A Discriminative Model for Age Invariant Face Recognition

Zhifeng Li[1], *Member, IEEE*, Unsang Park[2], *Member, IEEE*, and Anil K. Jain[2,3], *Fellow, IEEE*

[1]Shenzhen Institutes of Advanced Technology, Chinese Academy of Science, China
[2]Dept. of Computer Science and Engineering, Michigan State University, E. Lansing, MI 48823, USA
[3]Dept. of Brain and Cognitive Engineering, Korea University, Seoul 136-713, Korea

***Abstract*—Aging variation poses a serious problem to automatic face recognition systems. Most of the face recognition studies that have addressed the aging problem are focused on age estimation or aging simulation. Designing an appropriate feature representation and an effective matching framework for age invariant face recognition remains an open problem. In this paper, we propose a discriminative model to address face matching in the presence of age variation. In this framework, we first represent each face by designing a densely sampled local feature description scheme, in which scale invariant feature transform (SIFT) and multi-scale local binary patterns (MLBP) serve as the local descriptors. By densely sampling the two kinds of local descriptors from the entire facial image, sufficient discriminatory information, including the distribution of the edge direction in the face image (that is expected to be age invariant) can be extracted for further analysis. Since both SIFT-based local features and MLBP-based local features span a high-dimensional feature space, to avoid the overfitting problem, we develop an algorithm, called multi-feature discriminant analysis (MFDA) to process these two local feature spaces in a unified framework. The MFDA is an extension and improvement of the LDA using multiple features combined with two different random sampling methods in feature and sample space. By random sampling the training set as well as the feature space, multiple LDA-based classifiers are constructed and then combined to generate a robust decision via a fusion rule. Experimental results show that our approach outperforms a state-of-the-art commercial face recognition engine on two public domain face aging data sets: MORPH and FG-NET. We also compare the performance of the proposed discriminative model with a generative aging model. A fusion of discriminative and generative models further improves the face matching accuracy in the presence of aging.**

Manuscript received October 9, xxxx. This research was partially supported by The work described in this paper was partially supported through a grant from the Center for Identification Technology Research (CITeR) at West Virginia University and WCU (World Class University) program funded by the Ministry of Education, Science and Technology through the National Research Foundation of Korea (R31-10008) to Korea University.

Zhifeng Li is with the Shenzhen Institutes of Advanced Technology, Chinese Academy of Science, China.

U. Park and A. K. Jain are with the Department of Computer Science and Engineering Department, Michigan State University, East Lansing, MI 48824 USA (email: parkunsa@cse.msu.edu; jain@cse.msu.edu). A. K. Jain is also with the Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea.

***Index Terms*— Face recognition, age invariance, local feature representation, multi-feature discriminant analysis, discriminative model, generative model.**

I. Introduction

AUTOMATIC face recognition is an important yet challenging problem. This challenge can be attributed to (i) large intra-subject variations and (ii) large inter-user similarity. Fig. 1 shows some of the main intra-subject variations (pose, illumination, expression, and aging) commonly encountered in face recognition. Among these variations, aging variation is now beginning to receive increasing attention in the face recognition community. Designing an age-invariant face recognition method is necessary in many applications, particularly those that require checking whether the same person has been issued multiple government documents (e.g., passports and driver license) that include facial images [1], [2].

Published approaches to age invariant face recognition are limited. Most of the available algorithms dealing with facial aging problem are focused on age estimation [3]-[13] and aging simulation [14]-[18]. One of the successful approaches to age invariant face recognition is to build a 2D or 3D generative model for face aging [4], [14], [18]. The aging model can be used to compensate for the aging process in face matching or age estimation. These methods first transform the face images being compared to the same age as the gallery image using a trained aging model to compensate for the age effect (see Fig. 2). While the model based methods have been shown to be effective in age invariant face recognition, they have some limitations. First, construction of face models is difficult and sometimes they do not represent the aging process very well, especially when the training sample size is limited. Further, the facial aging process is very complex and, consequently, in order to construct the aging model, strong parametric assumptions are needed, which are often unrealistic in real-world face recognition scenarios. Second, for constructing the aging model, additional information in the form of the true ages of the training faces and the locations of landmark points on each face image are needed. A further constraint on the training set is that the images should be captured under controlled conditions (e.g., frontal pose,

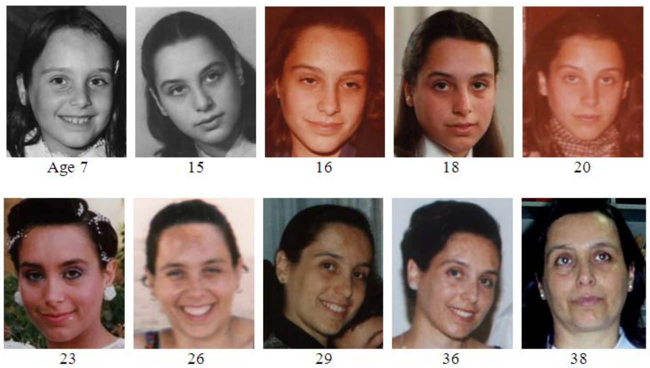


Figure 1. Example images showing intra-subject variations (e.g., pose, illumination, expression, and aging) for one of the subjects in the FG-NET database [29].

normal illumination, neutral expression). Unfortunately, such constraints are not easy to satisfy in practice, especially in scenarios where the face images being compared are subject to significant changes not only in aging, but also in other possible variations such as pose, illumination, and expression. In order to overcome these problems, approaches based on discriminative models have been proposed for the aging problem. Some of the representative works of discriminative models is [28], [38] which used gradient orientation pyramid (GOP) for feature representation, combined with support vector machine for verifying faces across age progression. Guo et al. [39] investigated the relationship between recognition accuracy and age gap, and reported the performance of two well known algorithms (PCA and EBGM) on a large data set. They also showed some improvement in matching by indexing the gallery based on demographic information (gender, race, height, and weight).

