Using Random Subspace to Combine Multiple Features for Face Recognition

Xiaogang Wang and Xiaoou Tang

Department of Information Engineering
The Chinese University of Hong Kong
Shatin, Hong Kong
{xgwang1, xtang}@ie.cuhk.edu.hk

Abstract

LDA is a popular subspace based face recognition approach. However, it often suffers from the small sample size problem. When dealing with the high dimensional face data, the LDA classifier constructed from the small training set is often biased and unstable. In this paper, we use the random subspace method (RSM) to overcome the small sample size problem for LDA. Some low dimensional subspaces are randomly generated from face space. A LDA classifier is constructed from each random subspace, and the outputs of multiple LDA classifiers are combined in the final decision. Based on the random subspace LDA classifiers, a robust face recognition system is developed integrating shape, texture, and Gabor wavelet responses. The algorithm achieves 99.83% accuracy on the XM2VTS database.

1. Introduction

Face recognition has drawn more and more attention in recent years. LDA or Fisherface [2][3] is a popular subspace based face recognition approach. According to the Fisher criteria, LDA determines a set of projection vectors maximizing the between-class scatter matrix and minimizing the within-class scatter matrix in the projective feature space. However, when dealing with the high dimensional face data, LDA often suffers from the small sample size problem. Since usually there are only a few samples in each face class for training, the within-class scatter matrix is not well estimated and may become singular. So the LDA classifier is often biased and sensitive to slight changes of the training set.

To address this problem, a two-stage PCA+LDA approach, i.e. Fisherface [2] is proposed. Using PCA, the high dimensional face data is projected to a low dimensional feature space and then LDA is performed in the low dimensional PCA subspace. Usually, the eigenfaces with small eigenvalues are removed in the PCA subspace. Since they may also encode some information helpful for recognition, their removal may introduce a loss of discriminant information. To construct a stable LDA classifier, the PCA subspace dimension has to be much smaller than the training set size. When the PCA subspace dimension is relatively high, the constructed LDA classifier is often biased and unstable. The projection vectors may be greatly changed by the slight disturbance of noise on the training set. So when the training set is small, some discriminative information has to be discarded in order to construct a stable LDA classifier.

In this paper, we use the random subspace method to overcome the small sample size problem for LDA. PCA is first applied to face data. A small number of eigenfaces are randomly selected to form a random subspace, and a LDA classifier is constructed from each subspace. In recognition, the input face image is fed to multiple LDA classifiers constructed from K random subspaces. The outputs are combined using a fusion methodology to make the final decision.

We apply this random subspace method to the integration of multiple features. In [4], Zhao et. al. pointed out that both holistic feature and local features are crucial for face recognition, and have different contributions. Three typical kinds of features, shape, texture and local Gabor wavelet responses are selected. They undergo a scale normalization and decorrelation by PCA to form a combined long feature vector. Then multiple LDA classifiers are generated by randomly sampling from this combined vector for recognition. Our approach is tested on the XM2VTS database [5]. It achieves 99.83% recognition accuracy, significantly outperforming conventional face recognition methods.

2. LDA in Reduced PCA Space

In this section, we briefly review the conventional LDA face recognition approach using two-stage PCA+LDA. For appearance-based face recognition, a 2D face image is viewed as a vector with length N in the high dimensional image space. The training set contains

Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR'04)
0-7695-2122-3/04 $ 20.00 © 2004 IEEE

M samples $\{x_i\}_{i=1}^{M}$ belonging to L individual classes $\{X_j\}_{j=1}^{L}$.

