 1 shows some face images from the ORL database.

In order to implement the algorithm computationally efficient, we first resized each face image into 46x56 pixels using the downsampling algorithm proposed in [28]. We respectively took the first five and four samples of each subject as training samples and used the other samples as test samples. Figures 2 and 3 show the rates of the classification errors obtained using our scheme, the first and second matrix-based LDA algorithms. In these two figures, the horizontal ordinate shows the number of the discriminant vectors used for feature extraction. The vertical ordinate shows the rate of the classification errors. In the experiment on Figure 2, weighting coefficients u_1, u_2 were set to 0.5 and 0.5, respectively. In the experiment on Figure 3, u_1, u_2 were set to 0.4 and 0.6, respectively. We see that our scheme, the first and second matrix-based LDA algorithms all obtain their lowest rates of classification errors when using a small number of discriminant vectors. Moreover, the lowest rate of classification errors of our scheme is less than those of the first and second matrix-based LDA algorithms. For example, when the first five samples of each subject were used as training samples and the other samples were used as test samples, the lowest rate of classification errors of our scheme is 8.5%, whereas mples and the other samples were used as test samples, the lowest rates of classification errors of the first and second matrix-based LDA algorithms are 14.5% and 11%, respectively. The experimental results also show that the two available matrix-based LDA algorithms have a clear difference in classification performance.

Fig. 1. Some face images from the ORL database. The images in the first, second and third rows are face images of three different subjects, respectively.

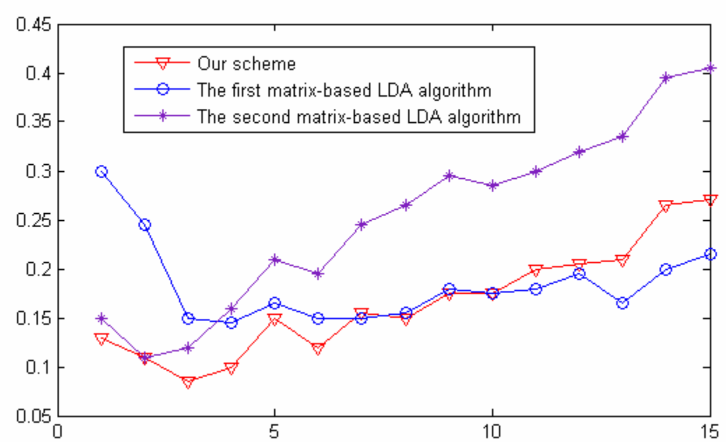

Fig. 2. The rate of the classification errors obtained using our scheme, the first and second matrix-based LDA algorithms. The horizontal ordinate shows the number of the discriminant vectors used for feature extraction. The vertical ordinate shows the rate of the classification errors. The first five samples per subject were used as training samples and the other samples were used as test samples.

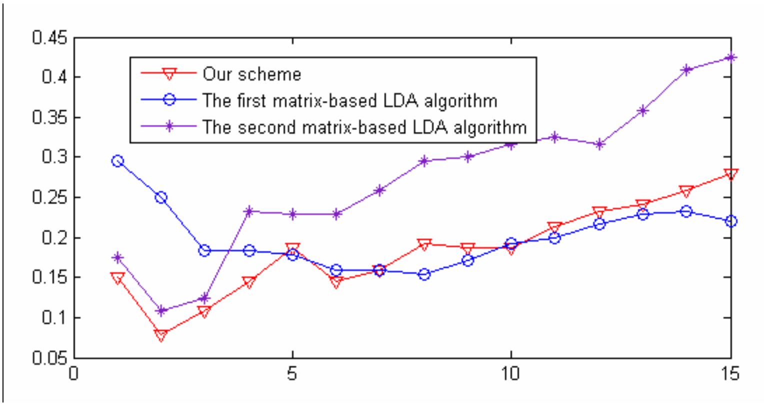

Fig. 3. The rate of the classification errors obtained using our scheme, the first and second matrix-based LDA algorithms. The horizontal ordinate shows the number of the discriminant vectors used for feature extraction. The vertical ordinate shows the rate of the classification errors. The first four samples per subject were used as training samples and the other samples were used as test samples.

5 Conclusion

Compared with classic LDA, the matrix-based LDA method can overcome the small sample size (SSS) problem. However, previous literatures neglect the fact that there are two available matrix-based LDA algorithms and usually use only one of the two algorithms to perform the experiment. By experimental analysis, this work point out the two available matrix-based LDA algorithms might have a difference in performance and the combination of them can obtain a better performance.

