Resampling LDA/QR and PCA+LDA for face recognition

Jun Liu, Songcan Chen

Department of Computer Science and Engineering,
Nanjing University of Aeronautics and Astronautics,
Nanjing 210016, P.R. China
maurice_nt@sohu.com, s.chen@nuaa.edu.cn

Abstract. Principal Component Analysis (PCA) plus Linear Discriminant Analysis (LDA) (PCA+LDA) and LDA/QR are both two-stage methods that deal with the small sample size (SSS) problem in traditional LDA. When applied to face recognition under varying lighting conditions and different facial expressions, neither method may work robustly due to limited number of training samples for each class in the training set. Recently, resampling, a technique that generates multiple subsets of samples from the training set, has been successfully employed to improve the classification performance of the PCA+LDA classifier. In this paper, stimulated by such success, we propose a resampling LDA/QR method to improve LDA/QR's performance. Furthermore, by analyzing the difference between LDA/QR and PCA+LDA and taking advantage of such difference, we incorporate LDA/QR and PCA+LDA in a combined framework by resampling for face recognition. Experimental results on AR dataset show that 1) resampling LDA/QR yields significantly higher classification performance than the original LDA/QR, and 2) resampling LDA/QR and resampling PCA+LDA in a combined framework further improves the classification compared to either resampling LDA/QR or resampling PCA+LDA.

1 Introduction

Linear Discriminant Analysis (LDA) [1-3] is a popular feature extraction in pattern recognition. It searches for a set of projection vectors onto which the data points of the same class are close to each other while requiring data points of different class to be far from each other, in other words, it calculates the projection matrix W that maximizes the Fisher's Linear Discriminant criterion as follows:

$$J_{FLD}(W_{opt}) = \arg\max_{W} \left|W^T S_b W\right| \Big/ \left|W^T S_w W\right| \tag{1}$$

here S_w is the within-class scatter matrix and S_b is the between-scatter matrix. It has been proved that if S_w is a non-singular matrix then the ratio of (1) is maximized when the column vectors of W are the eigenvectors of $S_w^{-1}S_b$. Unfortunately, in recognition task such as face recognition, S_w is typically singular, due to the fact that the number of the samples is much smaller than the dimension of sample space, i.e., the so-called Small Sample Size (SSS) problem.

Among the many methods that address this singularity problem, Principal Component Analysis (PCA) plus LDA (PCA+LDA) [1-3] and LDA/QR [4-5] are both methods that perform feature extraction in two sequential stages: in PCA+LDA, or Fisherface, the face samples are projected to a PCA [6] subspace in the first stage, and then LDA is applied secondly; in LDA/QR, the face samples are firstly projected to the range space of between-class matrix S_b through QR-decomposition [14], followed by LDA in the second stage. When applied to face recognition under varying lighting conditions and different facial expressions, neither methods may work robustly under relatively great facial variance, due to limited number of training samples for each class in the training set and the consequent biased estimates of both the within-class matrix S_w and the between-class matrix S_b.

To improve the performance of weak classifiers, a number of approaches have been studied in literature, including the bootstrapping-aggregating (bagging) method by Breiman [7], the boosting method by Freund and Schapire [8], random subspace method by Ho [9] and the resampling method by Lu and Jain [10]. Recently, Wang and Tang integrated bagging Null space LDA [12] and random subspace Fisherface for face recognition in [11], and Lu and Jain applied resampling PCA+LDA (R-PCA+LDA) for face recognition in [10]. However, to the best of our knowledge, similar techniques have not been applied to improve the classification performance of the LDA/QR classifier.

In this paper, we intend to improve the classification performance of LDA/QR utilizing the resampling technique, and propose the resampling LDA/QR (R-LDA/QR) method, which is based on the following three considerations: 1) LDA/QR has significantly lower costs in time and space than PCA+LDA [5], then if R-LDA/QR works pretty well, it will offer practitioners a more efficient tool compared to R-PCA+LDA; 2) As stated in [5], LDA/QR may encounter the problem of centroid sensitivity, which leads to a weak classifier. This phenomenon will typically take place when face images are under great facial variance, consequently, LDA/QR needs improvement; 3) in LDA based methods such as PCA+LDA and LDA/QR, both the within-class matrix S_w and the between-class matrix S_b should be utilized, then resampling rather than bagging should be employed [10].