2.1. PCA

PCA uses the Karhunen-Loeve Transform for face representation and recognition. A set of eigenfaces are computed from the eigenvectors of the ensemble covariance matrix C of the training set,

$$C = \sum_{i=1}^{M} (x_i - m)(x_i - m)^T , \quad (1)$$

where m is the mean of all samples. Eigenfaces are sorted by eigenvalues, which represent the variance of face distribution on eigenfaces. There are at most *M-1* eigenfaces with non-zero eigenvalues. Normally the K largest eigenfaces, $U = [u_1, \quad u_K]$, are selected to span the eigenspace, since they can optimally reconstruct the face image with the minimum reconstruction error. Low dimensional face features are extracted by projecting the face data x to the eigenspace,

$$w = U^T (x - m). \quad (2)$$

The features on different eigenfaces are uncorrelated, and they are independent if the face data can be modeled as a Gaussian distribution.

2.2. LDA

LDA tries to find a set of projecting vectors W best discriminating different classes. According to the Fisher criteria, it can be achieved by maximizing the ratio of determinant of the between-class scatter matrix S_b to the determinant of the within-class scatter matrix S_w,

$$W = \arg\max \left| \frac{W^T S_b W}{W^T S_w W} \right|. \quad (3)$$

S_b and S_w are defined as,

$$S_w = \sum_{i=1}^{L} \sum_{x_k \in X_i} (x_k - m_i)(x_k - m_i)^T , \quad (4)$$

$$S_b = \sum_{i=1}^{L} n_i (m_i - m)(m_i - m)^T , \quad (5)$$

where m_i is the mean face for class X_i with n_i samples. W can be computed from the eigenvectors of $S_w^{-1} S_b$ [6]. The rank of S_w is at most *M-L*. But in face recognition, usually there are only a few samples for each class, and *M-L* is far smaller than the face vector length *N*. So S_w may become singular and it is difficult to compute S_w^{-1}.

In the Fisherface method [2], the face data is first projected to a PCA subspace spanned by *M-L* largest eigenfaces. LDA is then performed in the *M-L* dimensional subspace, such that S_w is nonsingular. But in many cases, *M-L* dimensionality is still too high for the training set. When the training set is small, S_w is not well estimated. A slight disturbance of noise on the training set will greatly change the inverse of S_w. So the LDA classifier is often biased and unstable.

Different subspace dimensions are selected in other studies. In [3], the dimension of PCA subspace was chosen as 40% of the total number of eigenfaces. In [7], the selected eigenfaces contains 95% of the total energy. They all remove eigenfaces with small eigenvalues. However, eigenvalue is not an indicator of the feature discriminability. Since the PCA subspace dimension depends on the training set, when the training set is small, some discriminative information has to be discarded in order to construct a stable LDA classifier.

3. Multi-Features Fusion Using RSM

3.1. RSM Based LDA

In Fisherface, overfitting happens when the training set is relatively small compared to the high dimensionality of the feature vector. In order to construct a stable LDA classifier, we sample a small subset of features to reduce discrepancy between the training set size and the feature vector length. Using such a random sampling method, we construct a multiple number of stable LDA classifiers. We then combine these classifiers to construct a more powerful classifier that covers the entire feature space without losing discriminant information.

Unlike the traditional random subspace method that samples the original feature vector directly, our algorithm samples in the PCA subspace. We first apply PCA to the face training set. All the eigenfaces with zero eigenvalues are removed, and *M-1* eigenfaces $U_0 = \{u_1, \quad , u_{M-1}\}$ are retained as candidates to construct random subspaces. The dimension of feature space is first greatly reduced without any loss on discriminative information, since all the training samples have zero projections on the eigenfaces with zero eigenvalues.

Then, K random subspaces $\{S_i\}_{i=1}^{K}$ are generated. The dimension of random subspace is determined by the training set to make the LDA classifier stable. The

0-7695-2122-3/04 $ 20.00 © 2004 IEEE
IEEE COMPUTER SOCIETY

random subspace is composed of two parts. The first N_0 dimensions are fixed as the N_0 largest eigenfaces, and the remaining N_1 dimensions are randomly selected from $\{u_{M-N_0-1}, \ldots, u_{M-1}\}$. The N_0 largest eigenfaces encode much face structural information. If they are not included in the random subspace, the accuracy of LDA classifiers may be too low. Our approach guarantees that the LDA classifier in each random subspace has satisfactory accuracy. The N_1 random dimensions cover most of the remaining small eigenfaces. So the ensemble classifiers also have a certain degree of error diversity.

