Adaptive Generic Learning for Face Recognition from a Single Sample per Person

Yu Su[1,2] Shiguang Shan[2,3] Xilin Chen[2,3] Wen Gao[1,4]
[1] School of Computer Science and Technology, Harbin Institute of Technology, China
[2] Digital Media Research Center, Institute of Computing Technology, CAS, China
[3] Key Laboratory of Intelligent Information Processing, Chinese Academy of Sciences, China
[4] Institute for Digital Media, Peking University, China
{ysu, sgshan, xlchen, wgao}@jdl.ac.cn

Abstract

Real-world face recognition systems often have to face the single sample per person (SSPP) problem, that is, only a single training sample for each person is enrolled in the database. In this case, many of the popular face recognition methods fail to work well due to the inability to learn the discriminatory information specific to the persons to be identified. To address this problem, in this paper, we propose an Adaptive Generic Learning (AGL) method, which adapts a generic discriminant model to better distinguish the persons with single face sample. As a specific implementation of the AGL, a Coupled Linear Representation (CLR) algorithm is proposed to infer, based on the generic training set, the within-class scatter matrix and the class mean of each person given its single enrolled sample. Thus, the traditional Fisher's Linear Discriminant (FLD) can be applied to SSPP task. Experiments on the FERET and a challenging passport face database show that the proposed method can achieve better results compared with other common solutions to the SSPP problem.

1. Introduction

Face recognition from still images and video sequence has been an active research topic due to its scientific challenges and wide range of potential applications, such as biometric identity authentication, human-computer interaction, and video surveillance. The challenges of face recognition mainly come from the large variations in the visual stimulus due to illumination conditions, viewing directions, facial expressions, aging, and disguises. Within the past two decades, numerous face recognition methods have been proposed to deal with these challenging problems, as reviewed in the literature survey [1]. These methods can be roughly divided into two categories: geometric-based methods and appearance-based methods [2]. The former generally represent a face image by the relative position and other parameters of some distinctive features such as eyes, mouth, nose, and chin. In contrast, for appearance-based methods, a face image is treated holistically as a sample in the image space. Since 1990s, the appearance-based methods had dominated the face recognition field due to their good performance and simplicity. Among them, the most popular ones are subspace-based methods, such as the Eigenfaces (based on Principal Component Analysis, PCA) [3] and the Fisherfaces (based on Fisher's Linear Discriminant, FLD) [4].

However, the performance of the appearance-based methods is heavily affected by the number of training samples for each person [5]. More specifically, if the number of training samples is much smaller than the feature dimensionality, the estimation of intra-personal and inter-personal variations would be generally inaccurate. Especially, if only one training sample is available for each person, the intra-personal variations can not be estimated at all. In this case, for instance, FLD will degenerate to PCA. This is the so-called Single Sample Per Person (SSPP) problem in face recognition. This problem has prevented those discriminant analysis methods from successful application to real-world face recognition scenarios, such as e-passport and ID card verification. In these applications, there is usually only a single training sample for each person, since it is generally very difficult or even impossible to collect additional samples.

In the literature, some methods were proposed to deal with the SSPP problem, as reviewed in [6]. These methods can be coarsely divided into three categories: unsupervised learning, virtual sample generation and generic learning. We introduce them one by one briefly in the follows.

In the first category, the unsupervised techniques are used to circumvent the SSPP problem. These methods generally do not need the labels of the training samples and thus neglect the utilization of intra-personal variations. The most representative method of this category is PCA (or Eigenfaces), whose goal is to find the low dimensional subspace with maximum data variance. PCA had already become the baseline algorithm of face recognition and been extended to many different versions, such 2DPCA [7], $(PC)^2A$ [8] and Kernel PCA [9]. Besides the PCA-based methods, in [10], Tan et al. proposed another unsupervised method based on the Self-Organizing Map (SOM), and reported higher recognition rate than PCA. Although the unsupervised methods do not suffer from the SSPP problem, they only utilize the inter-personal variations and fail to make use of the intra-personal variations; therefore, their performance might be poor if the face images include large variations in expressions or lighting conditions.

In order to extract the discriminatory information embedded in the intra-personal variations, some researchers proposed to generate some extra samples for each person in the database. In [11], Martinez proposed a perturbation-based approach to generating virtual face images. In [12], Shan et al. extended Fisherfaces for SSPP problem by generating virtual face images via geometric transform and photometric changes. In [13], Huang et al. proposed a component-based method, in which each local facial region is moved in four directions to generate more training samples for each person. In [14], Chen et al. proposed to partition each face image into a set of sub-images with the same dimensionality, therefore obtaining multiple training samples for each person. With these extra training samples, the traditional FLD-based methods can be applied. In [15], Gao et al. utilized SVD to decompose each face image into two complementary parts: a smooth general appearance image and a difference image. The later is used to approximate the intra-personal variations. In addition, considering the 3D nature of the human face, many research efforts focus on generating the virtual face views with novel pose, lighting and expression, such as in [16, 17, 18]. Overall speaking, these methods are basically either heuristic or highly dependent on prior information about the human face. Furthermore, in real-world applications, determining what kind of and how many virtual samples need to be generated is not a trivial task.