In this paper, we address the age invariant face recognition problem by developing a new discriminative approach. We propose a learning algorithm that has the capability to not only address the aging variations, but also handle the other intra-user variations (e.g., pose, illumination, expression). Our discriminative model differs from the models in [28], [38] in both feature representation and classification, as shown in section II. Although global appearance based features have been widely used for face representation, it is now generally agreed [31], [48] that local image descriptors are more effective for face representation. Compared to the global appearance features, the local features inherently possess spatial locality and orientation selectivity. These properties allow the local feature representations to be robust to aging, illumination, and expression variations. Considering that the entire face image (which has high structural complexity) is difficult to be characterized by a single image descriptor, we use a patch-based local feature representation scheme (also called densely sampled local feature description) in this paper. We first divide the input face image into a set of overlapping patches with each patch represented by an appropriate image descriptor. In order to ensure local consistency, we use 50% overlap between the adjacent patches in our approach. We use both the Scale Invariant Feature Transform (SIFT) [20] and

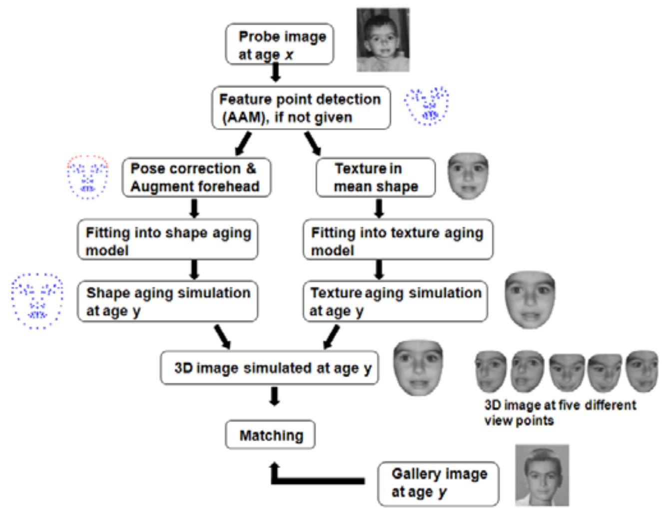


Figure 2. Schematic of the aging simulation process from age *x* to age *y* [18].

Multi-scale Local Binary Pattern (MLBP) [23] since both of these descriptors have been shown to be very successful in image representation [31]. For matching the set of large number of SIFT and MLBP local features effectively and efficiently, we develop a multi-feature discriminant analysis (MFDA) algorithm for dimensionality reduction. In MFDA, local descriptors are combined to construct a robust decision rule by a random subspace fusion model. Extensive experiments are conducted to validate the effectiveness of the proposed algorithm on two public domain face aging data sets: MORPH and FG-NET.

Our approach to match two face images of the same person acquired at different ages differs significantly from the previously published approaches. A major difference lies in the fact that our approach is a discriminative one while most of the other approaches construct generative models. A generative model considers the formation of the target subject's face to be controlled by a set of hidden parameters. Different faces of the same subject at different ages are generated under a similar structure with varying parameters. Accordingly, these parameters are used to characterize the target subject's face. Fig. 2 shows an example of the aging simulation process using a generative aging model proposed in [18]. However, the aging process which needs to be modeled is highly complex and there are multiple factors that affect the aging which are subject-specific and depend on the specific age range. This motivates our exploration of the discriminative model for age invariant face recognition. Our discriminative model also differs significantly from other existing discriminative models [28], [38] for the aging problem. The methods in [28], [38] were proposed for the face verification task, which is a binary recognition problem, while our approach is proposed for the face recognition task which is a multi-class recognition problem. Further, the methods in [28], [38] proposed the use of gradient orientation pyramid (GOP) for feature representation, followed by the support vector machine classifier for verification. Our approach, on the other hand, proposes densely sampled local feature description for feature representation, and further develops the MFDA for classification.

II. Discriminative Model

The proposed discriminative model consists of two components: densely sampled local feature description and multi-feature discriminant analysis (MFDA). We will describe each component of the framework in the following subsections.

A. Densely Sampled Local Feature Description

Compared to the global appearance features, local features have been shown to be more effective in representing face images at diverse scales and orientations and robust to geometric distortions and illumination variations [31]. Hence we adopt the local image descriptor-based technique for face representation.

We first divide the whole face images into a set of overlapping patches and then apply the selected local image descriptors to each patch. The extracted features from these patches are concatenated together to form a feature vector with large dimensionality for further analysis. Given a face image of size $H \times W$, it is divided into a set of $s \times s$ overlapping patches that overlap by r pixels. The number of horizontal (M) and vertical (N) patches obtained are

$$N = (W - s)/r + 1 \tag{1}$$

$$M = (H - s)/r + 1 \tag{2}$$

For each of the $M \times N$ patches, we compute a d-dimensional feature vector. These image feature vectors are concatenated into a single $M \times N \times d$-dimensional feature vector for a given face image.

Among the available local feature descriptors, SIFT [20] and LBP [23] have been shown to be the most effective for object recognition [31]. Based on their reported successes in face recognition literature [40]-[43], we choose both of them as feature descriptors in developing age invariant face recognition algorithm. The SIFT feature descriptor quantizes both the spatial location and orientation of image gradient within an $s \times s$ sized image patch, and computes a histogram in which each bin corresponds to a combination of specific spatial location and gradient orientation. The accumulation of the histogram bins is weighted by the gradient magnitude and a Gaussian decay function [20]. We use the extended LBP, MLBP [23], to describe the face at multiple scales, by computing the LBP descriptors computed at four different radii {1, 3, 5, 7}. An illustration of the local feature representation scheme is given in Fig. 3.

Although both SIFT and LBP have been successfully used in face recognition before, our adoption is different and novel for the aging problem. Traditionally, SIFT feature representation consists of two main parts: key point extraction, and feature descriptors. But in our study, we densely sample the SIFT feature descriptors from the entire facial image instead of only at a relatively small number of extracted key points. In other words, we do not perform key point extraction, but place a regular grid on the face. Such a strategy allows the definition of age invariant discriminatory information in the form of distribution of the edge direction in the face. The same is true for the use of MLBP in this paper. These extracted local features are well suited for age-invariant face recognition as supported by our experimental results.

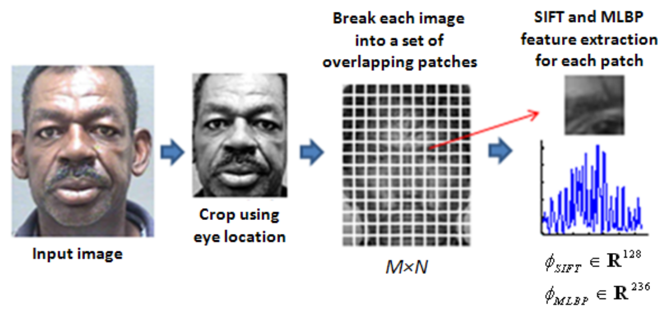
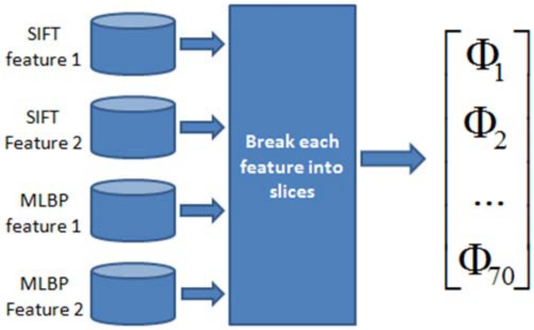


Figure 3. Illustration of local feature representation of a face image.