K LDA classifiers $\{C_k(x)\}$ are constructed from the K random subspaces. In recognition, the raw face data is projected to the K random subspace and fed to K LDA classifier. The output are combined using a fusion methodology to make the final decision.

3.2. Combining Multiple LDA Classifiers

Many methods on combining multiple classifiers have been proposed [9][10]. In this paper, we use two simple fusion rules to combine LDA classifiers: majority voting and sum rule.

3.2.1 Majority voting

Each LDA classifier $C_k(x)$ assigns a class label to the input face data, $C_k(x) = i$. We represent this event as a binary function,

$$T_k(x \in X_i) = \begin{cases} 1, & C_k(x) = i \\ 0, & otherwise \end{cases}. \tag{6}$$

By a majority voting, the final class is chosen as,

$$\beta(x) = \arg\max_{X_i} \sum_{k=1}^{K} T_k(x \in X_i). \tag{7}$$

3.2.2 Sum rule

We assume that $P(X_i \mid C_k(x))$ is the probability that x belongs to X_i under the measure of LDA classifier $C_k(x)$. According to the sum rule, the class for the final decision is chosen as,

$$\beta(x) = \arg\max_{X_i} \sum_{k=1}^{K} P(X_i \mid C_k(x)) \tag{8}$$

$P(X_i \mid C_k(x))$ can be estimated from the output of the LDA classifier. For LDA classifier $C_k(x)$, the center m_i of class X_i, and input face data x are projected to LDA vectors W_k,

$$w_k^i = W_k^T m_i \tag{9}$$

$$w_k^x = W_k^T x \tag{10}$$

$P(X_i \mid C_k(x))$ is estimated as

$$\hat{P}(X_i \mid C_k(x)) = \left(1 + \frac{(w_k^x)^T (w_k^i)}{\|w_k^x\| \cdot \|w_k^i\|}\right) / 2, \tag{11}$$

which has been mapped to [0,1].

3.3. Multi-Features Fusion

Random subspace based LDA essentially integrates different parts of the original high dimensional feature vector. It also can be extended to the integration of different kinds of features such as shape, texture, and Gabor responses. A face graph containing 35 fiducial points is designed as shown in Fig. 1. Using the method in Active Shape Model [11], we separate the face image into shape and texture. The shape vector V_s is formed by concatenating the coordinates of the 35 fiducial points after alignment. Warping the face image onto a mean face shape, the texture vector V_t is obtained by sampling intensity on the shape-normalized image. As described in Elastic Bunch Graph Matching [12], a set of Gabor kernels in five scales and eight orientations are convolved with the local patch around each fiducial point. The Gabor feature vector V_g combines 35×40 magnitudes of Gabor responses to represent the face local texture.

The multi-feature random subspace LDA face recognition algorithm is then designed as the following.

1. Apply PCA to the three features vectors respectively to compute the eigenvectors U_s, U_t, U_g and eigenvalues λ_i^s, λ_i^t, λ_i^g. All the eigenvectors with zero eigenvalues are removed.
2. For each face image, project each kind of feature to the eigenvectors and normalize them by the sum of eigenvalues, such that they are in the same scale.
$$w_j = U_j^T V_j / \sqrt{\sum \lambda_i^j}, \quad (j=s, t, g). \tag{12}$$
3. Combine w_t, w_s, w_g into a large feature vector, and apply PCA on this combined feature vector.
4. Apply the random subspace algorithm to the combined feature vectors to generate multiple LDA classifiers.
5. Combine these multiple classifiers.

While most other multi-feature integration systems are based on match score level or decision level by designing one classifier for each kind of feature, our

IEEE COMPUTER SOCIETY

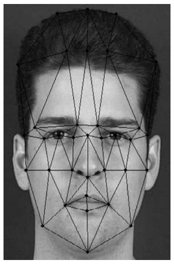

Figure 1. Face graph model.

integration approach starts from the feature level. Integration at feature level conveys the richest information, but it is more difficult, since different kinds of features are incompatible in scale and the new combined feature vector has a higher dimensionality. Our approach overcomes both problems.