Furthermore, PCA+LDA is quite different from LDA/QR, especially in their corresponding first stages, i.e., face samples are projected to an unsupervised PCA subspace in PCA+LDA, different from the supervised S_b's range space in LDA/QR. As a result, the projection matrices by PCA+LDA and LDA/QR may be quite different. Fig. 1 shows the degrees of the principal angels between the subspaces spanned respectively by their projection matrices on AR [13] face dataset (see the description of this dataset in section 5), which verifies experimentally that the projection matrices derived respectively by PCA+LDA and LDA/QR are quite different. Taking advantage of such difference, we incorporate PCA+LDA and LDA/QR in a combined framework by resampling for face recognition, and expect to gain improved classification performance.

The rest of this paper is organized as follows. In section 2, the LDA based methods such as PCA+LDA and LDA/QR are reviewed. In section 3, we briefly introduce the resampling technique and the R-PCA+LDA method. In section 4, we give our R-LDA/QR method and present the combined framework that incorporates PCA+LDA and LDA/QR by resampling technique. In section 5, experiments on AR face dataset are performed, and the effectiveness of the R-LDA/QR method and the combined framework that integrates R-PCA+LDA and R-LDA/QR are experimentally verified. In section 6, we draw a conclusion to this paper.

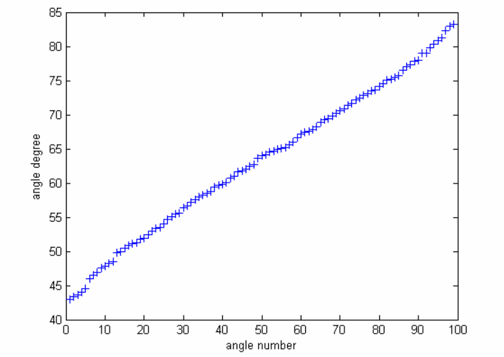


Fig. 1. Degrees of the principal angles between the subspaces respectively spanned by PCA+LDA and LDA/QR's projection matrices on AR face dataset.

2 Lineare Discriminant Analysis

In this section, we briefly review the LDA based methods such as PCA+LDA and LDA/QR. Let the training set be composed of C classes, where each class contains N_i samples. Then there a total of $N=N_1+N_2+\ldots+N_C$ samples. Let x^i_j be a d-dimensional column vector which denotes the j-th sample from the i-th class. The between-class scatter matrix S_b, the within-class scatter matrix S_w and the total scatter matrix S_t are respectively defined as:

$$S_w = \frac{1}{N}\sum_{i=1}^{C}\sum_{j=1}^{N}(x^i_j - m_i)(x^i_j - m_i)^T \tag{2}$$

$$S_b = \frac{1}{N}\sum_{i=1}^{C} N_i(m_i - m)(m_i - m)^T = H_b H_b^T \tag{3}$$

$$S_t = S_w + S_b \tag{4}$$

where m_i is the mean of samples in the i-th class, m the mean of all the samples and H_b is:

$$H_b = \frac{1}{\sqrt{N}}\left[\sqrt{N_1}(m_1 - m), \cdots \sqrt{N_C}(m_C - m)\right] \tag{5}$$

2.1 PCA+LDA

LDA wants to calculate the projection matrix by maximizing (1). However, when the SSS problem takes place, S_w will be typically singular. To remedy this problem, PCA+LDA, or Fisherface first projects face samples to the PCA subspace spanned by the N-C largest eigenfaces. Such PCA subspace, denoted as W_{PCA}, can be solved by an eigen-analysis to the total scatter matrix defined in (4). In the PCA subspace, the within-class matrix scatter and the between-class scatter matrices can be written as:

$$S_w^L = W_{PCA}^T S_w W_{PCA} \tag{6}$$

$$S_b^L = W_{PCA}^T S_b W_{PCA} \tag{7}$$

PCA+LDA then calculates the eigenvectors of $(S^L{}_w)^{-1} S^L{}_b$ corresponding to the largest C-1 eigenvalues. Put these C-1 eigenvectors as column vectors in W_{LDA}, $P^L=W_{PCA}W_{LDA}$ is the projection matrix yielded by PCA+LDA.