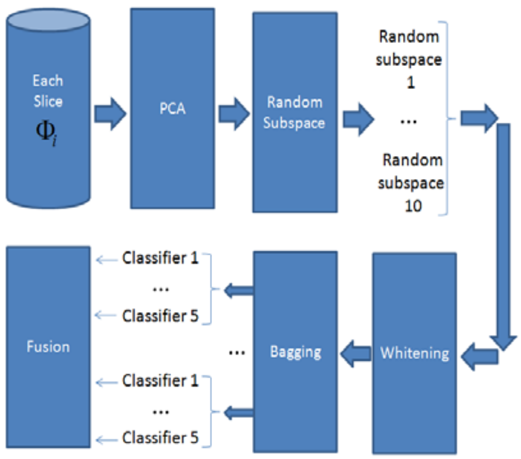
In order to extract the local features (SIFT and MLBP), each face is first normalized to 150×200 pixels, and then divided into either 88 overlapping patches (for patch size of 32×32) or 408 overlapping patches (for patch size of 16×16) Each patch is represented by a 128-dimensional SIFT feature vector or a 236-dimensional MLBP feature vector. Thus, the resulting feature dimensionality is very high and it is desirable to reduce the dimensionality using discriminant analysis. A straightforward approach would be to apply the well-known LDA (Linear Discriminant Analysis) to the SIFT and MLBP features separately and then fuse the outputs of the two classifiers, one based on SIFT and the other based on MLBP. However, this approach has some limitations. First, we would be fusing only two classifiers. Studies on multi-classifier system design [33] have shown that the choice of the number of classifiers is critical to the overall classifier stability and performance. Second, a single classifier constructed on a limited training data set is usually biased and unstable, especially when the original feature dimensionality is very high. In order to overcome these problems, we develop the multi-feature discriminant analysis (MFDA) framework to take advantage of the two different representations in a unified computational framework. The MFDA is an extension and improvement of the LDA using multiple features combined with two different random sampling methods in feature and sample spaces, as explained in the following sections.

B. Multi-Feature Discriminant Analysis (MFDA)

The LDA [25] is one of the most popular discriminant analysis scheme for face recognition. This can be evidenced by a variety of implement of LDA-based methods in face recognition literatures [24], [25], [26], [37], [45], [46], [47]. So we first briefly review the basic idea of the LDA. The LDA uses the within-class scatter matrix and the between-class scatter matrix to define a criterion function to measure the class separability. The within-class and between-class scatter matrices are defined as



(a) Break the local features into slices. There are total 70 slices for each face image.



(b) Train 50 different classifiers for each slice for fusion.

Figure 4. Block diagram of the multi-feature discriminant analysis (MFDA).

$$S_w = \sum_{i=1}^{c} \sum_{X_j \in C_i} (X_j - \mu_i)(X_j - \mu_i)^T \tag{3}$$

$$S_b = \sum_{i=1}^{c} (\mu - \mu_i)(\mu - \mu_i)^T \tag{4}$$

where μ_i denotes the mean of the class C_i, μ denotes the overall mean and c denotes the number of classes.

The objective of LDA is to obtain the optimal projection W_{opt}, which maximizes the ratio of the determinant of the between-class matrix to that of the within-class matrix, defined as:

Table I. Summary of local features used in our framework.

	SIFT feature 1	SIFT feature 2	MLBP feature 1	MLBP feature 2
Patch size	16 x 16	32 x 32	16 x 16	32 x 32
Number of patches	408	88	408	88
Number of slices	24	11	24	11
Number of total slices for each sample	70			

Figure 5. Example face images of the same subject at different ages (ranging from 42 to 46) in the MORPH database [19].

$$W_{opt} = \arg\max_{W} \frac{\left\| W^T S_b W \right\|}{\left\| W^T S_w W \right\|} \tag{5}$$

Mathematically, it is equivalent to computing the leading eigenvectors of $S_w^{-1} S_b$ [27].

LDA has been shown to be very successful in face recognition. However, if we directly use the LDA for discriminant analysis for age invariant face recognition, we will encounter the following problems. First, the high dimensionality of the input feature vector space in conjunction with relatively small size of the aging training set would drastically reduce the accuracy and stability of S_w. Furthermore, with only the class means taken into account for computing S_b, LDA fails to capture the boundary structure of the classes effectively from small training data that is typical of available aging databases.

A possible way to overcome the above problems is to use a random sampling technique to improve the performance of LDA. There are two popular random sampling methods: random subspace and bagging. In the random subspace method [21], multiple classifiers are constructed by randomly sampling the feature space. The decisions made by these individual classifiers are then combined to generate the final decision to strive for improved classification performance. In the bagging method [22], multiple training subsets are generated by randomly sampling the training set. A classifier is then constructed from each training subset, and the results of these multiple classifiers are integrated. In order to better address the "curse of dimensionality" problem [44], we utilize both random subspace and bagging schemes. First, in order to reduce the feature dimensionality, we apply the random subspace technique to sample the feature space to generate multiple subspaces with lower dimensionalities. Second, in order to utilize the classification boundary information, we select specific sample pairs from different classes to better estimate the between-class scatter matrix and the discriminant subspace. It has been shown that the inter-class sample pairs (sample pairs from different classes) near the classification boundary contain more discriminatory information and thus play more significant roles in the learning of the discriminative subspace [26], [27]. This inspires us to select a small set of inter-class sample pairs with smaller distances for constructing the between-class scatter matrix. To this end, we apply the bagging technique to randomly sample the inter-class sample pairs with small distances to generate multiple inter-class sample pair subsets for constructing multiple between-class scatter matrices. By combing the random subspace and bagging techniques, a random sampling based classification framework, called MFDA is developed in this paper. The framework is shown in Fig. 4, and the parameter values used in the feature representation are summarized in Table I.

Below, we summarize the entire procedure for MFDA.

Training stage:

1. For each SIFT or MLBP feature vector, break it into slices with feature from the patches of the same row in the image as one slice. As shown in Table I, the total number of slices is 70. For each slice, perform PCA on the training set and then keep all the eigenvectors with non-zero eigenvalues as candidates to construct 10 random PCA subspaces $\{S^i\}_{i=1}^{10}$, each spanned by 300 PCA dimensions. The first 200 dimensions are fixed according to the first 200 eigenvectors with the largest eigenvalues, which preserve most of the facial variation. The other 100 dimensions are randomly selected from the remaining eigenvectors, which are used to capture the local facial details.
2. In each reduced PCA subspace, estimate the within-class scatter matrix S_w and whiten it, in an attempt to remove the intra-personal variations. This is achieved by a whitening transformation matrix *T*, which is computed as follows:

$$T^T S_w T = I, \quad T = \Phi \Lambda^{-1/2} \tag{6}$$

where Φ is the eigenvector matrix of S_w, Λ is the eigenvalue matrix of S_w and *I* is the identity matrix.