4. Experiments

We conduct experiments on the XM2VTS face database [5]. There are 295 people, and each person has four frontal face images taken in four different sessions. Some examples are shown in Figure 2. In our experiments, two face images of each face class are used for training and reference, and the remaining two for testing. We adopt the recognition test protocol used in FERET [13]. All the face classes in the reference set are ranked. We measure the percentage of the "correct answer in top 1 match".

4.1. Random subspace based LDA

We first compare random subspace based LDA with conventional LDA approach using the holistic feature. In preprocessing, the face image is normalized by translation, rotation, and scaling, such that the centers of two eyes are in fixed positions. A 46 by 81 mask removes most of the background. Histogram equalization is applied as photometric normalization.

Figure 3 reports the accuracy of a single LDA classifier constructed from PCA space with different number of eigenfaces. Since there are 590 face images of 295 classes in the training set, there are 589 eigenfaces with non-zero eigenvalues. According to the Fisherface [2], the PCA space dimension should be $M\text{-}C=295$. However, the result shows that the accuracy is only 79% using a single LDA classifier constructed from 295 eigenfaces, because this dimension is too high for the training set. We observe that LDA classifier has the best accuracy 92.88%, when the PCA subspace dimension is set at 100. So for this data set, 100 seems to be a suitable dimension to construct a stable LDA

Figure 2. Examples of face images in XM2VTS database taken in four different sessions

classifier. In the following experiments, we choose 100 as the dimension of random subspaces to construct the multiple LDA classifiers.

First, we randomly select 100 eigenfaces from the 589 eigenfaces with nonzero eigenvalues. The result of combining 20 LDA classifiers using majority voting is shown in Figure 4. With random sampling, the accuracy of each individual LDA classifier is low, between 50% and 70%. Using majority voting, the weak classifiers are greatly enforced, and 87% accuracy is achieved. This shows that LDA classifiers constructed from different random subspaces are complementary of each other. In Figure 5, as we increase the classifier number K, the accuracy of the combined classifier improves, and even becomes better than the highest accuracy in Figure 3. Although increasing classifier number and using more complex combining rules may further improve the performance, it will increase the system burden.

A better approach to improve the accuracy of the combined classifier is to increase the performance of each individual weak classifier. To improve the accuracy of each individual LDA classifier, as proposed in Section 3.1, in each random subspace, we fix the first 50 basis as the 50 largest eigenfaces, and randomly select another 50 basis from the remaining 539 eigenfaces. As shown in Figure 6, individual LDA classifiers are improved significantly. They are similar to the LDA classifier based on the first 100 eigenfaces. This shows that $\{u_{51}, \ldots, u_{100}\}$ are not necessarily more discriminative than those smaller eigenfaces. These classifiers are also complementary of each other, so much better accuracy is achieved when they are combined.

4.2. Multiple Features Fusion

In Table 1, we report the recognition accuracy of integrating shape, texture, and Gabor features using random subspace LDA. Combining 20 classifiers using the sum rule, it achieves 99.83% recognition accuracy.

Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR'04)
0-7695-2122-3/04 $ 20.00 © 2004 IEEE
IEEE COMPUTER SOCIETY

For 590 testing samples, it misclassifies only one! For comparison, we also report the accuracies of some conventional face recognition approaches. Eigenface [14], Fisherface [2], and Bayes[15] are three subspace face recognition approaches based on holistic feature. Elastic Bunch Graph Matching described in [12] uses the correlation of Gabor features as similarity measure. Experiments clearly demonstrate the superiority of the new method.

5. Conclusion

In this paper, we propose to use random subspace based LDA to integrate multiple features. It effectively stabilizes the LDA classifier based on a small training set, and makes use of most discriminative features in the high dimensional space.