2.2 LDA/QR

Similar to PCA+LDA, LDA/QR is also a two-stage method. In the first stage, it aims to compute the optimal transformation G that maximizes the between-class distance, i.e.,

$$G = \arg\max_{G^T G = I} trace(G^T S_b G) \tag{8}$$

which can be solved by applying QR-decomposition [14] to H_b. Let $H_b=QR$ be the QR-decomposition on H_b, where Q is a d×(C-1) matrix with orthonormal columns and R a (C-1) ×C upper triangular matrix. It has been proven in [5] that $G=Q$ provides a solution to (8). Now, LDA/QR moves to the second stage to minimize the within-class distance in the space spanned by Q (or equivalently in the range space of S_b), where the within-class and between-class scatter matrices can be formulated as:

$$S_w^R = Q^T S_w Q \tag{9}$$

$$S_b^R = Q^T S_b Q \tag{10}$$

Put the eigenvectors of $(S^R_b)^{-1}S^R_w$ corresponding to the smallest C-1 eigenvalues as column vectors in W, $P^R=QW$ is the projection matrix yielded by LDA/QR.

At the end of this section, it should be noted that LDA/QR is much efficient than PCA+LDA for calculating the projection matrix, due to that: 1) in the first stage, the former employs QR-decomposition on a $d\times C$ matrix to calculate the range space, while the latter employs SVD [14] (keeping in mind that SVD is computationally more expensive than QR-decomposition) on a $N\times N$ matrix to calculate the PCA subspace; and 2) in the second stage, the former manipulates on an eigenvalue problem of the size $(C\text{-}1)\times(C\text{-}1)$ while the latter on an eigenvalue problem of the size $(N\text{-}C)\times(N\text{-}C)$. For detail, interested readers can refer to [5].

3 Resampling and resampling PCA+LDA

When such LDA based method as PCA+LDA is applied for face recognition under varying lighting conditions and different facial expressions, it may work weakly or not robustly due to limited number of training samples for each class in the training set. To improve the classification performance of PCA+LDA method under different facial variations, Lu and Jain proposed the resampling technique in [10] to improve the classification performance of the PCA+LDA classifier. Observing that both the intra- and inter-class information (namely, S_b and S_w) should be utilized, the resampling technique is not to sample the whole training set, but randomly samples each class, subject to the following two conditions: [10]

1) The number of sampled images for each subject in the subset is equal or as equal as possible;

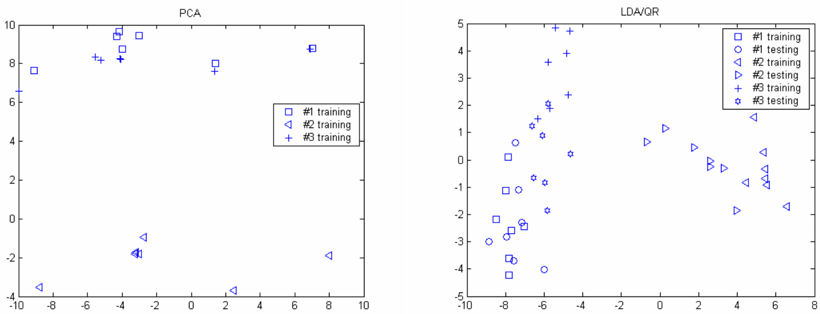
2) Each class is sampled in term of a uniform distribution.

After multiple subsets are generated from the whole training set by the resampling technique, the R-PCA+LDA method then trains a PCA+LDA classifier on each subset. When classifying an unknown test sample, it is presented to these trained multiple classifiers, and the classification results of all the classifiers are integrated to determine the class label by such integration strategies as majority voting and sum rule [10].

