Feedback-based Dynamic Generalized LDA for Face Recognition

Dahua Lin
Dept. of Information Engineering
The Chinese University of Hong Kong
Shatin, Hong Kong SAR
Email: dhlin4@ie.cuhk.edu.hk

Shuicheng Yan
Dept. of Information Engineering
The Chinese University of Hong Kong
Shatin, Hong Kong SAR
Email: scyan@ie.cuhk.edu.hk

Xiaoou Tang
Dept. of Information Engineering
The Chinese University of Hong Kong
Shatin, Hong Kong SAR
Email: xitang@microsoft.com

***Abstract*— Linear Discriminant Analysis(LDA) is widely-used in face recognition systems. However, with the traditional formulation, the available information in the training samples is not sufficiently utilized. In this paper, we present a new formulation, called Generalized LDA, where the scatter matrices are defined in a more flexible manner by identifying the fundamental principles of the scatter matrices construction. We further propose a novel framework called Feedback-based Dynamic Generalized LDA. It integrates the Generalized LDA and the dynamic feedback strategy for subspace analysis, in which the subspace is iteratively optimized by utilizing the feedback from the previous step. The comparative experiments demonstrate that the new framework achieves encouraging improvement on performances of both the face identification and the face verification.**

I. Introduction

Since Eigenface [1] is applied to face recognition, extensive researches have been conducted on linear subspace methods. Among them, Linear Discriminant Analysis (LDA), also known as Fisherface [2] [3] is the most popular one owing to its simplicity and effectiveness. Numerous improved implementations [4] [5] [6] [7] [8] of LDA have been proposed to enhance the generalization capability or the stability of the LDA algorithm. Though these algorithms have demonstrated their superiorities over the conventional implementation, their classification accuracies and adaptability are still insufficient when the variations of face images are complex. We argue that the limitation is due to the following two reasons:

First, the training process of LDA and most of its variations are formulated to optimize the trace-ratio of the two scatter matrices, which is a measure of separability. However, "best separability" is not always equivalent to "best classification". Actually, it can be proved that the projection matrix obtained under this formulation is optimal for classification only when the samples in all the classes satisfy Gaussian distribution with equal covariance. In reality, the probability distribution of the face sample space rarely meets such a strong condition, accordingly, the resultant basis is not always the optimal one.

Second, in the traditional LDA approaches, the construction of the scatter matrices is static. Though various approaches [6] [9] have been proposed to refine the scatter matrices; however, most of them only exploit a portion of important information, and thus cannot effectively capture the characteristics of the sample distribution. To fully make use of all the available information, an adaptive feedback mechanism is desirable.

In this paper, firstly, we reformulate the scatter matrices to provide a new perspective to interpret the scatter matrices and a more flexible mathematical form for scatter matrix construction. Secondly, based on this new formulation, a novel framework is proposed for discriminant subspace analysis. Different from the traditional LDA-based approaches, our framework seamlessly incorporates a powerful feedback mechanism. Briefly speaking, the initially constructed model is updated as follows: in each step, the learned model is applied to the training set, and the samples prone to be misclassified are selected to update the scatter matrices. Intuitively, the procedure adaptively converts the model from a "best-separability-model" to a "best-classification-model" by gradually increasing the emphasis on the data pairs with most important complementary information. A more effective utilization of discriminative information is achieved in this adaptive process.

The paper is organized as follows: In section 2, we review the LDA method and its variations, and then introduce our new formulation for discriminant subspace analysis. Section 3 elaborates on our feedback-based framework. Experiments and final conclusion are given in section 4 and section 5, respectively.

II. Reformulation of LDA

In this paper, each sample is represented as a d-dimensional column vector $\mathbf{x}$; $l(\mathbf{x}_k) \in \{1, 2, \ldots, C\}$ is a function indicating which class a sample belongs to, where C is the number of classes.