For the methods of the third category, a generic training set, in which each person has more than one training sample, is adopted to extract the discriminatory information. Then, this generic discriminatory information is directly used to identify the persons with only one training sample. Some previous works under this framework include [19, 20, 21, 22]. For example, in [19], Wang et al. presented a generic learning framework and adopted many feature extraction methods to extract the discriminatory information from a generic training set. In [21], Kim and Kittler proposed a solution to the pose-invariant face recognition problem from a single frontal face image by collecting a generic training set to extract a pose-invariant subspace.

The underlying assumption of the generic learning methods is that, both the intra-personal variations of different persons and the inter-personal variations for different populations (sets of persons) are similar, and thus can be approximated by estimating from a generic large population. However, this assumption is too strong in many cases, especially in the case of populations containing persons of different skin colors, ages, or even occupations. Therefore, we argue that the discriminatory information embedded in the generic training set should be adapted to identify other persons.

Based on the above analysis, in this paper, we propose the Adaptive Generic Learning (AGL) method for face recognition from SSPP. Unlike generic learning methods, AGL does not directly employ the discriminatory information learned from the generic training set, but adapts it to the persons to be identified (i.e., the persons in the gallery). Specifically, for each person (with only a single sample) in the gallery, AGL attempts to infer its intra-personal variations and mean by a predicting model learned from the generic training set. With the predicted intra-personal variations and mean of each enrolled person, the overall within-class and between-class scatter matrix can be estimated for the persons in the specific gallery, which thus makes FLD applicable to the SSPP scenario.

The proposed method is evaluated on both the FERET face database and a passport face database under the SSPP scenario. Experimental results show that the proposed AGL method significantly outperforms the traditional generic learning method as well as other common solutions to the SSPP problem.

2. AGL: Problem Description and Basic Ideas

As mentioned above, this paper aims at applying FLD to face recognition under SSPP scenario. For clarity, we first give a brief introduction to the traditional FLD and the generic learning based methods, then the basic idea of the proposed AGL is presented.

In the traditional FLD, the within-class scatter matrix ($\mathbf{S}_W$) and the between-class scatter matrix ($\mathbf{S}_B$) are used to measure the class separability. They are defined as,

$$\mathbf{S}_W = \sum_{i=1}^{C} \mathbf{S}_i \ , \ \mathbf{S}_i = \sum_{\mathbf{x} \in X_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T \ , \tag{1}$$

$$\mathbf{S}_B = \sum_{i=1}^{C} N_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \ , \tag{2}$$

where C is the number of classes; $\mathbf{S}_i$ is the within-class scatter matrix of the i-th class; N_i is the number of samples in class X_i; $\mathbf{m}_i$ is the mean vector of class X_i; $\mathbf{m}$ is the mean vector of all the samples. The FLD is then formulated as the solution of the following optimization problem:

$$\mathbf{W}_{opt} = \arg\max_{\mathbf{W}} \frac{\left|\mathbf{W}^T \mathbf{S}_B \mathbf{W}\right|}{\left|\mathbf{W}^T \mathbf{S}_W \mathbf{W}\right|} \ . \tag{3}$$

Mathematically, this ratio is maximized when the column vectors of the projection matrix $\mathbf{W}$ are the eigenvectors of $\mathbf{S}_W^{-1}\mathbf{S}_B$ [23] if $\mathbf{S}_W$ is non-singular.

As mentioned previously, in the case of SSPP scenario, for each enrolled person in the gallery, there is only one sample. Formally, we denote the gallery as $X^g = \{\mathbf{x}_k^g : k = 1, 2, ..., M\}$, where $\mathbf{x}_k^g$ is the face image (or sample) of the k-th person (or class) in the gallery. Evidently, with this gallery, FLD can not be applied because the $\mathbf{S}_W$ degenerates to a zero matrix in this case.

To address the SSPP problem, the generic learning method employs a generic training set. The set is formally denoted as $X^t = \{\mathbf{x}_{ij}^t : i = 1, 2, ..., C; j = 1, 2, ..., N_i\}$, where $\mathbf{x}_{ij}^t$ is the j-th face image of the i-th person in the generic training set, and N_i is the sample number of the i-th person. Clearly, in the generic training set, each person should have more than one training sample. Then, FLD model is learned on this generic training set and applied to identify the persons in the gallery. Evidently, the underlying assumption of this method is that, the within-class and between-class scatter matrix of the persons in the gallery are very similar to and thus can be approximated by those of the generic training set. To meet this assumption, the population and variations in the generic training set should be as similar as possible to those in the gallery and the unseen testing images, which is however hard to satisfy in real-world applications.

