Adaptive Discriminant Analysis for face recognition from Single Sample per Person

Meina Kan[1,2], Shiguang Shan[1,2], Yu Su[3], Xilin Chen[1,2], Wen Gao[4,5]

[1]Digital Media Research Center, Institute of Computing Technology, CAS, Beijing, China
[2]Key Laboratory of Intelligent Information Processing, Chinese Academy of Sciences, Beijing, China
[3]Department of Computer Science, University of Caen, France
[4]Institute for Digital Media, Peking University, Beijing, China
[5]School of Computer Science and Technology, Harbin Institute of Technology, Beijing, China

{mnkan, sgshan}@jdl.ac.cn, ysu@info.unicaen.fr, {xlchen, wgao}@jdl.ac.cn

***Abstract*—Discriminant analysis, especially Fisherface and its numerous variants, have achieved great success in face recognition. However, these methods fail to work for face recognition from Single Sample per Person (SSPP), since they need more than one sample per person to estimate the within-class scatter matrix. To break this inability of traditional discriminant analysis, our paper proposes Adaptive Discriminant Analysis (ADA). In our method, the within-class scatter matrix of each enrolled subject is estimated from his/her single sample, by inferring from a generic training set with multiple samples per person. The inference is inspired by a simple intuition that similar person follows similar within-class variations. Specifically, both kNN regression and Lasso regression are explored for this purpose. We evaluate our method on FERET database and a large real-world face database. The results are very impressive compared with dominant traditional solutions to SSPP problem.**

Keywords-single sample per person; adaptive; discriminant analysis; lasso regression; face recognition

I. INTRODUCTION

Within the past two decades, face recognition has received increasing attentions owing to its wide range of potential applications, such as identity authentication, society security, surveillance, human-computer interface and so on. However, great challenges have been confronted in current face recognition systems due to large appearance variations caused by illumination, expression, age, pose and so on. Numerous methods have been proposed to improve face recognition as reviewed in [1]. These methods can be roughly divided into two categories: geometric-based methods and appearance-based methods [2]. Geometric-based methods describe a face image by the relation of components such as eyes, mouth and nose. Appearance-based methods represent face by both shape and texture information. In recent years, appearance-based methods have become dominant methods for face recognition.

Most appearance-based methods employ statistical learning technology, which commonly requires as many training samples as possible for each person. So, performance of appearance-based methods might be heavily affected by the number of training samples for each person. More specifically, intra-personal and inter-personal variations would be estimated with large bias in case of only few samples for each person. Even worse, if only single sample per person is available, many methods, such as the most popular Eigenface [3] and Fisherfaces [4], will degrade seriously or even fail to work. This is the so-called *Single Sample per Person* (SSPP) problem [5]. Hereinafter, we call the gallery with only single sample per person as *single sample gallery*.

As a great challenge, the SSPP problem has become a big obstacle to many real-world applications, such as e-passport, watch list screening, since, in these scenarios, it is generally impossible to collect more than one sample per person. To solve SSPP problem, many methods have been developed recently [6]. These methods can be roughly divided into three categories according to what information is used when training the recognition model. Methods in the first category only exploit the single sample gallery, i.e., the gallery is just the training set; the second category generates virtual images for each single image to obtain multiple samples per person, where many machine learning technologies can be applied; for the methods in the third category, besides the single sample gallery, an auxiliary generic training set containing multiple samples per person is collected and exploited to learn the recognition model.

Among methods using the gallery as the only training set, some early methods use point feature of the single image for classification [7, 8]. The local region based methods are most typical [9-15]. The single image is partitioned to blocks, then blocks are treated in isolation or together, and finally are combined to make final decisions, such as HMM [10], SOM [11] and so on. Some others are holistic-feature based methods, like the popular PCA-like (or Eigenface [3]) method. This method estimates the total-class scatter just using all the samples in the single sample gallery. In this case, the total class scatter actually degenerates to inter-class scatter. Many algorithms have been proposed to extend the classical PCA, such as 2DPCA [16], $(PC)^2A$ [17, 18], and Kernel PCA [19]. In [20] the optical flow between images is used to define an unequal feature weight based distance measure. In [21] Ahonen et al. propose a LBP feature representation for classification.

This paper is partially supported by Natural Science Foundation of China under contracts No.60833013, No.61025010, No. 60803084, and No.60872077; National Basic Research Program of China (973 Program) under contract 2009CB320902; and Hi-Tech Research and Development Program of China under contract No.2009AA01Z317.

