Random Sampling LDA for Face Recognition

Xiaogang Wang and Xiaoou Tang

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

Abstract

Linear Discriminant Analysis (LDA) is a popular feature extraction technique for face recognition. However, It often suffers from the small sample size problem when dealing with the high dimensional face data. Fisherface and Null Space LDA (N-LDA) are two conventional approaches to address this problem. But in many cases, these LDA classifiers are overfitted to the training set and discard some useful discriminative information. In this paper, by analyzing different overfitting problems for the two kinds of LDA classifiers, we propose an approach using random subspace and bagging to improve them respectively. By random sampling on feature vector and training samples, multiple stabilized Fisherface and N-LDA classifiers are constructed. The two kinds of complementary classifiers are integrated using a fusion rule, so nearly all the discriminative information are preserved. We also apply this approach to the integration of multiple features. A robust face recognition system integrating shape, texture and Gabor responses is finally developed.

1. Introduction

LDA is a popular feature extraction technique for face recognition. It determines a set of projection vectors maximizing the between-class scatter matrix (S_b) and minimizing the within-class scatter matrix (S_w) in the projective feature space. But 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, S_w is not well estimated and may become singular.

To address this problem, a two-stage PCA+LDA approach, i.e. Fisherface [1] is proposed. Using PCA, the high dimensional face data is projected to a low dimensional feature space and then LDA is performed in this 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 discriminative information. To construct a stable LDA classifier, the PCA subspace dimension is dependent on 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.

Chen et al. [2] suggested that the null space spanned by the eigenvectors of S_w with zero eigenvalues contains the most discriminative information. A LDA method in the null space of S_w was proposed. It chose the projection vectors maximizing S_b with the constraint that S_w is zero. However, as explained in [2], with the existence of noise, when the training sample number is large, the null space of S_w becomes small, so much discriminative information outside this null space will be lost. The constructed classifier may also be over tuned to the training set.

In this paper, we propose an approach using random sampling to improve LDA based face recognition. Random subspace [3] and bagging [4] are two popular random sampling techniques to enforce weak classifiers. In the random subspace method, a set of low dimensional subspaces are generated by randomly sampling from the original high dimensional feature vector and multiple classifiers constructed in the random subspaces are combined in the final decision. In bagging, random independent bootstrap replicates are generated by sampling the training set. A classifier is constructed from each replicate, and the results of all the classifiers are finally integrated.

Both Fisherface and Null Space LDA (N-LDA) encounter the overfitting problem, but for different reasons. So we will improve them in different ways accordingly. In Fisherface, overfitting happens when the training set is small compared to the high dimensionality of the feature vector. We apply random subspace to reduce the feature vector dimension to reduce the discrepancy. In N-LDA, the null space is small when the training sample number is large. This problem can be alleviated by bagging, since each replicate has a smaller number of training samples. Both Fisherface and N-LDA discard some discriminative information. However, the

Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)
1063-6919/04 $20.00 © 2004 **IEEE**

two kinds of classifiers are also complementary, since they are computed in two orthogonal subspaces. We combine them using a fusion rule. Using random sampling, the constructed LDA classifiers are stable and multiple classifiers cover most of the face feature space, so less discriminative information is lost. We also apply this random sampling approach to the integration of multiple features. A robust face recognition system integrating shape, texture, and Gabor responses is developed.

2. LDA Based Face Recognition

In this section, we briefly review Fisherface and N-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 *M* samples $\{\vec{x}_i\}_{i=1}^{M}$ belonging to *L* individual classes $\{X_j\}_{j=1}^{L}$.

2.1. PCA

In PCA, a set of eigenfaces are computed from the eigenvectors of the ensemble covariance matrix *C* of the training set,

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

where $\vec{m}$ 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, $U = [\vec{u}_1, \ldots \vec{u}_K]$, are selected to span the PCA subspace, since they can optimally reconstruct the face image with the minimum reconstruction error. Low dimensional face features are extracted by projecting the face data $\vec{x}$ to the PCA subspace,

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

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

2.2. Fisherface

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 and 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,

$$S_w = \sum_{i=1}^{L} \sum_{\vec{x}_k \in X_i} (\vec{x}_k - \vec{m}_i)(\vec{x}_k - \vec{m}_i)^T , \quad (4)$$

$$S_b = \sum_{i=1}^{L} n_i (\vec{m}_i - \vec{m})(\vec{m}_i - \vec{m})^T . \quad (5)$$

where $\vec{m}_i$ is the mean face for class X_i with n_i samples. *W* can be computed from the eigenvectors of $S_w^{-1} S_b$ [5]. 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 [1], the face data is first projected to a PCA subspace spanned by the *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. In fact, the proper PCA subspace dimension depends on the training set. Usually, eigenfaces with small eigenvalues are removed. However, eigenvalue is not an indicator of the feature discriminability. When the training set is small, some discriminative information has to be discarded in order to construct a stable LDA classifier.