A. Review LDA

The traditional LDA approaches aim at finding a projection matrix $\mathbf{W}$ to maximize the separability, which is defined as the ratio between the trace of between-class scatter matrix $\mathbf{S}_b$

and that of the within-class scatter matrix $\mathbf{S}_w$.

$$\mathbf{W} = \underset{\mathbf{W}}{\text{argmax}} \frac{\text{tr}(\mathbf{W}^T\mathbf{S}_b\mathbf{W})}{\text{tr}(\mathbf{W}^T\mathbf{S}_w\mathbf{W})}, \quad (1)$$

$$\mathbf{S}_b = \sum_{i=1}^{C} n_i(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T, \quad (2)$$

$$\mathbf{S}_w = \sum_{i=1}^{C}\sum_{p=1}^{n_i}(\mathbf{x}_{ip} - \mathbf{m}_i)(\mathbf{x}_{ip} - \mathbf{m}_i)^T. \quad (3)$$

Here, N is the total number of training samples, n_i is the number of training samples in the i-th class, $\mathbf{x}_{ip}$ denotes the p-th sample in the i-th class, $\mathbf{m}_i$ is the mean vector of the samples belonging to the i-class, $\mathbf{m}$ is the mean vector of all training samples.

A variety of implementations of LDA [4] [5] [7] [8] to obtain transform matrix have been proposed. In this paper, we employ the Unified Subspace Analysis [7] [8], which first performs PCA for dimension reduction, and then a procedure composed of two diagonalization stages is applied to obtain the final projection matrix.

B. New Formulation

To gain a deeper insight into LDA, we first investigate the definition of the scatter matrices. From Eq.(2) and (3), we have

$$\mathbf{S}_b = \frac{1}{N}\sum_{i=1}^{C-1}\sum_{j=i+1}^{C} n_i n_j(\mathbf{m}_i - \mathbf{m}_j)(\mathbf{m}_i - \mathbf{m}_j)^T, \quad (4)$$

$$\mathbf{S}_w = \sum_{i=1}^{C}\frac{1}{n_i}\sum_{p=1}^{n_i-1}\sum_{q=p+1}^{n_i}(\mathbf{x}_{ip} - \mathbf{x}_{iq})(\mathbf{x}_{ip} - \mathbf{x}_{iq})^T. \quad (5)$$

Eq.(4) and Eq.(5) reveals the fundamental principles for scatter matrices construction:
1) A scatter matrix is the combination of the outer products of difference vectors. Increasing the trace of the scatter matrix involves increasing of distances of constituent sample pairs;
2) $\mathbf{S}_b$ is constructed with the pairs of samples from different classes (*Negative Sample Pairs*) to measure inter-class scattering; while $\mathbf{S}_w$ is constructed with the pairs of samples from the same classes (*Positive Sample Pairs*) to measure intra-class scattering.

In principle, any sample pair from different classes can be candidate for constructing $\mathbf{S}_b$, while any sample pair from the same class can be candidate for constructing $\mathbf{S}_w$, therefore the calculation of $\mathbf{S}_b$ as (4) is only one special instance, which only exploits the information in class centers.

Furthermore, the quantity of sample pairs is tremendous, and not all the sample pairs contribute to the learning of the discriminative subspace. Therefore, the construction of the scatter matrices should be based on the subset of "useful" sample pairs. Hence, we extends $\mathbf{S}_b$ to the *Negative Pair Scatter Matrix*, denoted as $\mathbf{S}_n$, and extends $\mathbf{S}_w$ to the *Positive Pair Scatter Matrix*, denoted as $\mathbf{S}_p$:

$$\mathbf{S}_n = \frac{1}{n_n}\sum_{(k_1,k_2)\in\Phi_n}(\mathbf{x}_{k_1} - \mathbf{x}_{k_2})(\mathbf{x}_{k_1} - \mathbf{x}_{k_2})^T, \quad (6)$$

$$\mathbf{S}_p = \frac{1}{n_p}\sum_{(k_1,k_2)\in\Phi_p}(\mathbf{x}_{k_1} - \mathbf{x}_{k_2})(\mathbf{x}_{k_1} - \mathbf{x}_{k_2})^T. \quad (7)$$

$$\Phi_n \subseteq \{(k_1,k_2)|l(\mathbf{x}_{k_1}) \neq l(\mathbf{x}_{k_2})\}, \quad (8)$$

$$\Phi_p \subseteq \{(k_1,k_2)|l(\mathbf{x}_{k_1}) = l(\mathbf{x}_{k_2})\}, \quad (9)$$

where n_n is number of negative pairs and n_p is number of positive pairs. $w(\mathbf{x}_{k_1}, \mathbf{x}_{k_2})$ is some form of weighting functions for adjusting the contribution of each constituent pair. It is easy to prove that $\mathbf{S}_b$ and $\mathbf{S}_w$ is a special case of $\mathbf{S}_n$ and $\mathbf{S}_p$, respectively.