An approach using multiple representations for each image is employed in [22]. In [23] Gao et al. decompose the image into two parts to estimate the within-class scatter and between-class scatter respectively. [24, 25] emphasize the neighboring information to obtain more discriminative low dimension feature representation for recognition. Although above methods do not suffer from the SSPP problem, however, most of them are unsupervised, and only exploiting inter-class variations without intra-class variations implies incompletely discriminative model.

In order to extract intra-class variations of the person with single sample, some researchers propose to synthesize virtual samples from the single sample of each person or partition a single image into sub-images. By doing like this, the single sample gallery is augmented to *multiple samples gallery*, thus many discriminant analysis methods can be applied. In [26, 27], new virtual images are obtain by learned information. [28-32] generate virtual face images via affine transformation, photometric changes, noise perturbation, shift and varying degrees of edge information. In [33], Huang et al. propose a component-based method that each face component is moved in four directions to generate virtual face images. In [34], Chen et al. directly partition the face images into several sub-images with the same dimensionality and treat these sub-images as multiple samples of each person. Some other researchers try to get the virtual images with novel pose, illumination and expression by first recovering the 3D face and then rendering it [35, 36]. Generally speaking, all above methods need prior knowledge to guide the generation of virtual images. Generating virtual images implies estimating "virtual" intra-personal variations, therefore, for these methods, how to guarantee the quality and reality of the virtual images is an open problem.

Intuitively, the faces of all human beings look similar, which implies different persons should also have similar intra-personal variations. Therefore, the within-class variations of the persons in the single sample gallery can be approximated by a generic training set containing as many persons as possible, which forms the motivation of the third category of solutions to SSPP [5, 37-42]. Typically as in [38], linear discriminant analysis model (both within-class and between-class scatter matrix) is learned on the generic training set and applied directly to the single sample gallery for feature extraction.

However, the variation distribution of the generic training set is often quite different from that of the single sample gallery. Therefore, the discriminant model learnt from the generic training set is more suitable to distinguish the persons in the generic training set but not those in the gallery. To address this drawback, Su et al. propose an adaptive generic learning method to estimate the within-class scatter of the single image [5], by the variance property that variance of independent random variables' sum equals to the sum of variance of each random variable.

In [5], the estimation of within-class scatter is obtained by all samples in the generic training set through least square regression. However in this paper, we propose that the within-class scatter of single sample can be estimated well just by several neighbors that are similar enough to the single sample.

Our method is based on a simple intuition: similar persons have similar intra-personal variations. So we can estimate the intra-personal variations of the subjects in the gallery by those of the persons in the generic training set who are similar enough to the single sample. Then, the variations of these neighboring generic subjects are combined to learn the final discriminant model. Specifically, two methods, k-Nearest Neighbors and Lasso Regression, are exploited to determine the neighbors and the weights for fusing their intra-personal variations.

The remainder of this paper is organized as follows. Section 2 describes the proposed Adaptive Discriminant Analysis with details about its basic idea, the formulation and the implementation. Section 3 presents the experimental evaluation of our method. Finally, conclusion is given in the last section.

II. Adaptive Discriminant Analysis for SSPP Face Recognition

In this section, we give a detailed description of the proposed Adaptive Discriminant Analysis (ADA). First, an overview of the basic idea is presented in subsection 2.1. Then how to infer the within-class scatter from single sample is formulated in the next subsection. Finally, the whole algorithm of ADA is summarized in subsection 2.3.

A. Basic Idea

Traditional Linear Discriminant Analysis (LDA) aims to find a set of most discriminative linear projections by maximizing the ratio of the determinant of the between-class scatter matrix to that of the within-class scatter matrix:

$$W_{opt} = \arg\max_{W} \frac{\left|W^t S_B W\right|}{\left|W^t S_W W\right|} \tag{1}$$

The within-class scatter matrix S_W and between-class scatter matrix S_B are defined as:

$$S_W = \sum_{i=1}^{c} \sum_{x \in X_i} (x - m_i)(x - m_i)^t \tag{2}$$

$$S_B = \sum_{i=1}^{c} N_i (m_i - m)(m_i - m)^t \tag{3}$$

where c is the number of classes in the training set, X_i is the set of images of the i-th person, N_i is the number of samples in X_i, m_i is the mean of all samples in X_i, and m is the mean of all the samples in the training set.