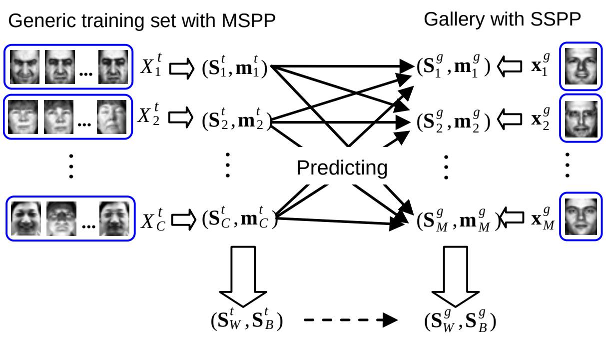

Fig.1. Illustration of the basic ideas of the proposed Adaptive Generic Learning. In the figure, MSPP means multiple samples per person and SSPP denotes single sample per person.

To solve the above problem, we propose the Adaptive Generic Learning method, whose basic idea is illustrated in Fig.1. Instead of directly using the within-class scatter matrix (hereinafter denoted as $\mathbf{S}_W^t$) and the between-class scatter matrix (hereinafter denoted as $\mathbf{S}_B^t$) of the generic training set, we propose to predict those of the gallery (hereinafter respectively denoted as $\mathbf{S}_W^g$ and $\mathbf{S}_B^g$).

With the predicted $\mathbf{S}_W^g$ and $\mathbf{S}_B^g$, FLD can be easily applied to recognize the persons in the gallery. It is worth pointing out that, the basic idea illustrated in Fig.1 is general enough to support various implementations. In the following section, we will present a specific implementation based on a linear-regression-like strategy.

3. Adaptive FLD Learning via Coupled Linear Representation

As mentioned previously, each class in the generic training set corresponds to a within-class scatter matrix and a class mean (i.e., X_i^t corresponds to ($\mathbf{S}_i^t$, $\mathbf{m}_i^t$)). Thus, given a generic training set with C classes, we can have C pairs of this correspondence. Consequently, our goal can be further formulated as: given the above C pairs of correspondence and a gallery sample $\mathbf{x}_k^g$, how to estimate the within-class scatter matrix $\mathbf{S}_k^g$ and the class mean $\mathbf{m}_k^g$ of this person? Evidently, this becomes a typical prediction problem and can be solved by many techniques. In this study, we present a relatively simple but effective method, named by us Coupled Linear Representation (CLR).

In the following, we first present the theoretical principle of CLR. Then, the detailed implementation is given.

3.1. Theoretical Principle

Let $\mathbf{X}_i^t (i = 1, ..., C)$ be the face image of the i-th person in the generic training set, which can be considered as a random vector. The expectation and covariance matrix of $\mathbf{X}_i^t$ are denoted as μ_i^t and Σ_i^t respectively. For a person $\mathbf{X}^g$ in the gallery, we can safely assume that it can be approximated by the linear combination of $\mathbf{X}_i^t (i = 1, ..., C)$:

$$\mathbf{X}^g \approx \sum_{i=1}^{C} w_i \mathbf{X}_i^t , \tag{4}$$

where w_i is the weight of $\mathbf{X}_i^t$. It is not difficult to derive that the expectation of $\mathbf{X}^g$ (denoted as μ^g) can be approximated by the linear combination of $\mu_i^t (i = 1, ..., C)$ with the same weights:

$$\mu^g \approx \sum_{i=1}^{C} w_i \mu_i^t . \tag{5}$$

As for the covariance matrix of $\mathbf{X}^g$ (denoted as Σ^g), the situation is more complex. According to the definition of covariance matrix, Σ^g can be computed as follows:

$$\Sigma^g = \mathrm{E}\left[\left(\mathbf{X}^g - \mu^g\right)\left(\mathbf{X}^g - \mu^g\right)^T\right]. \tag{6}$$