2.2. Null Space LDA

Chen et. al. [2] suggested that the null space of S_w, in which $W^T S_w W = 0$, also contains much discriminative information. It is possible to find some projection vectors *W* satisfying $W^T S_w W = 0$ and $W^T S_b W \neq 0$, thus the Fisher criteria in Eq. (3) definitely reaches its maximum value. A LDA approach in the null space of S_w was proposed. First, the null space of S_w is computed as,

$$V^T S_w V = 0 . \quad (6)$$

The between-class scatter matrix is projected to the null space of S_w,

$$\widetilde{S}_b = V^T S_b V . \quad (7)$$

The LDA projection vectors are defined as $W = V\Phi$, where Φ contains the eigenvectors of $\widetilde{S}_b$ with the largest eigenvalues.

N-LDA may also overfit the training set. The rank of S_w, $r(S_w)$ is bounded by $\min(M-L, N)$. Because of the existence of noise, $r(S_w)$ is almost equal to this bound. The dimension of the null space is $\max(0, N-M+L)$. As shown by experiments in [2], when the training sample number is large, the null space of S_w becomes small, thus much discriminative information outside this null space will be lost. An extreme case is that the training set is so large that we have *M-L=N*. Then no information can be obtained in this space, since the dimension of the null space is zero.

3. Random Sampling Based LDA for Face Recognition

The above LDA approaches have two common problems: the constructed classifier is unstable and much discriminative information is discarded. In this section, we use random sampling to improve LDA based face recognition. We construct many weak classifiers and combine them into a powerful decision rule. Although Fisherface and N-LDA share the same kind of problems, they are due to different reasons. So we design different random sampling algorithms to improve the two LDA methods. We then combine the two improved methods in a multi-classifier structure.

Although the dimension of image space is very high, only part of the full space contains the discriminative information. This subspace is spanned by all the eigenvectors of the ensemble covariance matrix *C* with nonzero eigenvalues. For the covariance matrix computed from *M* training samples, there are at most *M-1* eigenvectors with nonzero eigenvalues. On the remaining eigenvectors with zero eigenvalues, all the training samples have zero projections and no discriminative information can be obtained. Therefore for both random sampling algorithms, we first project the high dimension image data to the *M-1* dimension PCA subspace before random sampling.

3.1. Using Random Subspace to Improve Fisherface

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. The random subspace LDA algorithm contains the following steps:

At the training stage,

(1) Apply PCA to the face training set. All the eigenfaces with zero eigenvalues are removed, and *M-1* eigenfaces $U_t = \{u_1, \ldots, u_{M-1}\}$ are retained as candidates to construct the random subspaces.

(2) Generate *K* random subspaces $\{S_i\}_{i=1}^{K}$. Each random subspace S_i is spanned by $N_0 + N_1$ dimensions. The first N_0 dimensions are fixed as the N_0 largest eigenfaces in U_t. The remaining N_1 dimensions are randomly selected from the other $M-1-N_0$ eigenfaces in U_t.

(3) *K* LDA classifiers $\{C_i(x)\}$ are constructed from the *K* random subspaces.

At the recognition stage,

(1) The input face data is projected to the *K* random subspaces and fed to the *K* LDA classifiers in parallel.

(2) The outputs of the *K* LDA classifiers are combined using a fusion scheme to make the final decision.

This algorithm has several novel features. First, this is the first time that the random subspace method is applied to face recognition. Second, unlike the traditional random subspace method that samples the original feature vector directly, our algorithm samples in the PCA subspace. The dimension of feature space is first greatly reduced without loss on discriminative information. After PCA, the features on different eigenfaces are uncorrelated, thus are more independent. Better accuracy can be achieved if different random subspaces are more independent from each other.