In case of SSPP, Sw degenerates to $\boldsymbol{0}$, thus making LDA fail to work. To address this drawback, as illustrated in Fig. 1, we introduce an auxiliary generic training set with multiple samples per person. Then, given a gallery set with single sample per person, we propose a novel method to predict the within-class scatter matrix of each person in the gallery by inferring from its single sample. Briefly speaking, our method is based on a rational intuition: similar persons are prone to have similar intra-personal variations. So we can approximate the intra-personal variations of one person by those of the persons in the generic training set who are similar enough to him/her. In this study, two methods, k-Nearest Neighbors and Lasso Regression, are exploited to determine the neighbors and the weights for fusing their intra-personal variations. The next subsection details the procedure.

B. *Inferring Within-class Scatter of one Person from its Single Sample*

The key procedure in our method is inferring the within-class scatter of each person in the gallery from his/her single sample. As mentioned above, our basic idea is predicting from some similar persons selected from the generic training set, as illustrated in Fig. 2. Given a single sample of person *i*, our method performs two basic steps. The first step is finding similar persons (referred to as neighbors) in the generic training set for person *i*, as shown in Fig. 2(a). For each selected similar person, a weight is also computed for later use. Then, in the second step, the within-class scatter matrix of the *i*-th person is computed by a linear combination of the within-class scatter matrix of the selected neighbors in the first step, using the learnt weights, as shown in Fig. 2(b). The method is formally formulated as follows.

We denote the gallery with single sample per person as:

$$G = [x_1^g, x_2^g, \cdots, x_i^g, \cdots, x_M^g] \in R^{d \times M} \tag{4}$$

where M is the number of samples(also persons), d is the feature dimension, and x_i^g is the sample of the i-th person in the gallery.

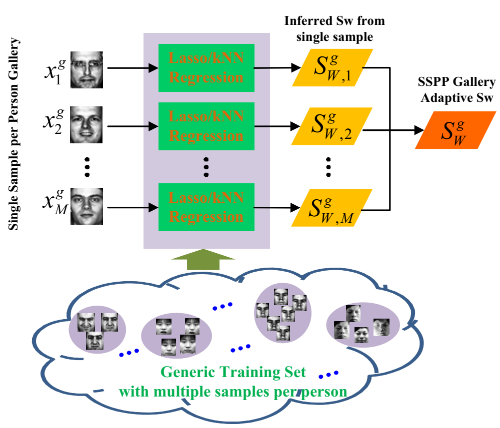


Figure 1. Illustration of the proposed Adaptive Discriminant Analysis.

As mentioned above, a generic training set containing multiple samples per person is needed in our method. We denote it as:

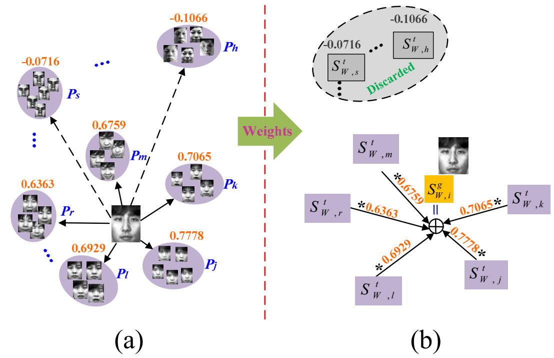


Figure 2. Illustration of inferring within-class scatter of one single sample. In this example, we keep five nearest neighbors for the single sample. (a) Step 1: finding k nearest neighbors in the generic training set for person *i*; (b) Step 2: inferring within-class scatter matrix by linear combination for person *i*.

$$\begin{aligned} A = [&x_{1,1}^t, x_{1,2}^t, \cdots, x_{1,N_1}^t; x_{2,1}^t, \cdots, x_{2,N_2}^t; \cdots; x_{i,1}^t, \cdots, \\ &x_{i,j}^t, \cdots, x_{i,N_i}^t; \cdots; x_{c,1}^t, \cdots, x_{c,N_c}^t] \in R^{d \times N} \end{aligned} \tag{5}$$

where x_{ij}^t is the j-th sample of the i-th person in the generic training set, c is the total number of the persons in the generic training set, N_i is the number of samples corresponding to the i-th person, and N is the total number of samples in the generic training set.

After defining the single sample gallery and the generic training set, we then detail the two steps for inferring within-class scatter of a single sample (please refer to Fig. 2).