Then the LDA can be reformulated as

$$\mathbf{W} = \underset{\mathbf{W}}{\text{argmax}} \frac{\text{tr}(\mathbf{W}^T\mathbf{S}_n\mathbf{W})}{\text{tr}(\mathbf{W}^T\mathbf{S}_p\mathbf{W})}. \quad (10)$$

The transform matrix can be solved similar to other LDA-based methods. We call the new formulation of LDA *Generalized Linear Discriminant Analysis*(*GLDA*).

III. Dynamic Feedback Framework

In this section, we introduce the Dynamic Feedback-based framework based on GLDA.

A. Initialization: K-Nearest Construction

As initialization, we should first construct a GLDA model. At this stage, we do not know precisely which sample pairs are actually useful, so we can select the sample pairs based on some heuristic rule. It has been shown that the pairs near the boundary bear more useful discriminative information and thus deserve emphasis [6]. Inspired by this rationale, we develop the K-nearest construction of negative pair scatter matrix $\mathbf{S}_n$, called *GLDA-KN*: for each sample, find K nearest samples that belong to other classes to form negative pairs.

$$\mathbf{S}_n = \frac{1}{KN}\sum_{k=1}^{N}\sum_{l=1}^{K}\frac{(\mathbf{x}_k - \mathbf{x}_{k_l})(\mathbf{x}_k - \mathbf{x}_{k_l})^T}{||(\mathbf{x}_k - \mathbf{x}_{k_l})||^2}, \quad (11)$$

where $\mathbf{x}_{k_1}, \mathbf{x}_{k_2}, \ldots, \mathbf{x}_{k_K}$ is the K nearest samples to $\mathbf{x}_k$ in the set $\{\mathbf{x}|l(\mathbf{x}) \neq l(\mathbf{x}_k)\}$. The denominator is used to normalize the contribution of each term.

As to the matrix $\mathbf{S}_p$, it can be initially constructed as traditional within-class scattering matrix.

B. Dynamic Procedure

The whole dynamic training procedure can be briefly described as follows:

---

1) Perform PCA on the training set, the dimension-reduced samples are denoted as $\{\mathbf{x}_k\}$, which are regarded as the training samples for GLDA model.
2) Construct the initial GLDA model using the GLDA-KN , the scatter matrices are denoted as $\mathbf{S}_n^{(0)}$ and $\mathbf{S}_p^{(0)}$, and the computed transform matrix is denoted as $\mathbf{W}^{(0)}$.

3) Iteratively optimize the model with dynamic feedback, for each step $t = 1, 2, \ldots$

   a) Transform all the training samples to the learned feature space: $\mathbf{y}_k^{(t)} = \mathbf{W}^{(t-1)}\mathbf{x}_k$.
   b) Calculate the distances for each pair of samples in the learned feature space.
   c) Compute Equal-Error-Rate (EER) on the training set. (EER is obtained by finding the threshold of distance value so that false accept rate and false reject rate are equal).
   d) Select the F negative pairs with smallest feature space distances, denoted as $\{(\mathbf{x}_{kn_{11}}, \mathbf{x}_{kn_{12}}), \ldots, (\mathbf{x}_{kn_{F1}}, \mathbf{x}_{kn_{F2}})\}$
   e) Select the F positive pairs with largest feature space distances, denoted as $\{(\mathbf{x}_{kp_{11}}, \mathbf{x}_{kp_{12}}), \ldots, (\mathbf{x}_{kp_{F1}}, \mathbf{x}_{kp_{F2}})\}$
   f) Apply the selected pairs to update scatter matrices