Third, our random subspace is not completely random. The dimension of the random subspace is fixed. It is determined by the training set to make the individual LDA classifier stable. Then, the dimensions of the random subspace are divided into 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. Although many multiple classifier

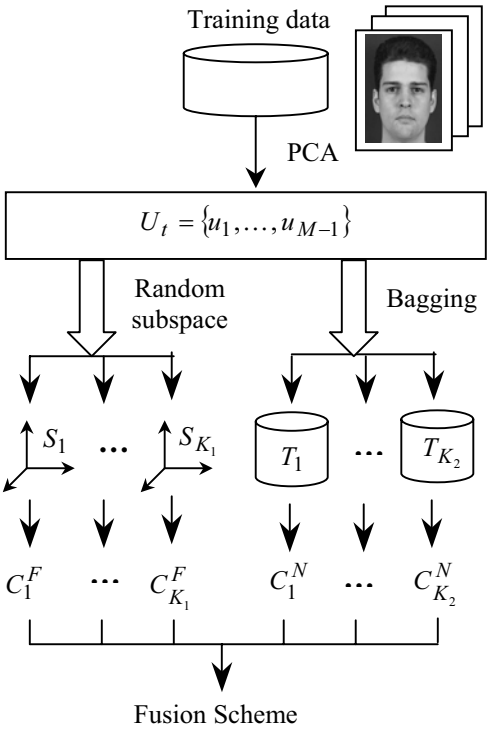

Figure 1. Integrate multiple Fisherface and N-LDA classifiers generated by random sampling. C_i^F is the LDA classifier constructed from the random subspace S_i, and C_i^N is the N-LDA classifier constructed from the bagging replicate T_i .

systems have been proposed to enforce weak classifiers, the fusion method will be more complicated if each individual LDA classifier is poor. In our approach, 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. Good performance can be achieved using very simple fusion rules such as majority voting.

3.2. Using Bagging to Improve Null Space LDA

Contrary to Fisherface, for N-LDA, the overfitting problem happens when the training sample number is large, since the null space is too small to contain enough discriminative information. This problem can be alleviated by bagging. In bagging, random independent bootstrap replicates are generated by sampling the training set, so each replicate has a smaller number of training samples. Based on this strategy, we propose the following algorithm:

(1) Apply PCA to the face training set with M samples for L classes. Project all the face data to the *M-1* eigenfaces $U_t = \{u_1, \ldots, u_{M-1}\}$ with positive eigenvalues.

(2) Generate K bootstrap replicates $\{T_i\}_{i=1}^K$. Each replicate contains the training samples of L_1 individuals randomly selected from the L classes.

(3) Construct a N-LDA classifier from each replicate and combine the multiple classifiers using a fusion rule.

Our algorithm randomly selects the individual classes, but does not randomly sample data within each class. This is because in face recognition usually there are a large number of people to be classified but there are very few samples in each class. For example, in our experiment, there are 295 people in the gallery and each person has only two samples for training. The N-LDA constructed from the replicate T_i not only can classify the L_1 individuals in this replicate effectively, but also can distinguish persons outside T_i , because human faces share similar intrapersonal variations [6]. The K classifiers can cover all the L classes in the training set.

3.3. Integrating Random Subspace and Bagging for LDA Based Face Recognition

While Fisherface is computed from the principal subspace of S_w , in which $W^T S_w W \neq 0$, N-LDA is computed from its orthogonal subspace in which $W^T S_w W = 0$. Both of them discard some discriminative information. Fortunately, the information retained by the two classifiers complements each other. So we combine the two sets of complementary multiple LDA classifiers generated by random sampling to construct the final classifier as illustrated in Figure 1.

Many methods on combining multiple classifiers have been proposed [7]. In this paper, we use two simple fusion rules to combine LDA classifiers: majority voting and sum rule. More complex combination algorithms may further improve the system performance.

In [8], Zhao et al. pointed out that both face holistic features and local features are critical for recognition and have different contributions. We apply this random sampling LDA approach to the integration of multiple features including shape, texture, and Gabor responses. Using the method in Active Shape Model [9], we separate the face image into shape $\vec{V}_s$ and texture $\vec{V}_t$. A set of Gabor features $\vec{V}_g$ are extracted as described in Elastic Bunch Graph Matching [10]. The multi-feature multi-

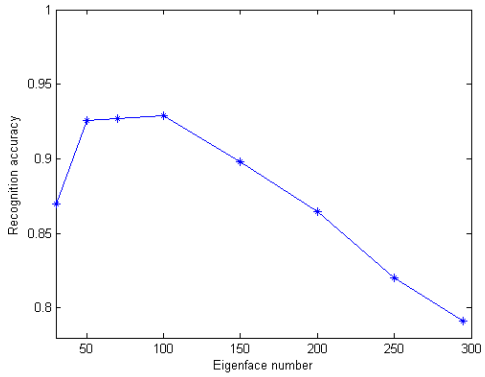

Figure 2. Recognition accuracy of Fisherface classifier using different number of eigenfaces in the reduced PCA subspace.