In this study, we use the simplest fusion rules to combine multiple LDA classifiers and also achieve notable improvement. Many more complex combination algorithms [9][10] have been proposed. They may further improve the performance. Using random subspace, a large set of LDA classifiers can be generated. Instead of combining them directly, it is helpful to select a small set of complementary LDA classifiers with high accuracy from them for combination. This is a direction for our further study.

Acknowledgement

The work described in this paper was fully supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region (Project no. CUHK 4190/01E and CUHK 4224/03E).

References

[1] T. Kam Ho, "The Random Subspace Method for Constructing Decision Forests," *IEEE Trans. on PAMI*, Vol. 20, No. 8, pp. 832-844, August 1998.

[2] P.N. Belhumeur, J. Hespanda, and D. Kiregeman, “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection”, *IEEE Trans. on PAMI,* Vol. 19, No. 7, pp. 711-720, July 1997.

[3] D. Swets, J. Weng, “Using Discriminant Eigenfeatures for Image Retrieval”, *IEEE Trans. on PAMI*, Vol. 16, No. 8, pp. 831-836, August, 1996.

[4] W. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips,(2000) "Face Recognition: A Literature Survey", *UMD CAR Technical Report CAR-TR948.*

[5] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, “XM2VTSDB: The Extended M2VTS Database,” *Proceedings of International Conference on Audio- and Video-Based Person Authentication*, pp. 72-77, 1999.

[6] K. Fukunnaga, “Introduction to Statistical Pattern Recognition”, Academic Press, second edition, 1991.

[7] H. Monn, and P.J. Phillips, “Analysis of PCA-Based Face Recognition Algorithms,” Empirical Evaluation Techniques in Computer Vision, K. W. Bowyer and P.J. Phillips, eds, IEEE Computer Society Press, Los Alamitos, CA, 1998.

[8] R. Huang, Q. Liu, H. Lu, and S. Ma, "Solving Small Sample Size Problem of LDA", *Proceedings of ICPR*, pp. 29-32, 2002.

[9] L. Xu, A. Krzyzak, C. Y. Suen, "Method of Combining Multiple Classifiers and Their Applications to Handwriting Recognition," IEEE Trans. on System, Man, and Cybernetics, Vol. 22, No. 3, 418-435, 1992.

[10] T. Kam Ho, J. Hull, S. Srihari, "Decision Combination in Multiple Classifier Systems," IEEE Trans. on PAMI, Vol. 16, No.1, pp. 66-75, Jan. 1994.

[11] A. Lanitis, C. J. Taylor, and T. F. Cootes, “Automatic Interpretation and Coding of Face Images Using Flexible Models,” *IEEE Trans. on PAMI*, Vol. 19, No. 7, pp. 743-756, July 1997.

[12] L. Wiskott, J.M. Fellous, N. Kruger and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No.7, pp. 775-779, July, 1997.

[13] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, “The FERET Evaluation,” in Face Recognition: From Theory to Applications, H. Wechsler, P. J. Phillips, V. Bruce, F.F. Soulie, and T. S. Huang, Eds., Berlin: Springer-Verlag, 1998.

[14] M. Turk and A. Pentland, “Face recognition using eigenfaces”, *Proceedings of IEEE, CVPR*, pp. 586-591, Hawaii, June, 1991.

[15] B. Moghaddam, T. Jebara, and A. Pentland, “Bayesian Face Recognition”, *Pattern Recognition*, Vol. 33, pp. 1771-1782, 2000.

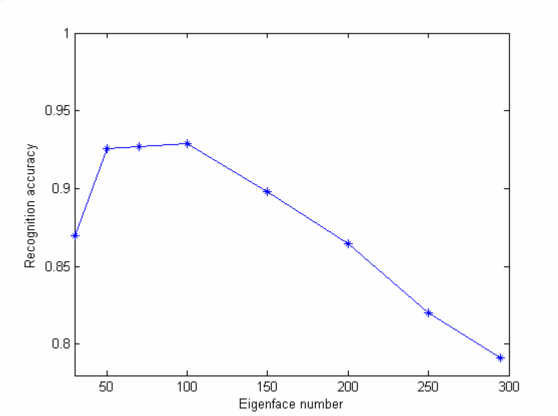

Figure 3. Recognition accuracy of LDA classifier using different number of eigenfaces in the reduced PCA subspace.

