Multiview Discriminative Learning for Age-Invariant Face Recognition

Diana Sungatullina[1], Jiwen Lu[2], Gang Wang[2,3], and Pierre Moulin[4]
[1]*Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Moscow, Russia*
[2]*Advanced Digital Sciences Center, Singapore*
[3]*School of EEE, Nanyang Technological University, Singapore*
[4]*Department of ECE, University of Illinois at Urbana-Champaign, IL USA*
Email: diana.sungatullina@gmail.com; jiwen.lu@adsc.com.sg; wanggang@ntu.edu.sg; moulin@ifp.uiuc.edu

***Abstract*— In this paper, we propose a new multiview discriminative learning (MDL) method for age-invariant face recognition, which is a challenging and important problem in many practical face recognition systems. Motivated by the fact that local appearance features are more robust to age variations, we first extract three different local feature descriptors including scale invariant feature transform (SIFT), local binary patterns (LBP) and gradient orientation pyramid (GOP) for each face image to exploit the discriminative information. Then, we develop a discriminative learning method with multiview feature representations, called MDL, to project different types of local features into a latent discriminative subspace where the intraclass variation of each feature is minimized, the interclass variation of each feature and the correlation of different features of the same person are maximized, simultaneously, such that more discriminative information can be boosted for recognition. Experimental results on the widely used MORPH and FG-NET face aging datasets are presented to show the efficiency of the proposed approach.**

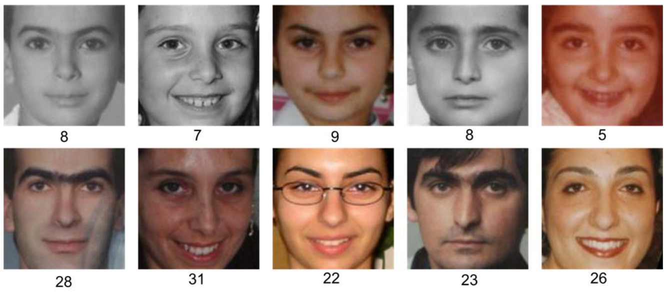

Fig. 1: Examples of face images from five different persons in the FG-NET database with different age values, where each column are face images of the same person captured at different age values and the number below each image is the age value of the person.

I. Introduction

Face recognition has been extensively investigated in computer vision and pattern recognition [15]–[18], [28]. However, most previous works aim to develop effective face representation and recognition algorithms which are robust to different imaging conditions, such as varying poses, illuminations and expressions. While a number of face recognition methods have been proposed in the literature [2], [26], [28], most of them cannot work well when facial images are collected under uncontrolled environments. Recently, a more unconstrained face database called "Labeled Faces in the Wild" (LFW) [8] was created and released, aiming at providing a data set applied to real face recognition applications. Since substantial face images were collected from the real world scenarios, the LFW dataset poses more natural variations rather than artificial controls or designs. While there have been some progresses achieved on this dataset in recent years, the age gap between gallery and probe images in the LFW database is limited. Hence, this dataset is not suitable to the research of age-invariant face recognition.

In this paper, we investigate the problem of age-invariant face recognition where there is an age gap between face images in the gallery and probe sets. The key challenge for age-invariant face recognition is how to reduce the feature gap between two facial images of the same person captured at different ages (e.g. in the childhood and adulthood, as shown in Fig. 1), such that more discriminative information can be exploited for recognition. Recent advances in face recognition have shown that local appearance features are more robust to age variations [13], [20]. Motivated by this finding, we extract three different local feature descriptors including scale invariant feature transform (SIFT) [14], local binary patterns (LBP) [1] and gradient orientation pyramid (GOP) [13] for each face image to exploit the discriminative information. Our purpose is to develop a discriminative learning method with multiview feature representations, to project different types of local features into a latent discriminative subspace where the intraclass variation of each feature is minimized, the interclass variation of each feature and the correlation of different features of the same person are maximized, simultaneously, such that more discriminative information can be boosted for recognition. Experimental results on the widely used MORPH and FG-NET face aging datasets are presented to demonstrate the effectiveness of our method.

The rest of this paper is organized as follows. Section II discusses related work. Section III details our proposed approach. Section IV provides the experimental results and Section V concludes the paper.

II. Related Work

While a large number of face recognition methods have been proposed in the literature, face recognition across age progression has not been extensively investigated. Most previous age-related face analysis studies focus on the problem of age estimation [4], [6], [7], [9], [10] and age simulation [5], [11], [20], [24], [25], [27], and only a few age-invariant face recognition algorithms have been proposed in recent years [3], [12], [13], [20]. These algorithms can be mainly classified into two categories: generative [20], [25] and discriminative [12], [13].