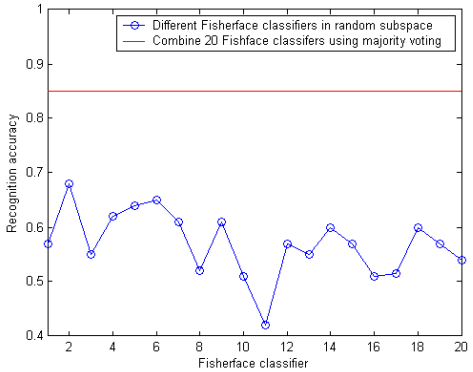

Figure 3. Recognition accuracy of combing 20 Fisherface classifiers constructed from random subspaces using majority voting. Each random subspace randomly selects 100 eigenfaces from 589 eigenfaces with non-zero eigenvalues.

calssifier face recognition algorithm is then designed as following:

(1) Apply PCA to the three feature 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.

$$\vec{w}_j = U_j^T \vec{V}_j / \sqrt{\sum \lambda_i^j}, \quad (j=s, t, g). \tag{8}$$

(3) Combine $\vec{w}_t$, $\vec{w}_s$, $\vec{w}_g$ into a large feature vector.

(4) Apply the random sampling algorithm as illustrated in Figure 1 to the combined feature vector to generate multiple LDA classifiers.

(5) Combine these multiple classifiers.

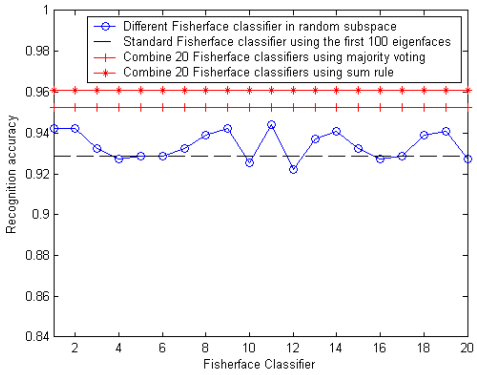

Figure 4. Recognition accuracy of combing 20 Fisherface classifiers constructed from random subspaces using majority voting and the sum rule. For each 100 dimensional random subspace, the first 50 dimensions are fixed as the 50 largest eigenfaces, and another 50 dimensions are randomly selected from the remaining 539 eigenfaces with non-zero eigenvalues.

4. Experiments

We conduct experiments on the XM2VTS face database [11]. There are 295 people, and each person has four frontal face images taken in four different sessions. In our experiments, two face images of each face class are selected for training and reference, and the remaining two are for testing. We adopt the recognition test protocol used in FERET [12]. 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 LDA

We first compare random subspace LDA with the 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. So the image space dimensionality is $46 \times 81 = 3726$. Histogram equalization is applied as photometric normalization.

Figure 2 reports the accuracy of a single LDA classifier constructed from PCA subspace 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 [1], the PCA subspace dimension should be *M-L=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

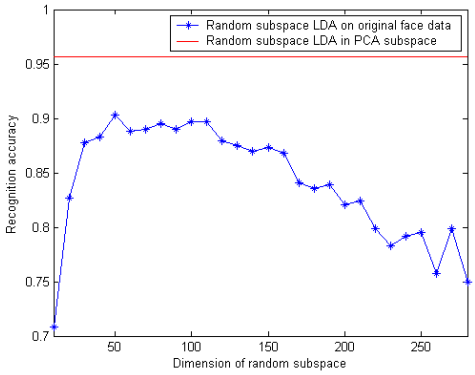

Figure 5. Recognition accuracy of random subspace LDA directly sampling on the original face data, using majority voting based on 20 random subspaces. The random subspace dimension for the new PCA based sampling is fixed at 100.

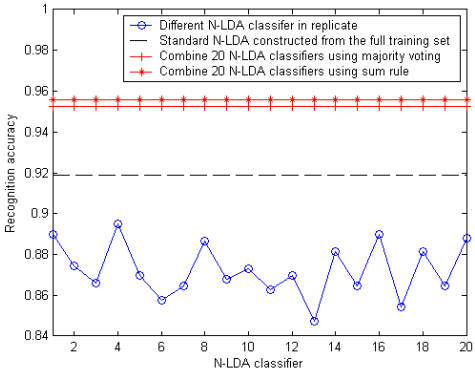

Figure 6. Recognition accuracy of combining 20 N-LDA classifiers constructed from bagging replicates using majority voting and sum rule. Each replicate contains 150 training people.

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 training set, 100 seems to be a suitable dimension to construct a stable LDA classifier. In the following experiments, we choose 100 as the dimension of the random subspaces to construct the multiple LDA classifiers.