References

1. Turk, M., Pentland, A.: Eigenfaces for Recognition. Journal of Cognitive Neurosicence 3(1), 71–86 (1991)
2. Turk, M., Pentland, A.: Face Recognition Using Eigenfaces. In: Proc. IEEE Conf. On Computer Vision and Pattern Recognition, pp. 586–591 (1991)
3. Beymer, D., Poggio, T.: Face recognition from one example view. In: Proceedings of the International Conference on Computer Vision, pp. 500–507 (1995)
4. Xu, Y., Yao, L., Zhang, D.: Improving the Interest Operator for Face Recognition. Expert System With Applications 36(6), 9719–9728 (2009)
5. Bowyer, K.W., Kyong, C., Flynn, P.: A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition. Comput. Vision Image Understanding 101(1), 1–15 (2006)
6. Xu, Y., Feng, G., Zhao, Y.: One improvement to two-dimensional locality preserving projection method for use with face recognition. Neurocomputing 73, 245–249 (2009)
7. Zhang, D., Song, F., Xu, Y., Liang, Z.: Advanced pattern recognition technologies with applications to biometrics. Medical Information Science Reference, IGI Global (2009)

8. Xu, Y., Zhong, A., Yang, Zhang, D.: LPP solution schemes for use with face recognition. Pattern Recognition, doi:10.1016/j.patcog.2010.06.016
9. Xu, Y., Yang, J.-Y., Jin, Z.: A novel method for Fisher discriminant Analysis. Pattern Recognition 37(2), 381–384 (2004)
10. Yu, H., Yang, J.: A direct LDA algorithm for high-dimensional data—with application to face recognition. Pattern Recognition 34(10), 2067–2070 (2001)
11. Chen, S., Zhu, Y., Zhang, D., Yang, J.-Y.: Feature extraction approaches based on matrix pattern: MatPCA and MatFLDA. Pattern Recognition Letters 26(8), 1157–1167 (2005)
12. Xu, Y., Zhang, D.: Represent and fuse bimodal biometric images at the feature level: complex-matrix-based fusion scheme. Optical Engineering 49(3) (2010), doi:10.1117/1.3359514
13. Belhumeur, P.N., Hespanha, J.P., Kriengman, D.J.: Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Machine Intell. 19(7), 711–720 (1997)
14. Foley, D.H., Sammon Jr., J.W.: An optimal set of discriminant vectors. IEEE Trans. Comput. 24(3), 281–289 (1975)
15. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET evaluation methodology for face-recognition algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(10), 1090–1104 (2000)
16. Jing, X.-Y., Zhang, D., Jin, Z.: Improvements on the uncorrelated optimal discriminant vectors. Pattern Recognition 36(8), 1921–1923 (2003)
17. Xu, Y., Yang, J.-Y., Jin, Z.: A novel method for Fisher discriminant Analysis. Pattern Recognition 37(2), 381–384 (2004)
18. Scholkopf, B., Smola, A.: Learning With Kernels. MIT Ppress, Cambridge (2002)
19. Scholkopf, B., Smola, A., Muller, K.R.: Nonlinear Component Analysis as a Kernel Eigenvalue Problem. Neural Computation 10(5), 1299–1319 (1998)
20. Xu, Y., Yang, J.-Y., Yang, J.: A reformative kernel Fisher discriminant analysis. Pattern Recognition 37(6), 1299–1302 (2004)
21. Mika, S., Rätsch, G., Weston, J., Schölkopf, B., Müller, K.-R.: Fisher discriminant analysis with kernels. In: Neural Networks for Signal Processing IX, pp. 41–48. IEEE, New York (1999)
22. Muller, K.-R., Mika, S., Rätsch, G., Tsuda, K., Schölkopf, B.: An introduction to kernel-based learning algorithms. IEEE Trans. Neural Network 12(1), 181–201 (2001)
23. Steve, B., Kian, L.: Nonlinear Fisher discriminant analysis using a minimum squared error cost function and the orthogonal least squares algorithm. Neural Networks 15(1), 263–270 (2002)
24. Cristianini, N., Shawe-Taylor, J.: Kernel Methods for Pattern Analysis. Cambridge Univ. Press, New York (2004), [2]
25. Scholkopf, B., Mika, S., Burges, C., Knirsch, P., Muller, K.R., Ratsch, G., Smola, A.: Input Space versus Feature Space in Kernel-Based Methods. IEEE Trans. on Neural Networks 10(5), 1000–1017 (1999)
26. Xu, Y., Zhang, D., Yang, J., Yang, J.-Y.: An approach for directly extracting features from matrix data and its application in face recognition. Neurocomputing 71(10-12), 1857–1865 (2008)
27. `http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html`
28. Xu, Y., Jin, Z.: Down-sampling face images and low-resolution face recognition. In: The Third International Conference on Innovative Computing, Information and Control, Dalian, China, June 18-20, pp. 392–395 (2008)