By putting Eq.4 and Eq.5 into the above equation, it can be reformulated as

$$\begin{aligned}\Sigma^g &\approx \mathrm{E}\left[\left(\sum_i w_i \mathbf{X}_i^t - \sum_i w_i \mu_i^t\right)\left(\sum_i w_i \mathbf{X}_i^t - \sum_i w_i \mu_i^t\right)^T\right] \\ &= \mathrm{E}\left[\left(\sum_i w_i \left(\mathbf{X}_i^t - \mu_i^t\right)\right)\left(\sum_i w_i \left(\mathbf{X}_i^t - \mu_i^t\right)\right)^T\right] \\ &= \sum_{i=1} w_i^2 \,\mathrm{E}\left[\left(\mathbf{X}_i^t - \mu_i^t\right)\left(\mathbf{X}_i^t - \mu_i^t\right)^T\right] \\ &\quad + \sum_i \sum_{j \neq i} w_i w_j \,\mathrm{E}\left[\left(\mathbf{X}_i^t - \mu_i^t\right)\left(\mathbf{X}_j^t - \mu_j^t\right)^T\right] \\ &= \sum_i w_i^2 \Sigma_i^t + \sum_i \sum_{j \neq i} w_i w_j \Sigma_{ij}^t \end{aligned} \quad , \qquad (7)$$

where Σ_{ij}^t denotes the cross covariance matrix of $\mathbf{X}_i^t$ and $\mathbf{X}_j^t$. Considering that $\mathbf{X}_i^t$ and $\mathbf{X}_j^t$ represent face images of different persons, they should not have any causality. In other words, $\mathbf{X}_i^t$ and $\mathbf{X}_j^t$ are irrelevant. Thus, the cross covariance matrix Σ_{ij}^t should be a zero matrix. With the above analysis, Eq.7 can be rewritten as:

$$\Sigma^g \approx \sum_i^C w_i^2 \Sigma_i^t \,. \qquad (8)$$

That is to say Σ^g also can be approximated by the linear combination of Σ_i^t $(i = 1, \ldots, C)$.

3.2. Implementation

With the above analysis, this subsection presents in detail how to predict the expectation (μ^g) and the covariance matrix (Σ^g) for each person $\mathbf{X}^g$ in the gallery. Considering that, in real-world applications, the number of the available face images for each person is limited, the mean and within-class scatter matrix for each person are adopted as the estimation of its expectation and covariance matrix respectively.

In order to predict $\mathbf{m}^g$ and $\mathbf{S}^g$ for each person in the gallery, the key issue is to learn the combination coefficients $\{w_i\}$. In this paper, $\mathbf{X}_i^t$ is represented by its mean vector $\mathbf{m}_i^t$ and $\mathbf{X}^g$ is represented by its single face image $\mathbf{x}^g$. Thus, Eq.4 can be rewritten as follows:

$$\mathbf{x}^g \approx \sum_{i=1}^C w_i \mathbf{m}_i^t \,. \qquad (9)$$

Once the coefficients $\{w_i\}$ are learned, the mean and within-class scatter matrix of $\mathbf{X}^g$ can be approximated as:

$$\mathbf{m}^g \approx \sum_{i=1}^C w_i \mathbf{m}_i^t = \mathbf{x}^g \,, \qquad (10)$$

$$\mathbf{S}^g \approx \sum_{i=1}^C w_i^2 \mathbf{S}_i^t \,. \qquad (11)$$

Eq.10 indicates $\mathbf{m}^g$ is estimated directly as $\mathbf{x}^g$. This is rational since $\mathbf{x}^g$ is the only representative sample of $\mathbf{X}^g$.

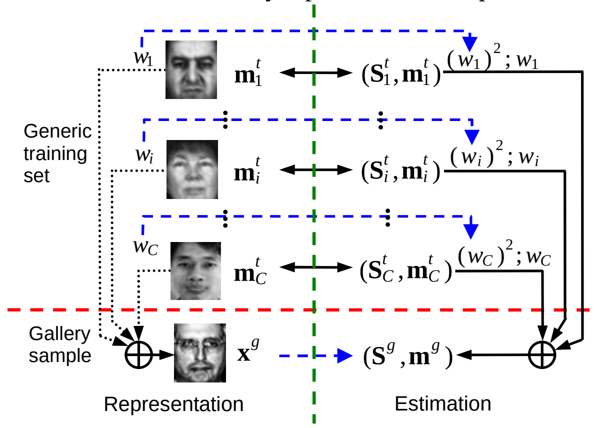

Fig.2. The estimation of $\mathbf{S}_k^g$ and $\mathbf{m}_k^g$ of the k-th person with single sample $\mathbf{x}_k^g$ in the gallery.

Then, how do we learn $\{w_i\}$? Fig.2 illustrates the overall framework of the proposed CLR which consists of two phases: representation and estimation. The former solves the coefficients of the linear combination; while the latter applies the coefficients to estimate the statistics (via Eq.10 and Eq.11). In the following, we describe how to solve the coefficients of the linear combination.