3. In each projected subspace above (after PCA and whitening), we construct 5 different between-class scatter matrices $\{S_b^j\}_{j=1}^{5}$ using the bagging technique. Each between-class scatter matrix, S_b^j, is calculated from the 2,000 inter-class pairs which are randomly selected from 10,000 inter-class pairs with the smallest distances among all the inter-class pairs,

$$S_b^j = \sum_{k=1}^{2000} \frac{(x_{k1} - x_{k2})(x_{k1} - x_{k2})^T}{\left\| x_{k1} - x_{k2} \right\|^2} \tag{7}$$

where (x_{k1}, x_{k2}) is the *k*-th selected inter-class pair from the subset of the 10,000 inter-class pairs with the smallest distances. Based on each between-class scatter matrix, we construct an LDA-based subspace classifier. In each projected subspace (10 in total for each slice), we generate 5 classifiers based on each $\{S_b^j\}_{j=1}^{5}$. In this way, we generate 50 different classifiers for each slice.

Testing stage:

1. For each testing sample (which is represented by four kinds of local features), obtain the 70 slices using the similar way as in the training stage.
2. Use the trained subspace classifiers to determine the classification outputs of these slices.
3. The outputs are first normalized by using the min-max score normalization scheme [32] and then combined by a simple score-sum based fusion rule to make the final decision.

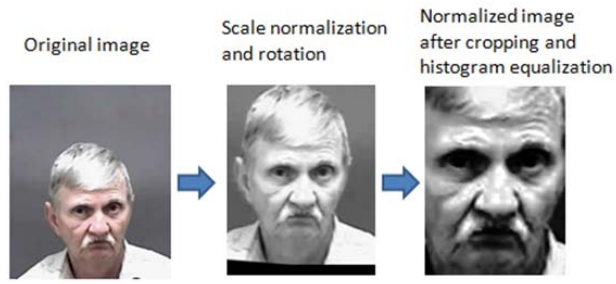


Figure 6. Schematic of the normalization process.

C. Discussion on MFDA

- The MFDA is proposed specifically for handling multiple feature sets with large dimensionality and with different scales and measurements. In our study, there are two kinds of local features (SIFT and MLBP), each with two different feature sets corresponding to two different patch sizes. In order to effectively handle these large numbers of features for enhanced performance, we need to overcome two problems: 1) different incompatibility in scale and measurement and 2) overfitting problem. The MFDA algorithm is not developed only to solve the traditional dimensionality reduction problem. In MFDA, different kinds of features are broken into slices and then scaled by PCA normalization, and the overfitting problem is solved by the random sampling. Our experimental results also support the effectiveness of the MFDA over LDA.
- The use of the bagging technique in the MFDA differs from the traditional random sampling based models [22], [37]. Instead of using the Bagging to randomly sample data within each class or randomly select a subset of classes, the MFDA uses bagging to choose a subset of specific inter-class sample pairs that are close to the classification boundary (in the projected subspace rather than the original feature space) for the construction of the between-class scatter matrix. Therefore, it is not completely random. The reason for adopting such a strategy is due to the fact that the number of inter-class sample pairs is very large, and not all the sample pairs contribute to the learning of discriminative model. Hence it is reasonable to choose a subset of “specific” inter-class sample pairs near the classification boundary as candidates to construct the between-class scatter matrix.
- By integrating the MFDA with the densely sampled local feature descriptors, the resulting discriminative model is well suited for age invariant face recognition problem due to the following reasons: (i) the densely sampled local feature description scheme is both an extension and a combination of the SIFT and MLBP. Therefore, it is expected to inherit the discriminative properties of these local description schemes, and furthermore have the capability in extracting age invariant features such as the distribution of edge direction in the face. (ii) MFDA has the capability to effectively combine the rich information conveyed by densely sampled SIFT and MLBP descriptors, which are complementary to some extent. The recognition

performance in the experiments section demonstrates the effectiveness of the proposed discriminative model for age invariant face recognition.

III. EXPERIMENTS

A. *Experiment on the MORPH database*

In this section, we report results on experiments on a large public domain face aging data set, which is an extended version of the MORPH database [19]. While there are several public domain face data sets (e.g., FERET [34], XM2VTS [35], AR [36]), only a few are constructed specifically for the aging problem. The lack of a large face aging database until recently limited the research on age invariant face recognition. There are two desired attributes of a face aging database: (i) large number of subjects, and (ii) large number of face images per subject captured at many different ages. In addition, it is desired that these images should not have large variations in pose, expression, and illumination. The dataset that we use in this paper, MORPH album 2, is the largest face aging dataset available in the public domain. It is composed of about 78,000 face images of 20,000 different subjects captured at different ages. While the number of subjects in this database is large, the number of face images per subject is rather small (an average of about 4 face images per subject). Fig. 5 shows example images of one subject at different ages from MORPH album 2 [19]. Notice that there are large pose, lighting, and expression variations along with the age variation in this database. Hence, it is crucial to design an appropriate feature representation scheme which is tolerant to such multiple variations.

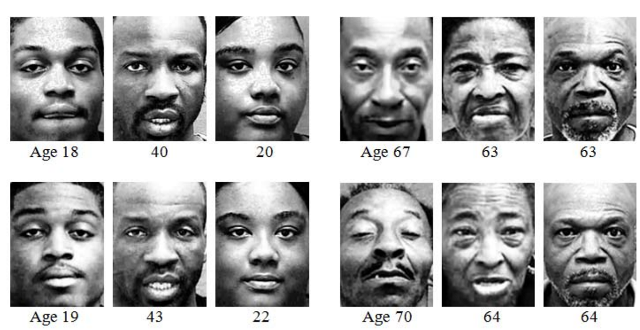
The MORPH album 2 data set is partitioned into a training set and an independent test set. For the training data used to learn the MFDA, we selected a subset of 20,000 face images from 10,000 subjects, with two images per subject. These two images were selected such that they had the largest age gap. The test data is composed of a gallery set and a probe set collected from the remaining 10,000 subjects. The gallery set is composed of 10,000 face images corresponding to the youngest age of these 10,000 subjects. The probe set is composed of 10,000 face images corresponding to the oldest age of these 10,000 subjects. Tables II and III show the statistics of the two publically available facial aging databases, FG-NET and MORPH, and the data set used in our experiments. Note that while FG-NET has only a small number of subjects, it has many more images per subject than MORPH. Also, FG-NET suffers from the fact that there is a large variation in expression, pose, and illumination among the images. To evaluate the recognition performance of our algorithm, all the facial images are automatically preprocessed through the following steps: (1) rotate the face image so that it is aligned with the vertical face orientation; (2) scale the face image so that the distance between the two eyes is the same for all the face images; (3) crop the face image tightly to remove the background and the hair region; (4) apply histogram equalization to the face image for photometric normalization. After preprocessing, each face image has a size of 150×200. Fig. 6 shows a schematic of the normalization process and Fig. 7 shows example images of training and testing data after preprocessing.