Step 1: finding neighbors and corresponding weights

In this step, we explore two alternative methods: kNN and Lasso regression. For kNN method, given x_i^g (the sample of the i-th person in the single sample gallery), first we simply compute its *k nearest neighbors* in the generic training set. Then, we assign a weight for each person in the generic training set. Formally, the weight for the j-th person is set as:

$$\omega_j^i = \begin{cases} s\left(x_i^g, \bar{x}_j^t\right), & \bar{x}_j^t \in kNN\ of\ x_i^g \\ 0\quad, & otherwise \end{cases} \tag{6}$$

where

$$\bar{x}_j^{\,t} = \frac{1}{N_j} \sum_{l=1}^{N_j} x_{j,l}^t \tag{7}$$

and $s(.)$ is the similarity measurement of person x and person y,.

KNN is simple yet without guarantee of non-correlation among the k neighbors. So, Lasso regression is adopted alternatively. Formally, the weights for all the c persons in the generic training set are simultaneously optimized by the following minimizing procedure [43]:

$$\left(\omega_1^i, \omega_2^i, \cdots, \omega_c^i\right) = \arg \min_{\omega_1, \omega_2, \cdots, \omega_c} \left(\left\| x_i^g - \sum_{j=1}^{c} \omega_j \overline{x}_j^t \right\|_2 \right)$$
$$subject\ \ to \sum_{j=1}^{c} \left\| \omega_j \right\|_1 \leq t \tag{8}$$

Here, it is worth pointing out that, in Lasso regression, we give no exact definition of nearest neighbors. Nevertheless, in the cost function of Lasso regression, the constraint term can lead to many *0* weights. All persons with non-zero weights can be regarded as "neighbors".

Step 2: inferring within-class scatter by linear combination

With the nearest neighbors and their corresponding weights computed, it is then quite simple to estimate the within-class scatter matrix $S_{W,i}^g$ of the *i*-th person with single sample x_i^g , by the following linear combination of the within-class scatter matrix of his/her neighbors:

$$S_{W,i}^g = \sum_{j=1}^{C} \omega_j^i S_{W,j}^t \tag{9}$$

where $S_{W,j}^t$ is the within-class scatter matrix of the *j*-th person in the generic training set.

C. Gallery Adaptive Discriminant Analysis

With the within-class scatter of each person in the single sample gallery set inferred by (9), the total gallery-adaptive within-class scatter matrix can be computed as follows:

$$S_W^g = \sum_{i=1}^{M} S_{W,i}^g \tag{10}$$

Meanwhile, the total between-class scatter suitable for the single sample gallery can also be calculated just by the samples in the gallery [5]:

$$S_B^g = \sum_{i=1}^{M} (x_i^g - m_g)(x_i^g - m_g)^t \tag{11}$$

$$m_g = \frac{1}{M} \sum_{i=1}^{M} x_i^g \tag{12}$$

Now, we can apply the LDA with the estimated within-class scatter and the actual between-class scatter as bellows:

$$W_{opt} = \arg \max_{W} \frac{\left| W^t S_B^g W \right|}{\left| W^t S_W^g W \right|} \tag{13}$$

A summarizing algorithm of the proposed Adaptive Discriminant Analysis is presented in Table. 1.

TABLE I. ALGORITHM OF ADAPTIVE DISCRIMINANT ANALYSIS

Algorithm. Adaptive Discriminant Analysis

Input: a single sample gallery *G* and a generic training set *A*.

$$G = [x_1^g, x_2^g, \cdots, x_i^g, \cdots, x_M^g] \in R^{d \times M}$$

$$A = [x_{1,1}^t, x_{1,2}^t, \cdots, x_{1,N_1}^t; x_{2,1}^t, \cdots, x_{2,N2}^t; \cdots; x_{i,1}^t, \cdots, x_{i,j}^t, \cdots, x_{i,N_i}^t; \cdots; x_{c,1}^t \cdots x_{c,N_c}^t] \in R^{d \times N}$$

Step 1: Compute the $S_{w,j}^t$ and $\overline{x}_j^t$ for each class in the generic training set.

Step 2: for each x_i^g in the single sample gallery:

a) Find its neighbors and corresponding weights by kNN or Lasso regression according to (6) or (8).
b) Estimate $S_{w,i}^g$ according to (9).

Step 3: Compute S_W^g by (10), and S_B^g by (11).

Step 4: Learn LDA model by (13).