4 Proposed methods

4.1 Resampling LDA/QR for face recognition

In the end of section 2, we have briefly analyzed that LDA/QR is more efficient than PCA+LDA. However, similar to PCA+LDA, it will work weakly under great facial variations, as can be observed from Figure 2, where we show the samples of three classes from AR face dataset projected by PCA and LDA/QR respectively. From Fig. 2 (a), we can see that the samples from the same class are not compact, e.g., the three samples from class 1 (denoted by square) are far away from the other four samples from the same class (denoted also by square) which cluster relatively compactly. Such phenomenon is due to the great facial variations within the person. In fact, these three samples are indeed under varying lighting conditions, while the other four are under the same lighting condition. Fig. 2 (b) illustrates the distributions of both the training and testing samples in the LDA/QR projected subspace, from which we can see that, although the samples from the same class in the training set group more compactly than those in Fig (a), the three testing samples (denoted by six-pointed star) from class 3 are misclassified as class 1.



(a) (b)

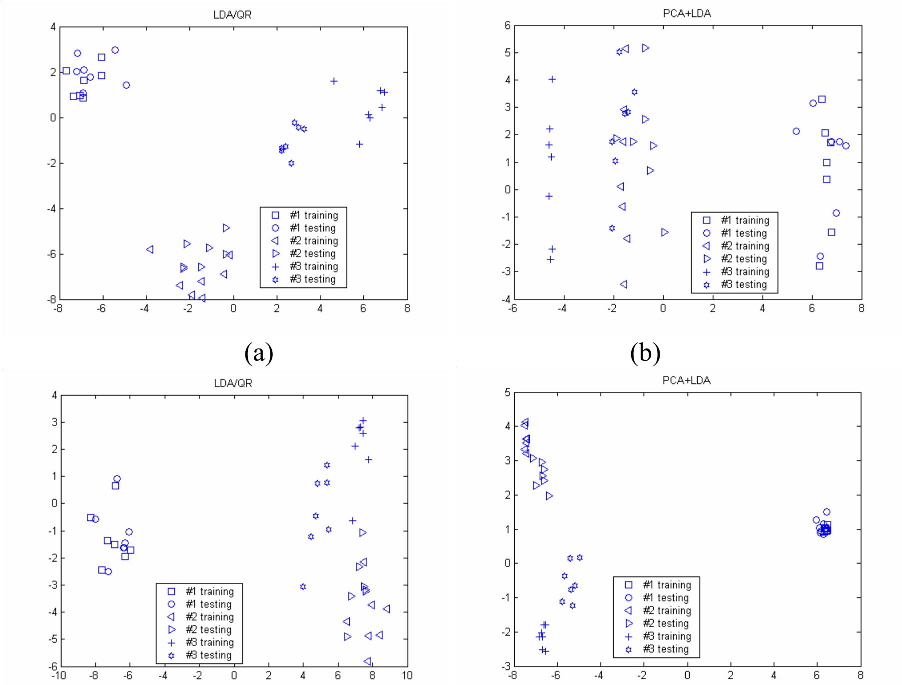
Fig. 2. Two dimensional display of samples from three classes in the AR dataset: (a) the training samples of the three classes are projected by the leading two eigenvectors of the total scatter matrix of all the training samples, (b) both the training and testing samples of the three classes are projected by the projection matrix of LDA/QR.

In order to improve the classification performance of LDA/QR classifier, we employ the resampling technique to generate multiple subsets, train a LDA/QR classifier on each subset, and integrate the classification results of these multiple classifiers using majority voting for classifying a given unknown sample. The effectiveness of the proposed method will be verified experimentally in section 5, which shows that R-LDA/QR can achieve significantly higher classification accuracy over LDA/QR using all the training samples and meanwhile comparable or even better classification accuracy than R-PCA+LDA. Thanks to the fact that LDA/QR is more efficient than PCA+LDA, naturally the proposed R-LDA/QR here will be more important for practitioners.