Let us define matrix $\mathbf{A} = [\mathbf{m}_1^t, \mathbf{m}_2^t, \ldots, \mathbf{m}_C^t] \in \Re^{d \times C}$ and column vector $\mathbf{w}_k = [w_1^k, w_2^k, \ldots, w_C^k] \in \Re^C$, where d is the dimensionality of the samples and C is the number of persons in the generic training set. Thus, for the k-th person in the gallery, Eq.9 can be rewritten as

$$\mathbf{x}_k^g = \mathbf{A}\mathbf{w}_k^T \,. \qquad (12)$$

$\mathbf{w}_k$ can be easily solved by using pseudo-inverse:

$$\mathbf{w}_k = \mathbf{A}^\dagger \mathbf{x}_k^g = \mathbf{A}^T (\mathbf{A}\mathbf{A}^T)^{-1} \mathbf{x}_k^g \,, \qquad (13)$$

where $\mathbf{A}^\dagger$ is the pseudo-inverse matrix of $\mathbf{A}$. With the learned coefficients, in the estimation phase, the $\mathbf{S}_k^g$ and $\mathbf{m}_k^g$ can be approximated as a linear combination of all the $\mathbf{S}_i^t$ and $\mathbf{m}_i^t$ respectively (by Eq.10 and Eq.11).

After the estimation of the $\mathbf{S}_k^g$ and $\mathbf{m}_k^g$ for each person in the gallery, the traditional FLD can be applied to obtain the low dimensional subspace, which is called by us the Adapted FLD subspace since it is adapted to discriminate the persons in the gallery.

For clarity, the complete procedure of the proposed adaptive FLD learning via coupled linear representation is summarized in *Alg*.1. After Adapted FLD subspace is obtained, both the gallery samples and the testing samples are projected into this subspace and then the nearest neighbor classifier is used for classification.

***Alg.* 1. Adaptive FLD Learning via Coupled Linear Representation**

- Given a generic training set $\mathbf{T}=[\mathbf{x}_{1,1}^{t},\mathbf{x}_{1,2}^{t},...,\mathbf{x}_{1,N_1}^{t},\mathbf{x}_{2,1}^{t}...,\ \mathbf{x}_{C,N_C}^{t}]\in R^{d\times N}$ and a galley $\mathbf{G}=[\mathbf{x}_{1}^{t},...,\mathbf{x}_{M}^{t}]\in R^{d\times M}$.
- Compute the $\mathbf{S}_i^t$ and $\mathbf{m}_i^t$ for each person in the generic training set.
- For each $\mathbf{x}_k^g$ in the gallery:
  1. solve coefficients vector $\mathbf{w}_k$ according Eq.13.
  2. estimate the $\mathbf{S}_k^g$ and $\mathbf{m}_k^g$ according to Eq.10 and Eq.11 respectively.
- Compute the $\mathbf{S}_W^g$ and $\mathbf{S}_B^g$ by using the estimated $\mathbf{S}_k^g$ and $\mathbf{m}_k^g$, according to Eq.1 and Eq.2.

4. Experiments

In this section, we adopt some publicly available large-scale face databases (XM2VTS [24], CAS-PEL [25] and FERET [26]) and a private passport face database to evaluate the proposed method and compared it with some popular methods dealing with the SSPP problem.

To simulate the SSPP problem, two testing sets are used in our experiments, i.e., the FERET face database and our private passport face database. Among them, the FERET face database consists of a gallery and four probe (i.e., testing) sets: *fafb*, *fafc*, *dup1* and *dup2*. The gallery includes 1196 persons with only a single image for each person. The images in *fafb* set are with expression variation, *fafc* set contains images with lighting variation, and the images in *dup1* and *dup2* sets were acquired some days later. Our passport face database contains 4658 passport face images of 2000 persons. To simulate an SSPP problem, one image is randomly selected for each person and to form the gallery, while other images are used for testing. Thus, the gallery consists of 2000 persons with a single image per person, and the probe set includes 2658 images. Due to privacy problem, the images are not exampled in the paper. However, 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 test forms a quite challenging scenario.

As for the generic training set, in real-word applications, it should contain as many persons and images as possible in order to include as much representative discriminative information as possible. However, considering the high computational cost of using a huge generic training set, we only build a generic training set of moderate scale. As mentioned in the introduction, the motivation of our method is that the discrimination information embedded in the generic training set should not be directly used to identify persons in the gallery, because there may exists large difference between these two image sets. Thus, in order to validate our method, only two public face databases (i.e., XM2VTS and CAS-PEAL) which are very different from FERET and our Passport databases are selected to build the generic training set. The XM2VTS face database contains 3440 images of 295 persons taken at one month interval with the slight head pose variations and illumination condition changes. The CAS-PEAL face database contains images with the variations due to pose, expression, lighting and so on. In our experiments, all the images in the XM2VTS face database and all the images in the CAS-PEAL training set (1200 frontal face images of 300 persons) are put together to form the generic training set. Consequently, the generic training set contains 4640 face images of 595 persons.

Fig.3. Examples of the normalized face image in the XM2VTS, CAS-PEAL and FERET face databases.