III. EXPERIMENTS

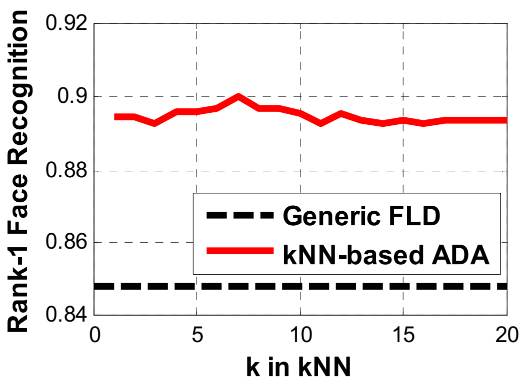
In this section, we evaluate our proposed ADA method dealing with the SSPP problem on two large databases, FERET and a passport-like private real-world face database. We first evaluate the influence of the main parameters in our method and then compare with existing methods for SSPP problem.

A. Databases for Evaluation

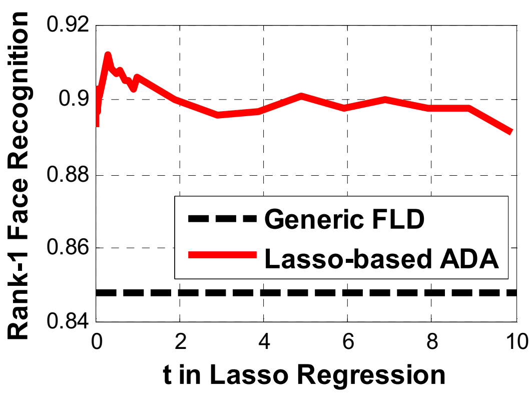
In our experiment, two testing databases are involved. One is FERET database [44], whose gallery contains 1196 images, one image per person. According to the FERET evaluation protocol, there are four probe sets: fafb, fafc, duplicate I (denoted as dupI), and duplicate II (denoted as dupII). The images in fafb set are with expression variations, fafc set contains images with lighting variations, and the images in dupI and dupII sets were acquired some days later.

The other testing database is our own private database, which contains passport-like face images. The gallery consists of 3000 persons with a single image per person, and the probe set includes 4190 images. Some of examples are shown in Fig. 3. Here, we must point out that, the images for the same person in this database were acquired at several years' interval and with various image acquiring devices. Therefore, this testing forms a quite challenging SSPP scenario.

Figure 3. Examples of the private database. Images in the same column correspond to the same person.



(a) kNN-based ADA



(b) Lasso-based ADA

Figure 4. Rank-1 face recognition rate corresponds to different parameters in our methods based on gray-intensity on fafb of FERET database. (a) result with different k for kNN. (b) result with different *t* for lasso regression.

For both testing databases, we use the same generic training set, which consists of images from two databases: XM2VTS [45] and the training set of CAS-PEAL [46]. The XM2VTS database contains 3440 images of 295 persons taken over a period of four months with slight head pose and illumination variations. The CAS-PEAL training set contains 1200 images of 300 persons. The images cover larger variations due to pose, expression, lighting and so on. So, by uniting XM2VTS and CAS-PEAL, we obtain a generic training set with 4640 images of 595 persons.

In our experiments, all face images are normalized to 40*50 pixels according to manual-labeled eye locations and preprocessed by histogram equalization.

B. *Influence of Parameters*

In our method, the main parameters needed to be determined are k for kNN-based ADA or *t* in Lasso-based ADA. In this subsection, we show how performance changes with different parameters on *fafb* which is the largest test set of FERET. As shown in Fig. 4, it can be seen that our method can achieve better performance in a wide range of parameter values. Note that, in the figure, Generic FLD stands for Fisher Linear Discriminant (FLD) method trained on the generic training set and tested on the probe set. From this figure, it can also be seen that k is better set to a number smaller than 10, while *t* in Lasso can be set to a value smaller than 2.0.

C. *Comparison with Other Methods*

In this subsection, we compare our method with other seven typical methods dealing with SSPP problem. Among them, PCA, $(PC)^2A$, LBP, SVD-based FLDA and Block FLD only use information from the single sample gallery; Generic FLD only uses the generic training set information without adaptation; AGL and our ADA utilizes the information from the single sample gallery to adapt the model learnt from generic training set. The basic information about these methods is described in the following.