$$\Delta\mathbf{S}_n^{(i)} = \sum_{j=1}^{F} \frac{(\mathbf{x}_{kn_{j1}} - \mathbf{x}_{kn_{j2}})(\mathbf{x}_{kn_{j1}} - \mathbf{x}_{kn_{j2}})^T}{||\mathbf{x}_{kn_{j1}} - \mathbf{x}_{kn_{j2}}||^2}, \quad (12)$$

$$\Delta\mathbf{S}_p^{(i)} = \sum_{j=1}^{F} \frac{(\mathbf{x}_{kp_{j1}} - \mathbf{x}_{kp_{j2}})(\mathbf{x}_{kp_{j1}} - \mathbf{x}_{kp_{j2}})^T}{||\mathbf{x}_{kp_{j1}} - \mathbf{x}_{kp_{j2}}||^2}, \quad (13)$$

$$\mathbf{S}_n^{(i)} = \mathbf{S}_n^{(i-1)} + \Delta\mathbf{S}_n^{(i)}, \mathbf{S}_p^{(i)} = \mathbf{S}_p^{(i-1)} + \Delta\mathbf{S}_p^{(i)}. \quad (14)$$

   g) Compute $\mathbf{W}^{(t)}$ by Eq.(10); and the iteration stops if the EER keeps unchanged for several steps.

The principle of each step is to identify the pairs that are prone to misclassification by evaluating the similarities in the transformed feature space instead of the original space. With the participation of the model, the evaluation can reveal the model's deficiency and select the pairs carrying the most important complementary information. As the procedure advances, the sample pairs selected in feedback stage will gradually dominates the composition of scatter matrices, thus the model is gradually adapted to the training set.

IV. Experiments

We conduct experiments on a mixed face database where the samples are selected from XM2VTS [10] and PURDUE [11]. There are totally 385 different people, each person has 5 samples with different expressions. For each person, 3 of the 5 samples are used for training, the other 2 samples are for testing. In the 2 testing samples, one serves as gallery sample, the other serves as probe sample.

In the stage of preprocessing, affine transform is performed on all the samples to fix the positions of the two eyes and the mouth center. Each image is cropped to a 64×72 image and then histogram equalization is employed to reduce the effect of illumination variation. After that, we use a binary mask to filter out background interference.

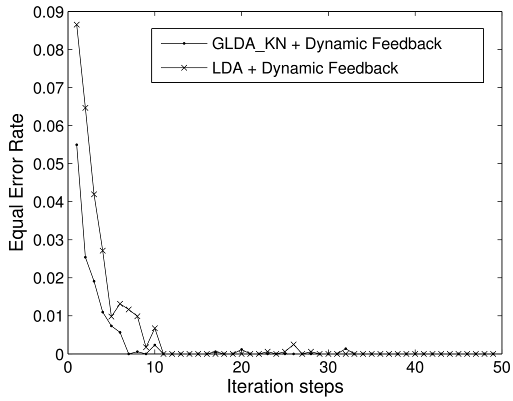

Fig. 1. The error rate vs the iteration number on training set for a) GLDA-KN + Dynamic feedback, and b) LDA + Dynamic feedback.

Each face is represented by a vector formed by scanning the pixel values from the image. Then, PCA model is trained and applied to reduce the space dimension. In our experiment, the 300 leading eigenvectors are selected with 98% of the variational energy retained in the principal subspace. And all LDA or GLDA models are trained based on the 300-dimension space.

Two models are trained with dynamic feedback framework, with one initialized by traditional LDA method, the other initialized by the GLDA-KN construction.

A. Performance in Training Stage

As mentioned above, at each iteration in the training stage, the equal error rate of applying the model to training set will be calculated. The trends of the empirical error rate is illustrated in figure 1. It shows that the iterative procedure drives the experimental error rate down to zero in a fairly fast speed, which indicates that our algorithm provides an effective way to optimize the classification performance of the model with respect to the training samples.

B. Performance on Testing set

The trained models are applied to an independent testing set to evaluate its generalization ability. Figure 2 shows the performance of the models in face identification. By using traditional LDA, the performance can only reach 83.64%, and a GLDA-KN model can reach 87.01%. As shown in Figure 2, in our proposed iterative procedure, both models gradually approach to an optimal state. The improvement is fairly encouraging: the two models achieves recognition accuracies at 94.51% and 94.83%, respectively. The recognition accuracy is raised by 11.8%.