We first investigate the effectiveness of the local features by using the traditional LDA (Fisherface method) [25] as the classification technique. The comparative results are shown in Fig. 8, where the cumulative match characteristic (CMC) curves are used for performance evaluation. Instead of asking "Is the top match correct?", the CMC plot answers the question, "Is the correct match in the top-*n* matches?" The number *n* is called the rank of the match which indicates how many gallery images have to be examined to get a desired level of performance. The relatively poor result of the LDA on pixel values (the traditional Fisherface method) with the rank-1 accuracy of ~35% clearly shows that the raw pixel intensity values are not suitable for face recognition across aging. On the other hand, the local descriptor features (SIFT and MLBP) significantly outperform the global features by a large margin (with the rank-1 accuracy of ~55%), regardless of which local descriptor or patch size is used. This confirms that the proposed local descriptor features are suitable for age invariant face recognition.

Next we investigate the performance of the proposed multi-feature discriminant analysis (MFDA) framework, and also compare this framework against (i) FaceVACS (a leading commercial face recognition engine) [30] and (ii) a generative face aging model proposed in [18]. We used the same test data set to compare the face recognition performance of generative and discriminative models. The generative model is trained using MORPH Album 1 (1,690 images of 632 subjects). The poor feature point detection on MORPH Album 2 made it difficult to use the images in Album 2 for training the generative model. The comparative results are shown in Fig. 9, from which we make the following observations.

(i) Compared to the results of LDA in Fig. 8, the multi-feature discriminant analysis (MFDA) significantly boosts the recognition performance. The best rank-1 accuracy for LDA in Fig. 8 is 60% compared to the 83.9% accuracy of MFDA in Fig. 9. This shows the effectiveness of the MFDA.

(ii) The MFDA algorithm gives better recognition



(a) Training set (b) Testing set

Figure 7. Example images after preprocessing. Each column shows two different samples of the same subject.

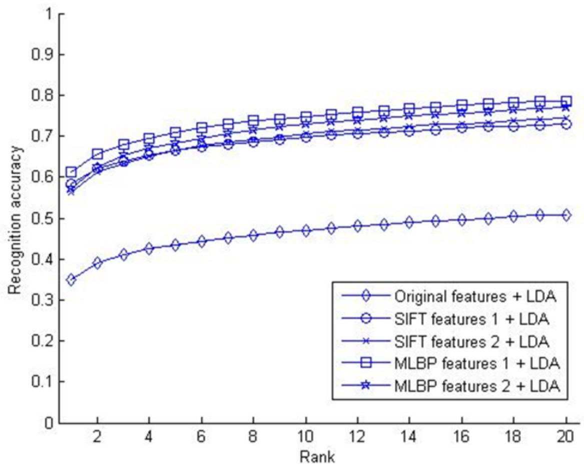


Figure 8. The cumulative match characteristic (CMC) curves with LDA applied to different feature representations: (i) Original features refer to raw pixel intensity values, (ii) SIFT feature 1 refers to SIFT features with patch size 16×16, (iii) SIFT feature 2 refers to the SIFT features with patch size 32×32, (iv) MLBP feature 1 refers to MLBP features with patch size 16×16, (v) MLBP feature 2 refers to MLBP features with patch size 32×32.

performance than the generative aging model [18]. The reason for the lower performance of the generative model compared to the proposed discriminative model is that the automatic facial landmark point detection which is required in the generative model, performs poorly on the extended MORPH database. This is because of the low image resolution (200×240 pixels) and large JPEG compression effect. Fig. 10 shows example face images with both successful and unsuccessful landmark detection results. The discriminative model does not need the landmarks; it only requires the coordinates of the two eyes for face alignment. The two eye coordinates are more robustly detected compared to the 68 landmark points needed by the generative model. This is one of the main advantages of the discriminative model over the generative model.

(iii) Both the generative and discriminative approaches outperform one of the best state-of-the-art face recognition system, FaceVACS. However, the discriminative approach offers more significant improvement (rank-1 accuracy of 83.9% compared to ~79% rank-1 accuracy of both generative model and FaceVACS).

(iv) A normalized score level fusion of discriminative and generative models further improves the recognition accuracy, but the improvement is marginal (rank-1 accuracy of 85.4% for the fused approach vs. 83.9% for the discriminative model). The recognition performance of generative model shown in Fig. 9 is the result of fusion of three different matching scores (i.e., original image, pose correction image, and aging simulation image) as was described in [18]. Therefore, the fusion of generative model and discriminative model already incorporates the fusion with scores for the original images. This shows the challenge associated with age invariant face recognition. As mentioned earlier, part of the challenge is that the available face aging datasets come with not just changes in the subject's age but also include

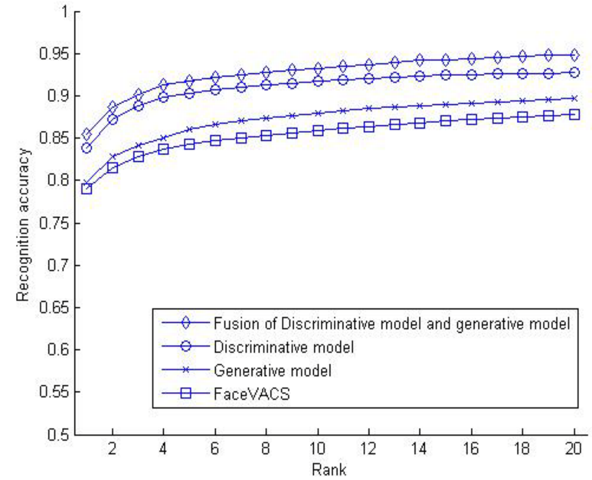


Figure 9. The cumulative matching characteristic (CMC) curves of different aging models.

variations due to pose, illumination and expression.