In the experiments, all the face images are aligned and normalized to the size of 40 by 50 pixels according to the manually located eye centers, with the histogram equalization as the illumination preprocessing. Some examples of the normalized face image are shown in Fig.3.

4.1. Evaluation of the Adapted FLD

In order to reduce the computational burden, before FLD adaptation, the dimensionality of the face images is reduced to 500 by using PCA (trained on the generic training set and about 96% energy reserved). The number of persons in both the generic training set and the gallery is larger than 500, thus the maximum dimension of the FLD subspace in our experiments is 500. For clarity, only the peak recognition rate of the optimal dimension is reported. It is worth pointing out that, in our experiments, after projecting images onto the FLD subspace, the similarity between two samples is measured by the normalized cross-correlation and the nearest neighbor classifier is used for classification.

Besides the Generic FLD (with $\mathbf{S}_B^t$ and $\mathbf{S}_W^t$) and the Adapted FLD (with $\mathbf{S}_B^g$ and $\mathbf{S}_W^g$), there are still two derived methods: one is with $\mathbf{S}_B^t$ and $\mathbf{S}_W^g$, the other is with $\mathbf{S}_B^g$ and $\mathbf{S}_W^t$. Table.1 gives the rank-1 recognition rate of these methods on both the FERET and the Passport face database.

Table 1. Rank-1 recognition rates of different combinations of $\mathbf{S}_B$ and $\mathbf{S}_W$ on both the FERET and passport face database. Bold font denotes the recognition rates of the best combination.

Different choices of S_W and S_B	FERET (%)				Passport (%)
	fafb	*fafc*	*dup1*	*dup2*	
$(\mathbf{S}_B^t;\mathbf{S}_W^t)$	84.1	67.5	47.5	23.5	35.8
$(\mathbf{S}_B^g;\mathbf{S}_W^t)$	76.5	63.4	44.0	20.5	31.0
$(\mathbf{S}_B^t;\mathbf{S}_W^g)$	75.1	57.2	43.4	23.5	32.8
$(\mathbf{S}_B^g;\mathbf{S}_W^g)$	**88.5**	**71.6**	**53.3**	**35.0**	**53.5**

It can be concluded from Table.1 that the Adapted FLD with $\mathbf{S}_B^g$ and $\mathbf{S}_W^g$ performs significantly better than other methods. Particularly, its accuracy is greatly improved compared with the Generic FLD with $\mathbf{S}_B^t$ and $\mathbf{S}_W^t$ computed directly on the generic training set. Especially on the Passport face database, the improvement of the recognition rate is more than 15 percents (from 35.8% to 53.5%).

4.2. Comparison with Other Methods

In this section, we compare the proposed method with other typical methods which can be used to deal with the SSPP problem, including PCA [3], $(PC)^2A$ [8], Local Binary Pattern (LBP) [27], the method in [14] (hereinafter denoted as Block FLD), as well as the Generic FLD.

We implement the above comparison methods ourselves and their setups are described as follows. For $(PC)^2A$ [8], there is only one free parameter α , the weight of projection-combined version of the face image. As reported in [8], the experiments demonstrate that the performance of $(PC)^2A$ is not sensitive to α when it is between 0.1 and 0.5. Thus, in our experiments, the weight is fixed to 0.3. As to the Block FLD [14], the critical parameter is the size of the image blocks. So, we try four different sizes (10×10, 10×25, 20×10 and 20×25) and report the results of the best one (10×25). Similarly, in LBP, the partition of the face image has great effect on its performance. In our experiments, we also try four different numbers of image blocks (16, 32, 40, 72) and report the best results (i.e., those of 72 blocks in our study). It should be pointed out that, all these methods except LBP are trained on the gallery samples.

In addition, in recent years, FLD is often combined with Gabor features to further improve the accuracy of the face recognition systems, as done in [28, 29, 30]. Therefore, we also validate the method combining the Adapted FLD with Gabor features (only the magnitude part). In our implementation, to reduce the dimensionality of Gabor features for FLD, we partition the face image into 4 blocks (each with the size of 20×25 pixels) and then train 4 Gabor-FLD classifiers respectively, which are finally combined by weighted sum rule. Hereinafter the method is denoted as Adapted Gabor-FLD. Obviously, the Generic FLD can also be enhanced by this strategy, which is denoted hereinafter as Generic Gabor-FLD hereinafter.

Table 2. Rank-1 recognition rates of our methods and other compared methods on FERET and Passport database. Bold font denotes the recognition rates of our methods.