1) *PCA*[3]**.** PCA is trained directly on the single sample gallery so as to get better performance, with 500 dimensions kept. 2) *$(PC)^2A$*[17]. The weight of projection-combined face images is set to 0.3 according to [17]. $(PC)^2A$ is also trained on the single sample gallery. 3) *LBP*[21]. Since the partition of images is important to LBP, we try different numbers of image blocks, and present the best result with 80 image blocks trained on the single sample gallery. 4) *SVD-based FLDA*[23]. Since the width and height of image are not equal in our experiment, so the within-class scatter degenerates to $Sw = (1/C)\sum_{k=1}^{C}\left[\left(A_k - \bar{A}_k\right)^T\left(A_k - \bar{A}_k\right)\right]$. 5) *Block FLD*[34]. For Block FLD, the key parameter is size of image blocks; we test four different sizes, and report the best result with size 10*25. 6) *Generic FLD*. Fisherface trained on the generic training set. Best result is represented with testing all dimensions 7) *FERET-Specific FLD*[4]. Fisherface trained on training set of FERET. In should be noted that, training set of FERET contains images collected in the same condition as the test set of FERET, not the images of the persons in test set. Models obtained by this method are employed to just simulate the ground truth model and show how the proposed ADA can approach the ground truth, so it should not be compared to other methods. 8) *AGL*[5]**.** The PCA dimension is set same as in [5] 9) *kNN ADA*. Our proposed method that uses k-Nearest Neighbors to find neighbors. We test several k and report the best result. Here, we take the simple cosine similarity to find neighbors and also as combination weights. 10) *Lasso ADA*. Our proposed method that uses Lasso Regression to find neighbors. We test several *t* and report the best result.

In addition, in recent years, LDA is often combined with Gabor features to further improve the accuracy of the face recognition systems [47-49]. Therefore, we also validate the method combining the ADA with Gabor features (only the magnitude part). In our implementation, to reduce the dimensionality of Gabor features for LDA, we partition the face image into 4 blocks (each with the size of 20×25 pixels) and then train 4 Gabor-ADA classifiers respectively, which are finally combined by sum rule. Hereinafter the method is denoted as Gabor-ADA. Obviously, the Generic FLD can also be enhanced by this strategy, which is denoted as Generic Gabor-FLD hereinafter.

Table 2 gives comparison results of above methods on FERET and our private real-world face databases based on gray and Gabor features. As shown in Table 2, we can see that 1) methods only using the single sample gallery or only the generic training set, such as PCA, $(PC)^2A$, LBP, Block FLD and Generic FLD, perform not very well on most testing sets; However, the adaptive-based methods performs much better, such as AGL and ADA. 2) the proposed ADA with an adaption gets an significant improvement especially on the more challenging private real-world database. 3) by combining with Gabor feature, ADA can be further improved. 4) Compared to the *FERET-Specific FLD* which is employed to stands for the 'ground truth model', ADA achieves a lower performance on fafb and dupI, however surprisingly performs higher on fafc and dupII. This illustrates that our ADA can estimate a comparable within-class scatter of single sample to that of the 'ground truth model'.

TABLE II. RANK-1 FACE RECOGNITION RATES ON FERET AND THE PRIVATE REAL-WORLD DATABASE. THE LAST THREE ROWS DENOTE RESULTS BASED ON GABOR FEATURES, OTHERS DENOTE RESULTS BASED ON IMAGE GRAY-INTENSITY. THE BOLD DENOTES RESULTS OF THE PROPOSED ADA.

Methods	FERET				Real-world database
	fafb	fafc	dup I	dup II	
PCA	0.896	0.134	0.399	0.150	0.168
(PC)2A	0.896	0.144	0.404	0.150	0.170
LBP	0.976	0.557	0.575	0.329	0.335
Block FLD	0.783	0.485	0.432	0.321	0.393
Generic FLD	0.841	0.675	0.475	0.235	0.263
SVD-based FLDA	0.833	0.253	0.314	0.120	0.202
AGL	0.890	0.720	0.515	0.350	0.455
KNN ADA	**0.901**	**0.748**	**0.525**	**0.368**	**0.520**
Lasso ADA	**0.912**	**0.758**	**0.519**	**0.372**	**0.508**
***FERET-Specific FLD**	**0.980**	**0.711**	**0.616**	**0.316**	**--**
Generic Gabor-FLD	0.964	0.892	0.673	0.474	0.465
Gabor-AGL	0.981	0.938	0.728	0.581	0.641
KNN Gabor-ADA	**0.981**	**0.943**	**0.738**	**0.603**	**0.632**
Lasso Gabor-ADA	**0.986**	**0.959**	**0.747**	**0.615**	**0.678**

* stands for the result of the 'ground truth model' which is a reference value used to measure the approximate of ADA.

IV. CONCLUSION

To break the inability of traditional discriminant analysis to deal with the single sample per person problem, we propose Adaptive Discriminant Analysis (ADA). In our method, the within-class scatter matrix of the single sample gallery is inferred by combining the within-class scatter matrix of a number of persons selected from the generic training set who are similar to the persons in the gallery. Thus, the learnt gallery-Adaptive Discriminant Analysis model not only exploits the between-class discriminative information among the gallery samples but also adaptively borrows the within-class scatter variations from the generic training set. Experiments on FERET and a large real-world dataset validate the proposed method.