First, we randomly select 100 eigenfaces from 589 eigenfaces with nonzero eigenvalues. The result of combining 20 LDA classifiers using majority voting is shown in Figure 3. 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.

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. Toward this, as illustrated in Section 3.1, in each random subspace, we fix the first 50 dimensions as the 50 largest eigenfaces, and randomly select another 50 dimensions from the remaining 539 eigenfaces. As shown in Figure 4, 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.

In Figure 5, we report the recognition accuracy of random subspace LDA directly sampling on the raw face data with different random subspace dimensions, just like the original random subspace method, instead of in the PCA subspace. It shows that our improved random subspace method in PCA subspace has a superior performance.

4.2. Bagging LDA

Figure 6 reports the performance of bagging based N-LDA. We generate 20 replicates and each replicate contains 150 people for training. As expected, the individual N-LDA classifier constructed from each replicate is less effective than the original classifier trained on the full training set. However, when the multiple classifiers are combined, the accuracy is significantly improved, and becomes much better than the original N-LDA.

4.3.Integration of Random Subspace and Bagging

Integrating the multiple Fisherface classifiers generated by random subspace and N-LDA classifiers generated by bagging the recognition accuracy can be further improved. We combine 10 Fisherface classifiers constructed from random subspaces and 10 N-LDA classifiers constructed from bagging replicates, and achieve an even better result as shown in Table 1.

In Table 1, we also report the recognition accuracy of integrating shape, texture, and Gabor features using random sampling LDA. Combining 20 classifiers using the sum rule, we achieve 99.83% recognition accuracy. For 590 testing samples, it misclassifies only one! For comparison, we also compute the accuracies of some conventional face recognition approaches in Table 1. Eigenface [13], Fisherface [1], and Bayesian analysis [14] are three subspace face recognition approaches based on

holistic feature. Elastic Bunch Graph Matching as described in [10] uses the correlation of Gabor features as similarity measure. The results clearly demonstrate the superiority of our new algorithm.

5. Conclusion

Random sampling is an effective technique to enforce weak classifiers. Both Fisherface and N-LDA encounter the overfitting problems in face recognition, however, for different reasons. So we improve them using different random sampling approaches, sampling on feature for Fisherface and sampling on training samples for N-LDA. The two kinds of complementary classifiers are then integrated in our system. Our approach effectively stabilizes the LDA classifier and makes use of all the discriminative information in the high dimensional space. In future study, we will investigate application of this random sampling approach to the unified subspace analysis [6] and face sketch recognition [15], and further compare with the Dual-Space LDA, which combines the two complementary LDA subspace at feature level [16].

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).

Table 1. Compare random sampling based LDA with conventional methods. R-LDA (1): random subspace based Fisherface; R-LDA (2): bagging based N-LDA; R-LDA (3): integrating random subspace and bagging based LDA.

Feature	Method	Accuracy
Holistic feature	Eigenface	85.59%
	Fisherface	92.88%
	Bayes	92.71%
	R-LDA (1)	96.10%
	R-LDA (2)	95.59%
	R-LDA (3)	97.63%
Texture	Euclid distance	85.76%
Shape	Euclid distance	49.50%
Gabor	EBGM	95.76%
Integration of multi-feature	R-LDA (3)	99.83%

Reference

[1] 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.

[2] L. Chen, H. Liao, M. Ko, J. Liin, and G. Yu, "A New LDA-based Face Recognition System Which can Solve the Samll Sample Size Problem," *Pattern Recognition*, Vol. 33, No. 10, pp. 1713-1726, Oct. 2000.

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

[4] L. Breiman, "Bagging Predictors," *Machine Learning*, Vol. 24, No. 2, pp. 123-140, 1996.

[5] K. Fukunnaga, "Introduction to Statistical Pattern Recognition," Academic Press, second edition, 1991.

[6] X. Wang and X. Tang, "A Unified Framework for Subspace Face Recognition," *in Proceedings of ICCV*, pp. 679-686, Nice, France, Oct. 2003.

[7] J. Kittler and F. Roli, (Eds): Multiple Classifier Systems.

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

[9] 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.

[10] 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.

[11] 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.

[12] 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.

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

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

[15] X. Tang and X. Wang, "Face Sketch Synthesis and Recognition," in *Proceedings of ICCV*, pp. 687-694, Nice, France, Oct. 2003.

[16] X. Wang and X. Tang, "Dual-Space Linear Discriminant Analysis for Face Recognition," in *Proceedings of CVPR*, Washington D.C., USA, June, 2004.