For generative methods, an age simulation approach is usually applied to transform one face image from one age to the target age such that the aging effect can be reduced. For example, Linitis *et al.* [11] proposed a statistical model to capture the variations of facial shape and texture over age progression. Ramanathan and Chellappa [24] presented a face growing model for face verification across age for people under the age of 18 years old. Park *et al.* [20] proposed a 3D aging model to compensate for the age variation to improve the face recognition performance. Suo *et al.* [25] presented a compositional and dynamic method to model human face aging. The compositional model represents faces in each age group by a hierarchical And-Or graph, in which the "And" nodes decompose a face into many local parts to describe facial details and Or nodes represent large diversity of faces by alternative selection. Wang *et al.* [27] presented an age simulation method by transforming facial shapes and textures of the person from the source age to the target age for face age-invariant face recognition. While these methods have explicitly addressed the aging effect in face recognition, they require more information about the captured face images, such as the actual age. Since both age estimation and age simulation are still open problems, there are still some instabilities for these generative age-invariant face recognition methods.

For discriminative models, robust feature descriptors and discriminative learning methods are usually applied to reduce the gap between face images collected at different ages. For example, Ramanathan and Chellappa [23] proposed a probabilistic eigenspace framework for face verification across age progression, where eigenspace techniques and the Bayesian model are combined to model the intrapersonal and interpersonal face differences. Ling *et al.* [13] proposed a gradient orientation pyramid (GOP) feature representation to describe facial image differences. Since only gradient orientation information is preserved for feature representation, it is more robust than other appearance features to age variation. More recently, Li *et al.* [12] presented a discriminative age-invariant face recognition framework by combining multiple local features, such as local binary pattern (LBP) [1] and SIFT [14], for face recognition. While these methods have achieved reasonably good performance in age-invariant face recognition, there are still two shortcomings: 1) most of them were only evaluated on small face datasets, and how their performance on large scale datasets remains unknown; 2) most of them used only a single feature for face representation, which may not be effective enough to extract sufficient discriminative information for recognition. To address this, we propose a new MDL method by making use of multiple local feature descriptors for enhanced age-invariant face recognition. Moreover, we conduct face recognition experiments on a large face dataset (26000 face images of 13000 subjects) to demonstrate the efficacy of the proposed method.

III. Proposed Approach

Our proposed approach consists of two components: local feature representation and multiview discriminative learning. We describe each component of the approach in the following subsections.

A. Local Feature Representation

Compared to global appearance features, local features have been shown to be more robust in representing face images versus different types of variations such as geometric distortions, illumination variations, and varying poses [12], [19]. Hence, we adopt local feature descriptors for face representation.

We divided each face image into several overlapping patches and then extracted local features from each patch. The extracted features were then concatenated together to construct a high-dimensional feature vector. Assume the original image size is $h \times w$, the image patch size is $a \times b$, the overlapping radius is r, the number of horizontal (m) and vertical (n) image patches are as follows:

$$m = \frac{h-a}{a-r} + 1 \tag{1}$$

$$n = \frac{w-b}{b-r} + 1 \tag{2}$$

For each image patch with size of $a \times b$, a d-dimensional local feature can be extracted. Then, these $m \times n$ feature vectors were concatenated into a long feature vector with size of $d{*}m{*}n$. In our experiments, the size of each whole image was 64×64 and the image was divided into 7×7 blocks with patch size 16×16 and overlapping radius $r = 8$. For each patch, we extracted three different local features: 128-dimensional SIFT [14], 256-dimensional LBP [1], and 512-dimensional GOP feature [13], respectively. Fig. 2 illustrates these local features of a face image.

Finally, we obtained three long feature vectors for each face image: 6272-dimensional SIFT feature, 12544-dimensional LBP feature and 25088-dimensional GOP feature. Since these features are high dimensional, we used PCA to project them into low-dimensional subspaces. In our implementations, each feature was projected into a 300-dimensional feature subspace.

B. Multiview Discriminative Learning (MDL)

Let $X_k = [x_{k1}, x_{k2}, \cdots, x_{kN}]$ the training set of the kth feature representation, where N is the number of samples in the training set and $1 \leq k \leq K$. In our scenario, $K = 3$ because there are three local features used. A

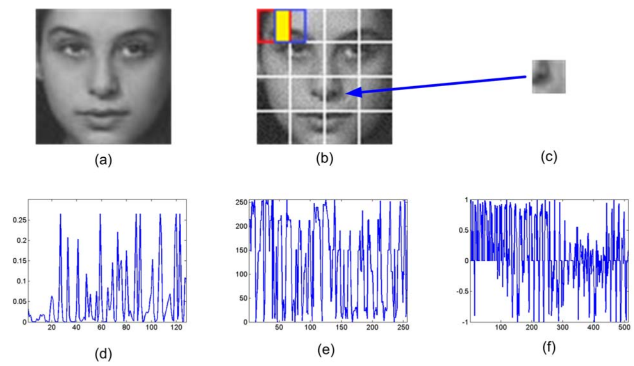

Fig. 2: Illustration of local feature representation of a face image. (a) Original aligned and cropped image. (b) Local patch division. (c) One divided local patch. (d) SIFT feature of the local patch. (e) LBP feature of the local patch. (f) GOP feature of the local patch.