Figure 3 shows the performance of the models in face verification, which also demonstrates the effectiveness of our algorithm: LDA and GLDA-KN initially yields equal error rates at 12.73% and 9.09% respectively. And the feedback-based procedure brings them down to 3.84% and 3.81%, the improvement is remarkable. The ratio of drop-down for error rate is up to 70% compared to traditional LDA!

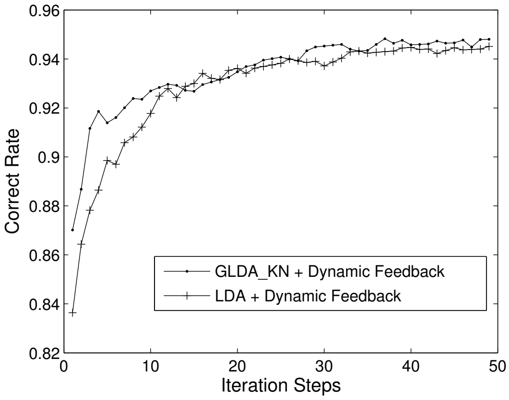

Fig. 2. Performance for face identification

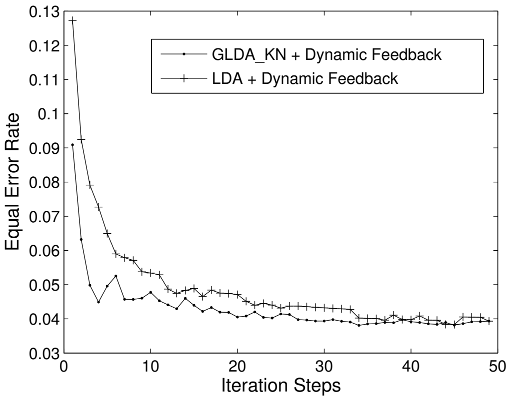

Fig. 3. Performance for face verification

The experiment results convincingly show that the dynamic feedback algorithm based on GLDA can effectively boost the performance of the model.

V. Conclusion

In this paper, a novel framework with feedback mechanism is presented, which is based on a generalized formulation of discriminant subspace analysis. Experiments show significant superiority of our proposed algorithm to the traditional LDA-based methods in both face identification and verification tasks.

ACKNOWLEDGEMENTS

The work described in this paper was fully supported by grants from the Research Grants Council of the Hong Kong SAR. The work was done while all the authors are with the Chinese University of Hong Kong.

References

[1] M.Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, 1991.

[2] D. L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. PAMI.*, vol. 18, no. 8, pp. 831–836, 1996.

[3] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. PAMI.*, vol. 19, no. 7, pp. 711–720, 1997.

[4] C. Liu and H. Wechsler, "Enhanced fisher linear discriminant models for face recognition," *Proceedings of ICPR'98*, vol. 2, pp. 2, 1998.

[5] X. Wang and X. Tang, "Dual-space linear discriminant analysis for face recognition," *Proceedings of CVPR'04*, vol. 2, pp. 564–569, 2004.

[6] M.Loog, R.P.W. Duin, and R. Haeb-Umbach, "Multiclass linear dimension reduction by weighted pairwise fisher criteria," *IEEE Trans. PAMI*, vol. 23, no. 7, pp. 762–766, 2001.

[7] X. Wang and X. Tang, "A unified framework for subspace face recognition," *IEEE Trans. PAMI*, vol. 26, no. 9, 2004.

[8] X. Wang and X. Tang, "Unified subspace analysis for face recognition," *Proceedings of ICCV'03*, 2003.

[9] K. Fukunaga, *Statistical Pattern Recognition*, Academic Press, 1990.

[10] J. Luettin and G. Maitre, "Evaluation protocol for the extended m2vts database (xm2vts).," *DMI for Perceptual Perceptual Artificial Intelligence*, 1998.

[11] A.M. Martinez and R. Benavente, "The ar face database," *CVC Technical Report 24*, 1998.