4.2 Incorporating LDA/QR and PCA+LDA by resampling technique for face recognition

After successfully applying the resampling to improve the classification performance of the LDA/QR method, a natural question is, “Can we benefit from integrating R-PCA+LDA and R-LDA/QR for face recognition?” From the theory of classifier combination [15], diversity plays a key factor to boost the generalization or classification performance of the combined classifiers. In the following, we give a qualitative diversity analysis for PCA+LDA and LDA/QR to be combined.

In the introduction, we have show that LDA/QR and PCA+LDA are different from each other in deriving the projection matrix, especially in their respective first stages, namely, the supervised range space of S_b versus the unsupervised PCA subspace. And also in the introduction, we exhibit in Fig. 1 that the projection matrices yielded respectively by LDA/QR and PCA+LDA are quite different. More specifically, about 60 (out of 99) principal angles between their projection matrices are over 60 degree, and the minimum principal angle is 42.9 degree.



(a) (b)

(c) (d)

Fig. 3. Two dimensional display of samples from three classes respectively by LDA/QR and PCA+LDA, where pictures in the same row correspond to the samples from same three classes: (a) and (b) illustrate the case that PCA+LDA classifies some testing samples wrongly while LDA/QR discriminates the three classes correctly, (c) and (d) show the case that LDA/QR some testing samples wrongly while PCA+LDA discriminates correctly.

Furthermore, we will show that there are cases that PCA+LDA classifies the testing samples wrongly and meanwhile LDA/QR classifies correctly, and vice versa. We take the samples of three classes from the AR dataset as analytical example, and report the experimental results in Fig. 3. Fig. 3(b) shows that PCA+LDA discriminates the testing samples from class 3 (denoted by six-pointed star) wrongly, however, LDA/QR discriminates the same testing samples correctly in Fig. 3 (a). Similarly, Fig. 3(c) shows the case that LDA/QR classifies some samples wrongly and meanwhile PCA+LDA classifies correctly in Fig. 3(d).

From the above discussion, we know that PCA+LDA and LDA/QR are both different and complementary to each other to some extent, thus have diversity. Then they can be integrated in a combined framework for face recognition. Fig. 4 shows the framework proposed, where T_i corresponds to the i-th generated subset by resampling technique, C^L_i and C^R_i are the accordingly built PCA+LDA and LDA/QR classifiers. In classifying an unknown sample, the classification results of all the classifiers are integrated using majority voting to determine its class label.

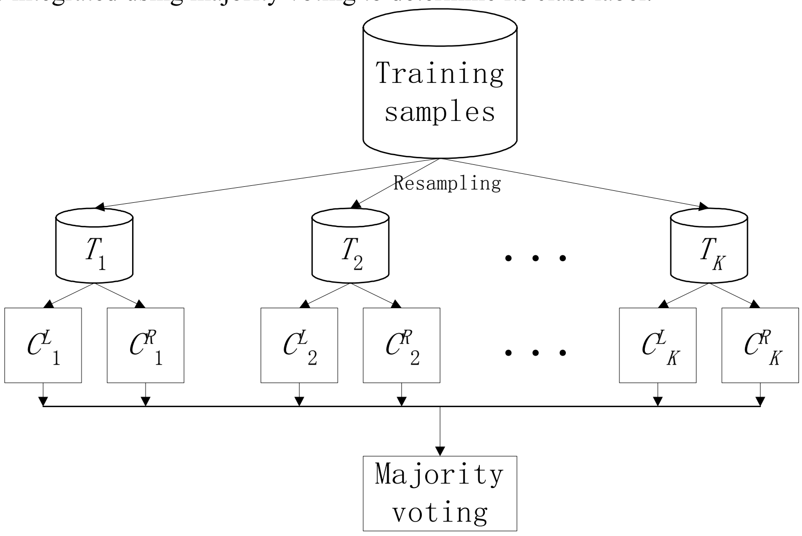


Fig. 4. Combined framework by integrating R-PCA+LDA and R-LDA/QR for face recognition.