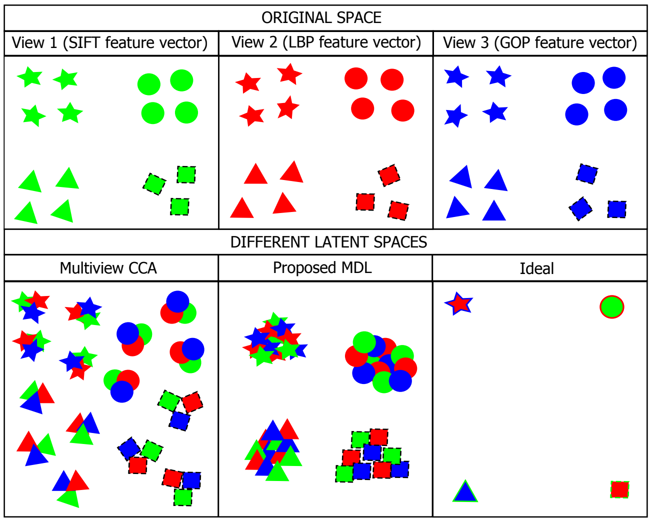

Fig. 3: A simple toy example to show the basic idea of our proposed MDL method. Samples denoted by the green, red and blue colors represent the SIFT, LBP and GOP features, respectively. Moreover, the samples represented by the same shape indicate that they are from the same class. The top row illustrates the original distribution of face samples in the SIFT, LBP and GOP feature spaces, respectively. The bottom row are the expected latent space learned by multiview CCA, our method and the ideal approach, respectively. In this figure, dashed outline squares are face samples from the an unseen class (not used in training). Ideally, we expect different classes (seen and unseen) to be well separated while all the within-class samples collapse to a point.

natural method to handle such multiview data representation is to concatenate different feature vectors of the same face together as a new vector and then apply some discriminative learning algorithm directly on the concatenated vector. However, this concatenation is not physically meaningful because each feature has a specific statistical property, and such concatenation ignores the diversity of multiple feature representations and thus cannot efficiently explore the complementary nature of different features. To address this, we propose a new learning algorithm, MDL, which learns a latent low-dimensional subspace by projecting different features into a common feature space, such that

1) The correlations of different feature representations of each sample are maximized, because different features of each sample share the same semantic label.
2) The within-class variation of each feature is minimized and the between-class variation of each feature is maximized, such that more discriminative information can be exploited in the latent space.

Fig. 3 shows the basic idea of our MDL method.

To achieve the above goal, we formulate our MDL method as the following optimization problem:

$$\begin{aligned} [w_1^*, w_2^*, w_3^*] &= \arg\max_{w_1,w_2,w_3} J_1(w_1,w_2,w_3) \\ &\quad + J_2(w_1,w_2,w_3) \\ \text{subject to} &\quad J_3(w_1,w_2,w_3) = 1 \end{aligned} \tag{3}$$

where

$$J_1 = \mu_1 w_1^T A_1 w_1 + \mu_2 w_2^T A_2 w_2 + \mu_3 w_3^T A_3 w_3 \tag{4}$$

$$\begin{aligned} J_2 &= 2\mu_{12} w_1^T X_1 X_2^T w_2 + 2\mu_{13} w_1^T X_1 X_3^T w_3 \\ &+ 2\mu_{23} w_2^T X_2 X_3^T w_3 \end{aligned} \tag{5}$$

$$\begin{aligned} J_3 &= \mu_1 w_1^T B_1 w_1 + \mu_2 w_2^T B_2 w_2 + \mu_3 w_3^T B_3 w_3 \\ &+ (1-\mu_1) w_1^T C_1 w_1 + (1-\mu_2) w_2^T C_2 w_2 \\ &+ (1-\mu_3) w_3^T C_3 w_3 \end{aligned} \tag{6}$$

where A_k and B_k are the between-class scatter and the within-class scatter matrices of the kth feature representation, $C_k = X_k D_k X_k^T$ is covariance matrix, where $D_k = I_k/N_k$ with N_k equal to number of samples and I_k is identity matrix for kth view. Here $1 \leq k \leq 3$, μ_1, μ_2, μ_3 are parameters to balance the contributions of different features to compute the within-class and between-class variations. μ_{12}, μ_{13} and μ_{23} are three parameters to weight the correlation of different features, and they are empirically set as 1, 1, and 1, respectively. w_1, w_2 and w_3 are projections for three different features and map these features into a common latent subspace.