In order to further validate the effectiveness of the proposed method, we conducted an additional experiment to explore the robustness of the MFDA with respect to the training set. For this experiment, we first divided the entire training set into two subsets according to the intra-subject age gap, as shown in Table IV. For each subset, we randomly selected 2,000 subjects with two images per subject in the training set. This gives two different training subsets each with 4,000 images from 2,000 subjects. For the first training subset, the intra-subject age gap is 0. For the second training subset, the average intra-subject age gap is 1.5 years. We compared the recognition performance on the same test set based on these two different training sets. Fig. 11 shows that the performance of MFDA on the test set using training subset #2 (with an age gap in the two images of each subject) is slightly better than the training subset #1. This shows the proposed MFDA takes advantage of the aging information available in the training set for improved face recognition performance in the presence of age variations.

Fig. 12 shows some successful and failed examples of face matching. The failure of the discriminative method is mainly due to large pose changes, rather than the age change. It seems that FaceVACS is more effective in handling the pose variations while our method is more effective in handling the

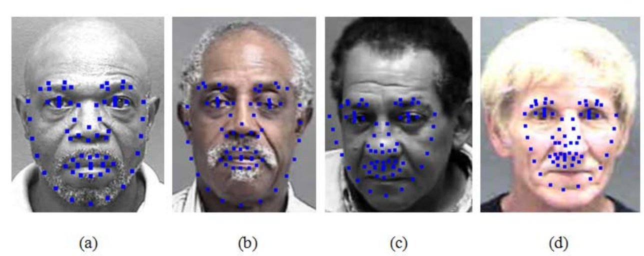


Figure 10. Facial landmark detection results. Blue points represent the 68 landmark points corresponding to those in Fig. 2. Figures in (a) and (b) show images where automatic landmark detection is successful; (c) and (d) show images where automatic landmark detection is not successful.

Table II. Statistics of public domain facial aging databases

DATABASE		#sub	#sub./gender		#sub./ethnicity					Total # image	#img/ sub	age range
			male	female	white	African-american	Hisp.	Asian	Other			
FG-NET		82	48	34	82	0	0	0	0	1,002	12.2	0~69
MORPH	Album1	632	525	107	172	459	0	0	1	1,690	2.7	15~68
	Album2	20,569	17,019	3,550	4,576	14,549	1,290	118	9	78,207	3.8	15~77

Table III. Summary of database (MORPH-Album2) used to train and test the proposed MFDA algorithm

Training dataset (20,000 images from subject 10001 to 20000)			Test dataset (20,000 images from the first 10,000 subjects)					
			Probe			Gallery		
#images	#subjects	Age range	#images	#subjects	Age range	#images	#subjects	Age range
20,000 (two/ subject)	10,000	15~76	10,000	10,000	20~77	10,000	10,000	18~76

Table IV. Constructing two different training subsets to evaluate the sensitivity of MFDA with respect to the training set.

	Training data
Subset #1	2,000 subjects are randomly selected from 2,636 subjects (age gap in the two images/subject is 0)
Subset #2	2,000 subjects are randomly selected from 7,364 subjects (the average age gap in the two images/subject is 1.5 years)

Table V. A comparison of age invariant face recognition methods.

	Approach	Database (# subjects, # images) in probe and gallery	Rank-1 recognition accuracy reported (%)
Lanitis et al. (2002) [14]	Build an aging function in terms of PCA coefficients of shape and texture	Private database (12,85)	68.5%
Ramanathan et al. (2006) [10]	Shape growth modeling up to age 18	Private database (109,109)	15.0%
Wang et al. (2006) [11]	Build an aging function in terms of PCA coefficients of shape and texture	Private database (NA,2000)	63.0%
Geng et al. (2007) [4]	Learn aging pattern on concatenated PCA coefficients of shape and texture across a series of ages	Public domain FG-NET (10,10)	38.1%
Park et al. (2010) [18]	Learn aging pattern based on PCA coefficients in separate 3D shape and texture spaces from the given 2D database	Public domain FG-NET (82,82)	37.4%
		Public domain MORPH Album 1 (612,612)	66.4%
		Public domain MORPH Album 2 (10000,20000)	79.8%
Proposed discriminative model	Use discriminative analysis method with densely sampled local descriptors	Public domain FGNET (82,82)	47.50%
		Public domain MORPH Album 2 (10000,20000)	83.9%

aging variations. Compared to FaceVACS, the proposed discriminative model is more adaptable to the age invariant face recognition problem. This is also supported by the results shown in Figs. 9 and 11. Finally, by integrating the discriminative model and the generative model, some failed matches can be corrected. This shows the advantage of a fusion framework, which integrates the generative and discriminative models.

B. *Experiment on the FG-NET database*

In order to verify the generality of the proposed discriminative model, we conducted an additional experiment on the FGNET database to compare our discriminative approach with the FaceVACS. The FGNET database is composed of 1,002 face images from 82 different subjects. In our experiment, we chose all the face images for performance evaluation. In order to keep the training data and testing data separated, the leave-one-out strategy is used in our study. The comparative results are shown in Fig. 13, from which we can clearly see that the proposed discriminative model outperforms the FaceVACS by a clear margin. This further validates the effectiveness of the proposed discriminative model.

IV. Conclusions and Future Work

A discriminative model for age invariant face recognition is proposed. The proposed approach addresses the face aging problem in a more direct way without relying on a generative aging model. This obviates the need of a training set of subjects that differ only in their age with minimal variations in illumination and pose, which is often a requirement to build a generative aging model. We first represent each face with a patch-based local feature representation scheme. In order to overcome the large feature dimensionality problem, we adopt a multi-feature discriminant analysis (MFDA) method to refine the feature space for enhanced recognition performance. Experimental results on two public domain databases (MORPH and FGNET) show the effectiveness of the proposed method. Our performance surpasses that of a commercial state-of-the-art face recognition engine. As shown in Table V, this is a very large evaluation of facial aging study reported in the literature.

Facial aging is a challenging problem that will require continued efforts to further improve the recognition performance. There are several directions for future work. First, since the generative model and the discriminative model offer somewhat complementary information, it is worthwhile to improve the fusion framework for enhanced performance. Second, as shown in Fig. 12, the proposed discriminative model is vulnerable to pose changes. A method more tolerant to pose changes should be studied in future work.