5 Experiments

To evaluate the effectiveness of the proposed R-LDA/QR and the combined framework, we carry out experiments on AR face dataset [13]. This face dataset consists of over 3200 images of frontal images of faces of 126 subjects. Each subject has 26 different images which were grabbed in two different sessions separated by two weeks, 13 images in each session were recorded. For the 13 images, the first one is of neutral expression, the second to the fourth are of smile, anger and scream expression, the others are either light or scarf variation. In our experiments here, we use the 1400 gray level images from 100 objects, where each object has 14 images. More specifically, we use the first seven images from the first session to construct the training set, and the first seven images from the second session for testing. Fig. 5 illustrates one image from AR dataset in our experiments. The 1400 images are preprocessed by Martinez [13] with a resolution of 165×120. Here, for computational convenience, we resize them to 66×48 and the gray level values are rescaled to [0 1].

As for experimental setting, we generate 60 subsets from the whole training set, namely, K equals 60 in Figure 4. We dynamically integrate the classification results of the first k (k=1,2, …, K) classifiers through majority voting, and denote the according classification accuracy as Acc_k. When reporting the classification accuracy of a given method's final result (e.g., R-PCA+LDA, R-LDA/QR, and the combined framework by R-PCA+LDA and R-LDA/QR), the mean result of Acc_{41} to Acc_{60} is used and meanwhile the corresponding standard variance is also reported.

When using the resampling technique to generate multiple subsets, we randomly select M samples from each class, where M ranges from 3 to 5.

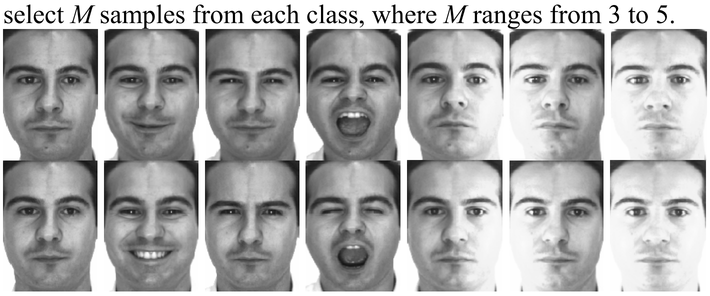


Fig. 5. Images from one person in the AR face dataset. Images in the first row are from the first session and used for training, while images in the second row are from the second session and used for testing.

5.1 Resampling LDA/QR

In this subsection, we report the classification performance of R-LDA/QR classifier. And for comparison, the classification results of R-PCA+LDA are also given.

The results are given in the left column of Fig. 6, from which we can get the following three observations:

1). R-LDA/QR significantly improves the classification performance of the original LDA/QR which operates on the whole training set. More specifically, when M is set to 3, R-LDA/QR has classification accuracy of 89.2%, 3.5 percentages higher than 85.7% of LDA/QR. Similar observation can also be obtained when M is set to 4 or 5.

2). R-LDA/QR can achieve comparable or even higher classification accuracies as R-PCA+LDA. Then considering the fact that LDA/QR is much more efficient than PCA+LDA, the proposing of R-LDA/QR is important, since R-LDA/QR offers practitioners a both efficient and effective classification tool.

3). The LDA/QR classifier in each subset is relatively weak. However, R-LDA/QR can achieve good classification performance. We attribute these to: a) LDA/QR is of centroid sensitivity [5], and resampling leads to more biased estimate of the class centroid than making use of all the training samples. As a result, the classification performance of each LDA/QR classifier built on the resampled subset deteriorates; however b) resampling also brings such an important advantage that the LDA/QR built on each subset is diverse and complementary to each other, which leads to R-LDA/QR's improved classification accuracy over LDA/QR.