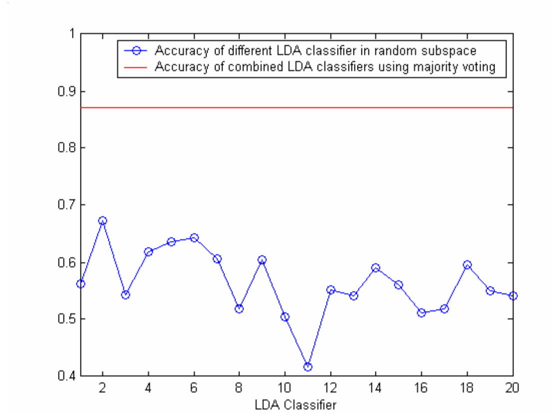

Figure 4. Recognition accuracy of combing 20 LDA classifiers constructed from random subspaces using majority voting. Each random subspace randomly selects 100 eigenfaces from 589 eigenfaces with non-zero eigenvalues.

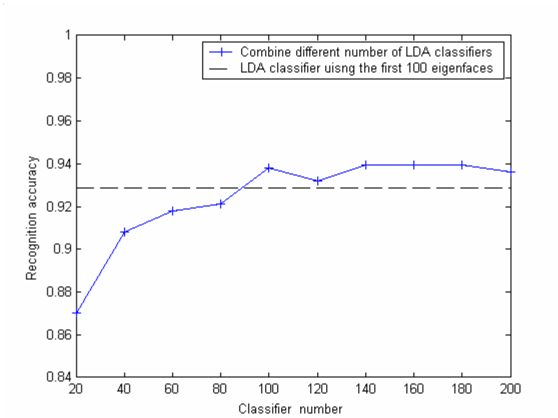

Figure 5. Accuracy of combining different number of LDA classifiers constructed from random subspaces using majority voting. Each random subspace randomly selects 100 eigenfaces from 589 eigenfaces with non-zero eigenvalues.

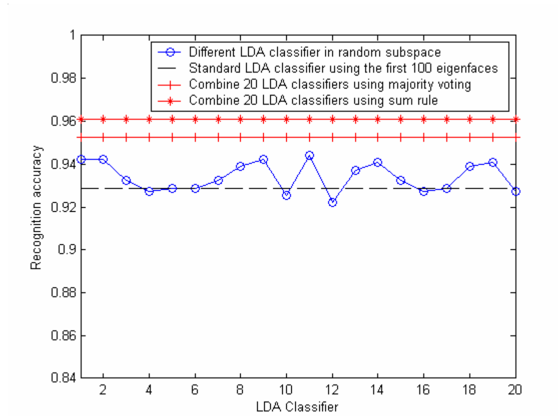

Figure 6. Recognition accuracy of combing 20 LDA classifiers constructed from random subspaces using majority voting and the sum rule. For each 100 dimension random subspace, the first 50 dimensions are fixed as the 50 largest eigenfaces, and another 50 dimensions are randomly selected from the remaining 539 eigenfaces with non-zero eigenvalues.

Table 1. Compare recognition accuracy of random subspace based LDA (R-LDA) with conventional methods

Feature	Holistic feature				Shape	Texture	Gabor	Integration of Multi-feature
Method	Eigenface	Fisherface	Bayes	R-LDA	Euclid	Euclid	EBGM	R-LDA
Accuracy	85.59%	92.88%	92.71%	96.10%	49.5%	85.76%	95.76%	99.83%

Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR'04)
0-7695-2122-3/04 $ 20.00 © 2004 IEEE

IEEE COMPUTER SOCIETY