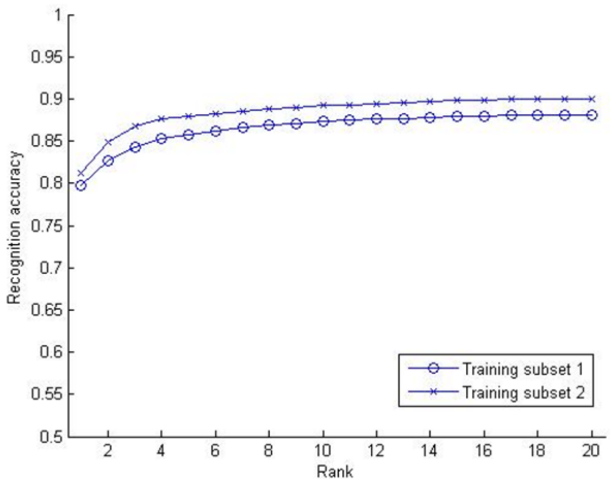


Figure 11. Cumulative matching characteristics (CMC) curves with two different training subsets.

ACKNOWLEDGMENT

The authors are grateful to Dr. Karl Ricanek of UNCW for providing the MORPH face aging database. All correspondence should be directed to Anil K. Jain.

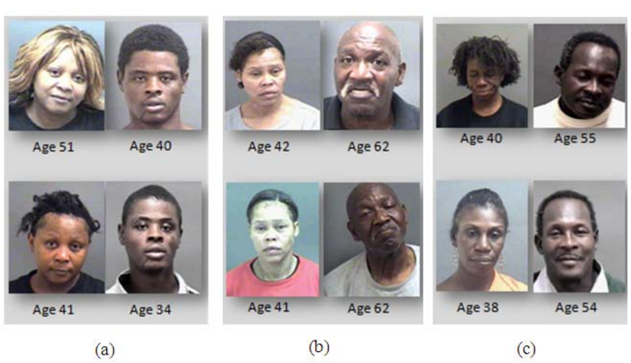


Figure 12. Example retrieval results at rank-1. The first row shows the probe images and the second row shows the gallery images retrieved at rank-1. (a) Two examples where both FaceVACS and the generative methods fail, while the discriminative method succeeds. (b) Two examples where the discriminative method fails, while the faceVACS and the generative method succeed. (c) Two examples where all the three methods (FaceVACS, generative model, and discriminative model) fail, but the fusion of the generative and discriminative models succeeds.

REFERENCES

[1] M. Albert, K. Ricanek, and E. Patterson, “A review of the literature on the aging adult skull and face: Implications for forensic science research and applications,” J. Forensic Science Int’l, vol. 172, no 1, October 2007, pp. 1-9.

[2] N. Ramanathan and R. Chellappa, “Computational Methods for modeling facial aging: A survey,” J. Visual Languages and Computing, vol. 20, no. 3, June 2009, pp. 131-144.

[3] Y. Fu and T. S. Huang, “Human age estimation with regression on discriminative aging manifold,” IEEE Trans. Multimedia, vol. 10, no. 4, pp. 578-584, Jun. 2008.

[4] X. Geng, Z. Zhou, and K. Smith-Miles, “Automatic age estimation based on facial aging patterns,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 29, no. 12, pp. 2234-2240, Dec. 2007.

[5] G. Guo, Y. Fu, C. Dyer, and T. Huang, “Image-based human age estimation by manifold learning and locally adjusted robust regression,” IEEE Trans. Image Process., vol. 17, no. 7, pp. 1178-1188, Jul. 2008.

[6] G. Guo, G. Mu, Y. Fu, and T. Huang, “Human age estimation using bio-inspired features,” in IEEE Conf. Computer Vision and Pattern Recognition, 2009, pp. 112-119.

[7] Y. Kwon and N. da Vitoria Lobo, “Age classification from facial images,” Computer Vision and Image Understanding, vol. 74, no. 1, pp. 1-21, 1999.

[8] A. Lanitis, C. Draganova, and C. Christodoulou, “Comparing different classifiers for automatic age estimation,” IEEE Trans. Syst., Man, Cybern, vol. 34, no. 1, pp., 621-628, Feb. 2004.

[9] A. Montillo and H. Ling, “Age regression from faces using random forests,” in IEEE Int. Conf. Image Processing, Cairo, Egypt, 2009.

[10] N. Ramanathan and R. Chellappa, “Face verification across age progression,” IEEE Trans. Image Process., vol. 15, no. 11, pp. 3349-3361, Nov. 2006.

[11] J. Wang, Y. Shang, G. Su, and X. Lin, “Age simulation for face recognition,” in Int. Conf. Pattern Recognition, pp. 913-916, 2006.

[12] S. Yan, H. Wang, X. Tang, and T. Huang, “Learning auto-structured regressor from uncertain nonnegative labels,” in Int’l Conf. Computer Vision, pp. 1-8, 2007.

[13] S. Zhou, B. Georgescu, X. Zhou, and D. Comaniciu, “Image based regression using boosting method,” in IEEE Int. Conf. Computer Vision, 2005, vol. 1, pp. 541-548.

[14] A. Lanitis, C. Taylor, and T. Cootes, “Toward automatic simulation of aging effects on face images,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 4, pp. 442-455, Apr. 2002.

[15] J. Suo, S. Zhu, S. Shan, and X. Chen, “A compositional and dynamic model for face aging,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 32, no. 3, pp. 385-401, March 2010.

[16] J. Suo, X. Chen, S. Shan, and W. Gao, “Learning long term face aging patterns from partially dense aging databases,” in Int. Conf. Computer Vision, pp. 622-629, 2009.

[17] N. Tsumura, N. Ojima, K. Sato, M. Shiraishi, H. Shimizu, H. Nabeshima, S. Akazaki, K. Hori, and Y. Miyake, “Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin,” ACM Trans. Graph., vol. 22, no. 3, pp. 770-779, 2003.

[18] U. Park, Y. Tong, and A. K. Jain, “Age Invariant Face Recognition, ” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 5, pp. 947-954, May 2010.

[19] K. Ricanek Jr and T. Tesafaye, “Morph: A Longitudinal Image Database of Normal Adult Age-Progression,” in Int’l Conf. Automatic Face and Gesture Recognition, pp. 341-345, 2006.

[20] D. Lowe, “Distinctive image features from scale-invariant keypoints,” Int’l Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

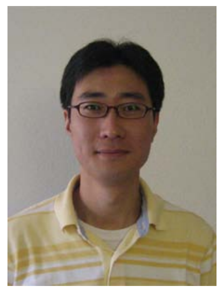
[21] T. K. Ho, “The random subspace method for constructing decision forests,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 8, pp. 832-844, 1998.

[22] L. Breiman, “Bagging predictors,” Machine Learning, vol. 24, no. 2, pp. 123-140, 1996.

[23] T. Ojala, M. Pietikainen, and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971-987, 2002.