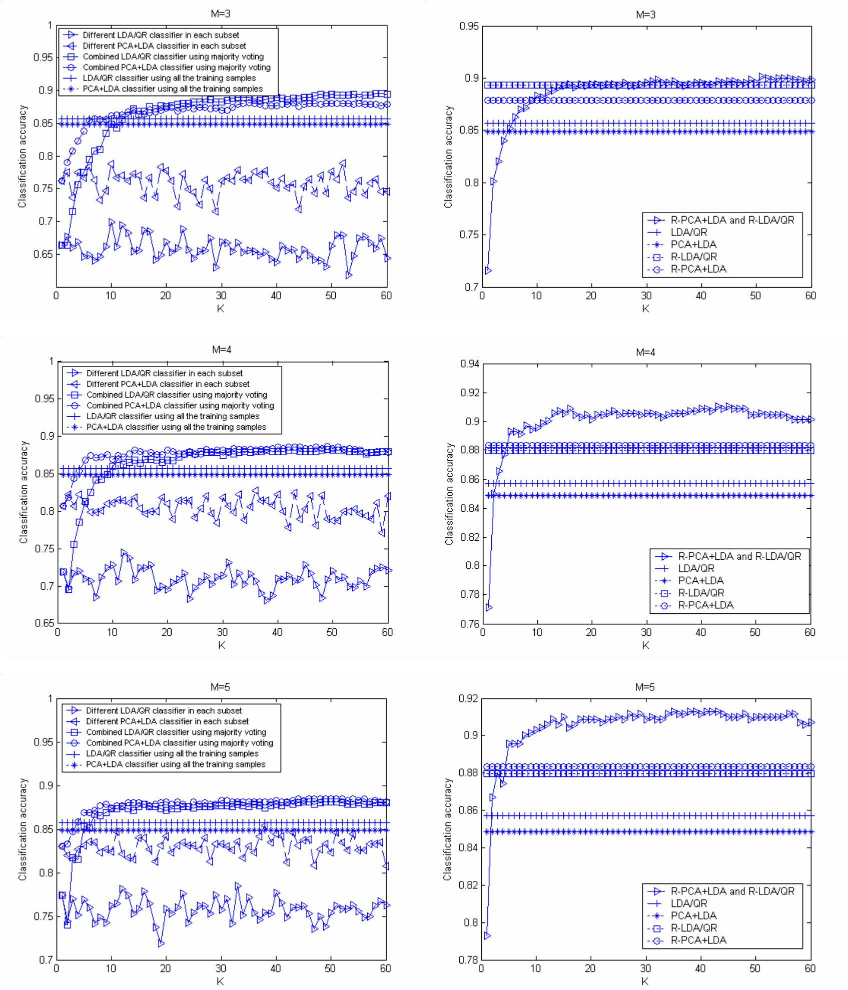


Fig. 6. Classification accuracies of R-LDA/QR, R-PCA+LDA and R-LDA/QR plus R-PCA+LDA

5.2 Integrating resampling LDA/QR and resampling PCA+LDA for face recognition

As described in section 4.2, LDA/QR and PCA+LDA are different and complementary to some extent. In this section, we will report the classification performance of combining LDA/QR and PCA+LDA in a combined framework by the resampling technique (R-LDA/QR plus R-PCA+LDA). The experimental results are reported in the right column of Fig. 6, from which we can see R-LDA/QR plus R-PCA+LDA further improves the classification performance when compared to LDA/QR, PCA+LDA, R-LDA/QR and R-PCA+LDA.

Table 1. Classification accuracy of different methods

	M		
	3	4	5
resampling LDA/QR	89.2[a] (2.81[b])	88.0(1.28)	88.0(1.25)
resampling PCA+LDA	87.9(1.25)	88.3(2.74)	88.3(1.22)
R-LDA/QR plus R-PCA+LDA	89.8(2.04)	90.6(3.01)	91.0(1.91)
LDA/QR	85.7		
PCA+LDA	84.9		

[a]The classification accuracy (%) for R-LDA/QR, R-PCA+LDA and R-LDA/QR plus R-PCA+LDA reported here are the mean values of Acc_{41} to Acc_{60}

[b]Value in the parenthesis ($\times 10^{-3}$) is the according standard derivation of Acc_{41} to Acc_{60}

For convenience of comparing R-LDA/QR plus R-PCA+LDA and other methods, we list the classification accuracies of all the methods in Table 1. When M equals 4, R-LDA/QR plus R-PCA+LDA yields a classification accuracy of 90.6%, 2.6, 2.3, 4.9 and 5.7 percentages respectively higher than R-LDA/QR, R-PCA+LDA, LDA/QR and PCA+LDA. Similar observation can also be made when M equals 3 or 5, which shows that R-LDA/QR plus R-PCA+LDA is a powerful combined framework that integrates R-LDA/QR and R-PCA+LDA for face recognition.