Methods	FERET (%)				Passport (%)
	fafb	*fafc*	*dup1*	*dup2*	
PCA	87.4	10.3	38.9	12.8	20.6
$(PC)^2A$	87.9	12.4	38.6	13.2	20.4
LBP	97.5	49.0	59.0	37.2	45.1
Block FLD	73.3	50.0	41.3	33.8	42.3
Generic FLD	84.1	67.5	47.5	23.5	35.8
Adapted FLD	**88.5**	**71.6**	**53.3**	**35.0**	**53.5**
Generic Gabor-FLD	96.4	89.2	67.3	47.4	53.8
Adapted Gabor-FLD	**97.9**	**92.3**	**70.8**	**54.7**	**63.5**

Note: since we do not exploit the FERET training set, the results on FERET dataset should not be compared with the results reported in previous literature on this database. The testing scenario in our paper is much challenging.

Table.2 gives the comparison results of these methods on both the FERET and Passport database. Note that, in the table, for all the subspace-based methods, only the peak recognition rates at the optimal subspace dimension are reported. From Table.2, we can reach several observations: 1) the PCA-based methods are worst; 2) Generic FLD performs worse than LBP, whereas our Adapted FLD outperforms LBP significantly on the Passport database and performs comparably to LBP on the FERET probe sets; 3) by using Gabor filters, the performances of FLD-based methods are greatly improved; 4) by using Gabor features, the proposed Adapted Gabor-FLD method outperforms other methods significantly on all the testing sets.

It is worth pointing out that, in recent years, some methods reported impressive results (e.g., in [31]) on the FERET database which is better than those in this paper.

However, it must be noted that those results are not comparable with ours directly, since they used an evaluation protocol quite different from ours. To learn the recognition model, those methods generally made use of the FERET training set, which has similar sample distribution with the testing dataset and thus facilitates the testing. However, our method does not exploit the FERET training set at all, thus forms a more challenging testing scenario.

5. Discussion

The proposed method essentially relates in some sense to the methods based on virtual sample generation (e.g. [12]). Specifically, both methods utilize some prior information to generate novel discriminatory information more specific to the persons under consideration. However, the differences between these two methods are also evident. Firstly, our method does not explicitly generate any virtual sample but directly estimate the within-class scatter matrix of each person. Secondly, our method employs a learning-based method to utilize the prior information embedded in the generic training set, whereas in the methods of virtual sample generation, the prior information is often utilized in an explicit manner (e.g. using 3D face model or illumination models).

Additionally, the traditional FLD often suffers from the small sample size (SSS) problem, i.e., the amount of training samples are not sufficient to guarantee the non-singularity of the within-class scatter matrix. Obviously, the SSPP problem is actually an extreme case of the SSS problem. Consequently, the proposed method can also solve the SSS problem. For example, as in this paper, we can directly replace the “real” but singular within-class scatter matrix (computed on the gallery) with the estimated non-singular within-class scatter matrix (adapted from the generic training set). Another alternative is to combine the estimated within-class scatter matrix and the “real” one to make the latter non-singular.

Besides generic FLD, the basic idea of adaptive generic learning can also be used to improve other generic learning methods. For example, in Bayesian face recognition [20], the intra-personal variations and inter-personal variations are also modeled on a generic training set. So, it is rational to adapt them to a specific gallery. For this purpose, we might need to respectively model the intra-personal variations of each person in the generic training set, from which the intra-personal variations of persons in the gallery can be estimated in the similar way of this paper.

6. Conclusion and Future Work

Real-world face recognition systems often suffer from the single sample per person problem, which makes many supervised learning methods fail to extract the discriminatory information. To deal with this problem, in this paper, we propose to adapt the within-class and between-class scatter matrices computed from a generic training set to the persons to be identified by coupled linear representation method. Experimental results on the FERET and a passport face databases demonstrate that the proposed method outperforms other relevant methods.

Though promising results have been achieved, as a preliminary study, the proposed method still has large space to extend. For instance, when solving the combination coefficients, the class mean is adopted to represent each class. However, in most real applications, a class can hardly be well modeled by its mean. Thus, there should be better methods to model the classes and solve the combination coefficients. Additionally, as discussed in Section 5, the method can be smoothly applied to solve SSS problem or adapted to other non-FLD face recognition models.

7. Acknowledgement

This paper is partially supported by Natural Science Foundation of China under contracts No.60772071, No.60833013, and No. U0835005; National Basic Research Program of China (973 Program) under contract 2009CB320902; Grand Program of International S&T Cooperation of Zhejiang Province S&T Department under contract No. 2008C14063. Portions of the research in this paper use the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office.

8. References

[1] W. Zhao, R. Chellappa, P. Phillips and A. Rosenfeld, “Face recognition: A literature survey”, *ACM Computing Surveys*, vol. 35, no. 4, pp. 399-458, 2003.

[2] R. Brunelli and T. Poggio, “Face recognition: feature versus templates”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042-1052, 1993.