[24] D. L. Swets and J. Weng, “Using discriminant eigenfeatures for image retrieval,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, pp. 831-836, Aug. 1996.
[25] P. Belhumeur, J. Hespanda, and D. Kiregeman, “Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720, July 1997.
[26] Z. Li, D. Lin, and X. Tang, “Nonparametric Discriminant Analysis for Face Recognition,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, no. 4, pp. 755-761, Apr. 2009.
[27] K. Fukunaga, Statistical Pattern Recognition, Academic Press, 1990.
[28] H. Ling, S. Soatto, N. Ramanathan, and D. Jacobs, “A Study of Face Recognition as People Age,” in Int’l Conf. on Computer Vision, 2007.
[29] FG-NET Aging Database, http://www.fgnet.rsunit.com/.
[30] FaceVACS Software Developer Kit, Cognitec Systems GbmH, http://www.cognitec-systems.de.
[31] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 10, pp. 1615-1630, Oct. 2005.
[32] A. K. Jain, K. Nandakumar, and A. Ross, “Score normalization in multimodal biometric systems,” in Audio- and Video-based Biometric Person Authentication, pp. 1049–1058, 2005.
[33] F. Roli, G. Giacinto, and G. Vernazza, “Methods for designing multiple classifiers systems,” Proc. of Second International Conference on Multiple Classifier Systems, pp. 78-87, 2001.
[34] P. Philips, H. Moon, S. Ryzvi, and P. Rauss, “The Feret evaluation methodology for face-recognition algorithms,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 10, pp. 1090-1104, 2000.
[35] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Matitre, “XM2VTSDB: The extended M2VTS database,” in Int’l Conf. Audio and Video-based Biometric Person Authentication, 1999.
[36] A. M. Martinez and R. Benavente, “The AR Face Database,” CVC Technical Report #24, Purdue University, June 1998.
[37] X. Wang and X. Tang, “Random Sampling for Subspace Face Recognition,” International Journal of Computer Vision, vol. 70, no. 1, pp. 91-104, 2006.
[38] H. Ling, S. Soatto, N. Ramanathan, and D. Jacobs, “Face verification across age progression using discriminative methods,” IEEE Trans. Information Forensic and Security, vol. 5, no. 1, pp. 82-91, Mar. 2010.
[39] G. Guo, G. Mu, and K. Ricanek, “Cross-age face recognition on a very large database: The performance versus age intervals and improvement using soft biometric traits,” in Int’l Conf. on Pattern Recognition, pp. 3392-3395, 2010.
[40] C. Fernandez and M. Vicente, “Face recognition using multiple interest point detectors and SIFT descriptors,” in Int’l Conf. on Automatic Face and Gesture Recognition, pp. 1-7, 2008.
[41] C. Geng and X. Jiang, “Face recognition using SIFT features,” in Int’l Conf. on Image Processing, pp. 3313-3316, 2009.
[42] T. Ahonen, A. Hadid, and M. Pietikainen, “Face Recognition with Local Binary Patterns,” Lecture Notes in Computer Science, pp. 469-481, 2004.
[43] T. Ahonen, A. Hadid, and M. Pietikainen, “Face Description with Local Binary Patterns: Application to Face Recognition,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 28, no. 12, pp. 2037-2041, Dec. 2006.
[44] S. Raudys and A. Jain, “Small sample size effects in statistical pattern recognition: recommendations for practitioners,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, no. 3, Mar. 1991.
[45] X. Wang and X. Tang. “A Unified Framework for Subspace Face Recognition”. IEEE Transaction on Pattern Analysis and Machine Intelligence, vol 26, no. 9, pp 1222-1228, 2004.
[46] X. Wang and X. Tang, “Dual-space linear discriminant analysis for face recognition,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 564-569, 2004.
[47] X. Wang and X. Tang, “Random sampling LDA for face recognition,” Proc. IEEE Conf. Computer Vision and Pattern Recognition,pp. 259-265, 2004.
[48] Z. Cao, Q. Yin, X. Tang, and J. Sun, “Face recognition with learning-based descriptor,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 2707-2714, 2010.

Zhifeng Li (M’06) received the Ph. D. degree from the Chinese University of Hong Kong in 2006. After that, he was a postdoctoral fellow at the Chinese University of Hong Kong and Michigan State University for several years. He is currently an associate professor at Shenzhen Institutes of Advanced Technology, Chinese Academy of Science. His research interests include computer vision, pattern recognition, and multimodal biometrics. He is a regular member of the program committee for some of the major computer vision conferences (e.g., International Conference on Computer Vision, International Conference on Computer Vision and Pattern Recognition, and European Conference on Computer Vision). He also has served as a reviewer for a number of major journals (e.g., IEEE Transactions on Pattern Analysis and Machine Intelligence, International Journal of Computer Vision, IEEE Transactions on Circuits and Systems for Video Technology, IEEE Transactions on Information Forensics and Security, etc.).



Unsang Park (S’08–M’08) received the B.S. and M.S. degrees from the Department of Materials Engineering, Hanyang University, South Korea, in 1998 and 2000, respectively. He received the M.S. and Ph.D. degrees from the Department of Computer Science and Engineering, Michigan State University, in 2004 and 2009, respectively. From 2009, he was a Postdoctoral Researcher in the Pattern Recognition and Image Processing Laboratory, Michigan State University, East Lansing. His research interests include biometrics, video surveillance, image processing, computer vision, and machine learning.

Anil K. Jain (S'70-M'72-SM'86-F'91) is a university distinguished professor in the Department of Computer Science and Engineering at Michigan State University, East Lansing. His research interests include pattern recognition and biometric authentication. He served as the editor-in-chief of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (1991–1994). The holder of six patents in the area of fingerprints, he is the author of a number of books, including Handbook of Fingerprint Recognition (2009), Handbook of Biometrics (2007), Handbook of Multibiometrics (2006), Handbook of Face Recognition (2005), BIOMETRICS: Personal Identification in Networked Society (1999), and Algorithms for Clustering Data (1988). He served as a member of the Defense Science Board and The National Academies committees on Whither Biometrics and Improvised Explosive Devices. Dr. Jain received the 1996 IEEE TRANSACTIONS ON NEURAL NETWORKS Outstanding Paper Award and the Pattern Recognition Society best paper awards in 1987, 1991, and 2005. He is a fellow of the AAAS, ACM, IAPR, and SPIE. He has received Fulbright, Guggenheim, Alexander von Humboldt, IEEE Computer Society Technical Achievement, IEEE Wallace McDowell, ICDM Research Contributions, and IAPR King-Sun Fu awards. ISI has designated him a highly cited researcher. According to Citeseer, his book Algorithms for Clustering Data (Englewood Cliffs, NJ: Prentice-Hall, 1988) is ranked #93 in most cited articles in computer science.