6 Conclusion

In this paper, we introduce the resampling technique to improve the classification performance of LDA/QR classifier, and propose a new R-LDA/QR classifier. Experimental results on AR face dataset with variations on both lighting and facial expression changes show that: 1) R-LDA/QR can achieve significantly higher classification accuracy than LDA/QR which operates on the whole training set, and meanwhile 2) R-LDA/QR can achieve comparable or even higher classification accuracy than R-PCA+LDA classifier proposed in [10]. Note that LDA/QR is more efficient than PCA+LDA, hence R-LDA/QR will offer practitioners a powerful and efficient classifier.

By analyzing that LDA/QR and PCA+LDA are different and complementary to some extent and taking advantage of such difference and complement, we propose a combining framework that integrates LDA/QR and PCA+LDA by resampling technique for face recognition. Experimental results on the AR face dataset show that R-LDA/QR plus R-PCA+LDA yields higher classification accuracy respectively than R-LDA/QR, R-PCA+LDA, LDA/QR, and PCA+LDA..

Acknowledgement

We thank Natural Science Foundation of China under Grant Nos. 60473035 for support.

References

1. Swets, D.L., and Weng, J., Using discriminant eigenfeatures for image retrieval, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 18, no. 8, pp. 831-836, August 1996.
2. Belhumeur, P.N., Hespanha, J.P., and Kriegman, D.J., Eigenfaces vs. fisherfaces: Recognition using class specific linear projection, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp.711-720, 1997.
3. Liu, C., and Wechsler, H., Robust Coding Schemes for Indexing and Retrieval from Large Face Databases, IEEE Transactions on Image Processing, vol.9, no.1, pp. 132-136, 2000.
4. Ye, J., Li, Q., LDA/QR: an efficient and effective dimension reduction algorithm and its theoretical foundation, Pattern Recognition, vol. 37, no. 4, pp. 851-854, 2004.
5. Ye, J., Li, Q., A two-stage Linear Discriminant Analysis via QR-Decomposition, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 27, no. 6, pp.929-941, 2005.
6. Turk, M.A., and Pentland, A.P., Face Recognition Using Eigenfaces, Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 586-591, 1991.
7. Breiman, L., Bagging predictors, Machine Learning, 24(2):123–140, 1996.
8. Freund, Y., and Schapire, R. E., Experiments with a new boosting algorithm, Proc. International Conference on Machine Learning, pp. 148–156, 1996
9. Ho, T. K., The random subspace method for constructing decision forests, IEEE Trans. Pattern Analysis and Machine Intelligence, 20(8):832–844, 1998.
10. Lu, X., and Jain, A. K., Resampling for Face Recognition, Proc. of AVBPA, 2003.
11. Wang, X., and Tang, X., Random sampling LDA for face recognition, Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, 2004.
12. Chen, L. F., Liao, H. Y. M., Ko, M. T., Lin, J-C, and Yu, G. J., A new LDA-based face recognition system which can solve the small sample size problem, Pattern Recognition, vol. 33, pp. 1713-1726, 2000.
13. Martinez, A.M., and Benavente, R., The AR Face Database, CVC Technical Report #24, June 1998.
14. Golub, G.H., Loan, C.F.V., Matrix Computations, 3rd Edition, The Johns Hopkins University Press, Baltimore, MD, USA, 1996.
15. Dietterich, T. G., Machine Learning Research: Four Current Directions, AI Magazine, 18(4), 97-136, 1997.