REFERENCES

[1] W. Y. Zhao, R. Chellappa, P. J. Phillips, and A. P. Rosenfeld, "Face recognition: A literature survey," ACM Computing Surveys, vol. 35, pp. 399-458, 2003.

[2] R. Brunelli and T. Poggio, "Face recognition: features versus templates," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, pp. 1042-1052, 1993.

[3] M. Turk and A. Pentland, "Eigenfaces for Recognition," Journal of Cognitive Neuroscience, vol. 3, pp. 71-86, 1991.

[4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, pp. 711-720, 1997.

[5] Y. Su, S. Shan, X. Chen, and W. Gao, "Adaptive generic learning for face recognition from a single sample per person," in IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 3699-2706.

[6] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Face recognition from a single image per person: A survey," Pattern Recognition, vol. 39, pp. 1725-1745, 2006.

[7] K.-M. Lam and H. Yan, "An analytic-to-holistic approach for face recognition based on a single frontal view," IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 20, pp. 673-686, 1998.

[8] Y. Gao and Y. Qi, "Robust visual similarity retrieval in single model face databases," Pattern Recognition, vol. 38, pp. 1009-1020, 2004.

[9] A. M. Martinez, "Recognizing Imprecisely Localized, Partially Occluded, and Expression Variant Faces from a Single Sample per Class," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp. 748-763, 2002.

[10] H.-S. Le and H. Li, "Recognizing frontal face images using Hidden Markov models with one training image per person," in International Conference on Pattern Recognition, 2004, pp. 318-321.

[11] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft k-NN ensemble," IEEE Transactions on Neural Networks, vol. 16, pp. 875-886, 2005.

[12] H. R. Kanan, K. Faez, and Y. Gao, "Face recognition using adaptively weighted patch PZM array from a single exemplar image per person," Pattern Recognition, vol. 41, pp. 3799-3812, 2008.

[13] H. R. Kanan and Y. Gao, "Recognition of expression variant faces from one sample image per enrolled subject," in IEEE International Conference on Image Processing (ICIP), 2009, pp. 3309-3312.

[14] H. R. Kanan and M. S. Moin, "Face recognition using Entropy Weighted Patch PCA Array under variation of lighting conditions from a single sample image per person," in Information, Communications and Signal Processing, Macau, 2009, pp. 1-5.

[15] H. R. Kanan and K. Faez, "Recognizing faces using Adaptively Weighted Sub-Gabor Array from a single sample image per enrolled subject," Image and Vision Computing, vol. 28, pp. 438-448, 2010.

[16] J. Yang, D. Zhang, A. F. Frangi, and J.-y. Yang, "Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and

Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, pp. 131-137, 2004.

[17] J. Wu and Z.-H. Zhou, "Face recognition with one training image per person," Pattern Recognition Letters, vol. 23, pp. 1711-1719, 2002.

[18] S. Chen, D. Zhang, and Z.-H. Zhou, "Enhanced (PC)2A for face recognition with one training image per person," Pattern Recognition Letters, vol. 25, pp. 1173-1181, 2004.

[19] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem," Neurocomputing, vol. 10, pp. 1299-1319, 1998.

[20] A. M. Martinez, "Recognizing expression variant faces from a single sample image per class," in IEEE Conference on Computer Vision and Pattern Recognition, 2003, pp. I-353 - I-358.

[21] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face Recognition with Local Binary Patterns," in European Conference on Computer Vision, Lecture Notes in Computer Science, 2004, pp. 469-481.

[22] F. D. l. Torre, R. Gross, S. Baker, and B. V. K. V. Kumar, "Representational oriented component analysis (ROCA) for face recognition with one sample image per training class," in IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 266-273.

[23] Q.-x. Gao, L. Zhang, and D. Zhang, "Face recognition using FLDA with single training image per person," Special Issue on Advanced Intelligent Computing Theory and Methodology in Applied Mathematics and Computation, vol. 205, pp. 726-734, 2008.

[24] L. Qiao, S. Chen, and X. Tan, "Sparsity preserving discriminant analysis for single training image face recognition," Pattern Recognition Letters, vol. 31, pp. 422-429, 2010.