[3] M. Turk and A. Pentland, “Eigenfaces for recognition”, *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.

[4] P. Belhumeur, J. Hespanha and D. Kriegman, “Eigenfaces vs Fisherfaces: recognition using class specific linear projection”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 7, pp. 711-720, 1997.

[5] A. Jain and B. Chandrasekaran, “Dimensionality and sample size considerations in pattern recognition practice”, *Handbook of Statistics*, P.R. Krishnaiah, L.N. Kanal (Eds.), vol. 2, 1982, pp. 835–855.

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

[7] J. Yang, D. Zhang, A. Frangi and J. 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.
[8] J. Wu and Z. Zhou, "Face recognition with one training image per person", *Pattern Recognition Letters*, vol. 23, no. 14, pp. 1711-1719, 2002.
[9] B. Schölkopf, A. Smola and K. Müller, "Nonlinear component analysis as a kernel eigenvalue problem", *Neural Computation*, vol. 10, no. 5, pp. 1299-1319, 1998.
[10] X. Tan, S. Chen, Z. Zhou and F. Zhang, "Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft kNN ensemble", *IEEE Transactions on Neural Networks*, vol.16, no.4, pp. 875-886, 2005.
[11] A.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, no. 6, pp. 748-763, 2002.
[12] S. Shan, B. Cao, W. Gao, D. Zhao, "Extended fisherface for face recognition from a single example image per person", *IEEE International Symposium on Circuits and Systems*, vol. 2, pp. 81-84, 2002.
[13] J. Huang, P. Yuen, W. Chen and J. Lai, "Component-based LDA method for face recognition with one training sample", *Proceeding of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 120-126, 2003.
[14] S. Chen, J. Liu and Z. Zhou, "Making FLDA applicable to face recognition with one sample per person", *Pattern Recognition*, vol. 37, no. 7, pp. 1553-1555, 2004.
[15] Q. Gao, L. Zhang and D. Zhang, "Face recognition using FLDA with single training image per-person", *Applied Mathematics and Computation*, vol. 205, pp. 726-734, 2008.
[16] D. Beymer and T. Poggio, "Face recognition from one example view", *Proceedings of the IEEE International Conference on Computer Vision*, pp. 500-507, 1995.
[17] T. Vetter, "Synthesis of novel views from a single face image", *International Journal of Computer Vision*, vol. 28, no. 2, pp. 102-116, 1998.
[18] T. Sim and T. Kanade, "Combining models and exemplars for face recognition: an illuminating example", *Proceedings of the CVPR Workshop on Models versus Exemplars in Computer Vision*, 2001.
[19] J. Wang, K. Plataniotis, J. Lu and A. Venetsanopoulos, "On solving the face recognition problem with one training sample per subject", *Pattern Recognition*, vol. 39, pp. 1746-1762, 2006.
[20] B. Moghaddam, T. Jebara and A. Pentland, "Bayesian face recognition", *Pattern Recognition*, vol. 33, no. 11, pp. 1771-1782, 2000.
[21] T. Kim and J. Kittler, "Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 318-327, 2005.
[22] A. Majumdar and R. Ward, "Pseudo-fisherface method for single image per person face recognition", *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 989-992, 2008.
[23] K. Fukunaga, "Introduction to statistical pattern recognition", pp. 448-449, Academic Press, Boston, 2nd edition, 1990.
[24] K. Messer, J. Matas, J. Kittler, J. Luettin and G. Maitre, "XM2VTSDB: The extended M2VTS database", *Proceedings of the Second International Conference on Audio and Video-based Biometric Person Authentication*, pp.72-77, 1999.
[25] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang and D. Zhao, "The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations", *IEEE Transactions on System Man, and Cybernetics (Part A)* , vol. 38, no. 1, pp. 149-161, 2008.
[26] P. Phillips, H. Wechsler, J. Huang and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms", *Image and Vision Computing*, vol. 16, no. 5, pp. 295-306, 1998.
[27] A. Timo, H. Abdenour and P. Matti, "Face recognition with local binary patterns", *Proceedings of the European Conference on Computer Vision*, pp. 469-481, 2004.
[28] 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, no. 4, pp. 467-476, 2002.
[29] S. Li, W. Liu, D. Lin and X. Tang, "Nonparametric subspace analysis for face recognition", *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 961-966, 2005.
[30] L. Shen, L. Bai and M. Fairhurst, "Gabor wavelets and general discriminant analysis for face identification and verification", *Image and Vision Computing*, vol. 25, no. 5, pp. 553-563, 2007.
[31] X. Tan and B. Triggs, "Fusing Gabor and LBP feature sets for kernel-based face recognition", *Proceeding of the IEEE International Workshop Analysis and Modeling of Faces and Gestures*, pp.235-249, 2007.