[25] W. Deng, J. Hu, J. Guo, W. Cai, and D. Feng, "Robust, accurate and efficient face recognition from a single training image: A uniform pursuit approach," Pattern Recognition, vol. 43, pp. 1748-1762, 2010.

[26] D. Beymer and T. Poggio, "Face recognition from one example view," International Conference on Computer Vision, pp. 500-507, 1995.

[27] A. Sharma, A. Dubey, P. Tripathi, and V. Kumar, "Pose invariant virtual classifiers from single training image using novel hybrid-eigenfaces," Neurocomputing, vol. 73, pp. 1868-1880, 2010.

[28] H.-C. Jung, B.-W. Hwang, and S.-W. Lee, "Authenticating Corrupted Face Image Based on Noise Model," in IEEE International Conference on Automatic Face and Gesture Recognition, 2004, pp. 272-277.

[29] D. Zhang, S. Chen, and Z.-H. Zhou, "A new face recognition method based on SVD perturbation for single example image per person," Applied Mathematics and Computation, vol. 163, pp. 895-907, 2005.

[30] J. Liu, S. Chen, Z.-H. Zhou, and X. Tan, "Single Image Subspace for Face Recognition," in Analysis and Modeling of Faces and Gestures, Lecture notes in Computer Science, 2007, pp. 205-219.

[31] A. Majumdar and R. K. Ward, "Single image per person face recognition with images synthesized by non-linear approximation," in International Conference on Image Processing, 2008, pp. 2740-2743.

[32] S. Shan, B. Cao, W. Gao, and D. Zhao, "Extended Fisherface for face recognition from a single example image per person," in IEEE International Symposium on Circuits and Systems, 2002, pp. II-81 - II-84.

[33] J. Huang, P. C. Yuen, W.-S. Chen, and J. H. Lai, "Component-based LDA Method for Face Recognition with One Training Sample," in IEEE International Workshop on Analysis and Modeling of Faces and Gestures, 2003, pp. 120-126.

[34] S. Chen, J. Liu, and Z.-H. Zhou, "Making FLDA applicable to face recognition with one sample per person," Pattern Recognition, vol. 37, pp. 1553-1555, 2004.

[35] T. Vetter, "Synthesis of Novel Views from a Single Face Image " International Journal of Computer Vision, vol. 28, pp. 103-116, 1998.

[36] P. Niyogi, F. Girosi, and T. Poggio, "Incorporating prior information in machine learning by creating virtual examples " Proceedings of the IEEE vol. 86, pp. 2196-2209, 1998.

[37] J. Wang, K. N. Plataniotis, and A. N. Venetsanopoulos, "Selecting discriminant eigenfaces for face recognition," Pattern Recognition Letters, vol. 26, pp. 1470-1482, 2005.

[38] J. Wang, K. N. Plataniotis, J. Lu, and A. N. Venetsanopoulos, "On solving the face recognition problem with one training sample per subject," Pattern Recognition, vol. 39, pp. 1746-1762, 2006.

[39] L. Zhang and D. Samaras, "Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, pp. 351-363, 2006.

[40] A. Majumdar and R. K. Ward, "Pseudo-Fisherface method for single image per person face recognition," in IEEE International Conference on Acoustics, Speech and Signal Processing, 2008, pp. 989-992.

[41] L. Zhu, Y. Jiang, and L. Li, "Making discriminative common vectors applicable to face recognition with one training image per person," in IEEE Conference on Cybernetics and Intelligent Systems, 2008, pp. 385-387.

[42] S. Chen, C. Sanderson, S. Sun, and B. C. Lovell, "Representative feature chain for single gallery image face recognition," in International Conference on Pattern Recognition, 2008, pp. 1-4.

[43] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction: Springer, 2001.

[44] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," Image and Vision Computing, vol. 16, pp. 295-306, 1998.

[45] K. Messer, J. Matas, J. Kittler, J. Lüttin, and G. Maitre, "XM2VTSDB: The Extended M2VTS Database," in Second International Conference on Audio and Video-based Biometric Person Authentication, 1999, pp. 72-77.

[46] W. Gao, et al., "The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations," IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, vol. 38, pp. 149-161, 2008.

[47] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," IEEE Transactions on Image Processing, vol. 11, pp. 467-476, 2002.

[48] L. Shen, L. Bai, and M. Fairhurst, "Gabor wavelets and General Discriminant Analysis for face identification and verification," Image and Vision Computing, vol. 25, pp. 553-563, 2007.

[49] Z. Li, W. Liu, D. Lin, and X. Tang, "Nonparametric subspace analysis for face recognition," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, pp. 961-966.