The expression in Eq. (3) can be simplified to the following form:

$$\begin{aligned} \begin{bmatrix} w_1^* \\ w_2^* \\ w_3^* \end{bmatrix} &= \arg\max_{w_1,w_2,w_3} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}^T F \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \\ \text{subject to} &\quad \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}^T (G+H) \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = 1. \end{aligned} \tag{7}$$

where

$$F = \begin{bmatrix} \mu_1 A_1 & \mu_{12} X_1 X_2^T & \mu_{13} X_1 X_3^T \\ \mu_{12}(X_1 X_2^T)^T & \mu_2 A_2 & \mu_{23} X_2 X_3^T \\ \mu_{13}(X_1 X_3^T)^T & \mu_{23}(X_2 X_3^T)^T & \mu_3 A_3 \end{bmatrix} \tag{8}$$

$$G = \begin{bmatrix} \mu_1 B_1 & 0 & 0 \\ 0 & \mu_2 B_2 & 0 \\ 0 & 0 & \mu_3 B_3 \end{bmatrix} \tag{9}$$

$$H = \begin{bmatrix} (1-\mu_1)C_1 & 0 & 0 \\ 0 & (1-\mu_2)C_2 & 0 \\ 0 & 0 & (1-\mu_3)C_3 \end{bmatrix} \tag{10}$$

Here, F is symmetric, G and H are symmetric definite matrices.

Let $w = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$. Eq. (7) can be further rewritten as

$$\begin{aligned} w^* &= \arg\max_w w^T F w \\ \text{subject to} \quad & w^T(G+H)w = 1. \end{aligned} \tag{11}$$

The projection w can be easily obtained by solving the following generalized eigenvalue problem

$$Fw = \lambda(G+H)w. \tag{12}$$

Let $w^1, w^2, \cdots, w^q$ be the eigenvectors of Eq. (12) corresponding to the q largest eigenvalues ordered according to $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_q$. A transformation matrix $W = [w^1, w^2, \cdots, w^q]$ is the projection matrix of our MDL method, where $w^i = \begin{bmatrix} w_{1i} \\ w_{2i} \\ w_{3i} \end{bmatrix}$.

In our implementations, each feature after PCA preprocessing was projected into 100-dimensional feature subspace ($q = 100$).

IV. Experiments

In this section, we evaluate our proposed method by conducting age invariant face recognition experiments on the widely used FG-NET and MORPH face databases. The following describes the details of the experiment setups and results.

A. *Experiments on the MORPH Database*

While there are some publicly available face datasets such as FERET [22], FGRC [21] and LFW [8] created and released for the face recognition research, only a few of them have addressed the aging problem until some recent attempts on age invariant face recognition. There are two desirable attributes for a face aging database: 1) a large number of subjects, and 2) a large number of face images per subject captured at many different age values. In addition, face images should not have large variations in pose, expression, and illumination such that the aging variation can be well modeled. In this study, the MORPH (version 2) face aging data set was used because it is the largest face aging dataset available in the public domain. It contains about 78000 face images of over 13000 different subjects captured at different ages. In our study, we partitioned the MORPH dataset into a training set and an independent testing set. For the training set, we selected a subset of 10000 face images from 5000 subjects and each subject has two images. These two images were selected such that they had the largest age gap. The test set consists of a gallery set and a probe set collected from the remaining 16000 face images corresponding to the youngest age of these 80000 subjects. The probe set is composed of 8000 face images corresponding to the oldest age of these 8000 subjects.

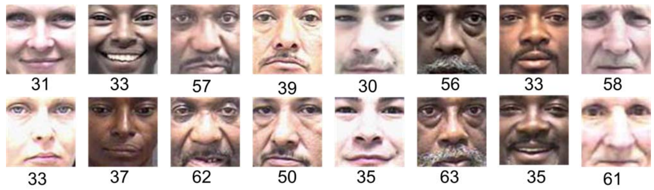

Fig. 4: Examples of face images from ten different persons in the MORPH database with different age values, where each column are two face images of the same person captured at different age values and the number below each image is the age value of the person. The left two column are face images of two female subjects and the remaining columns are face images of six male subjects, respectively.

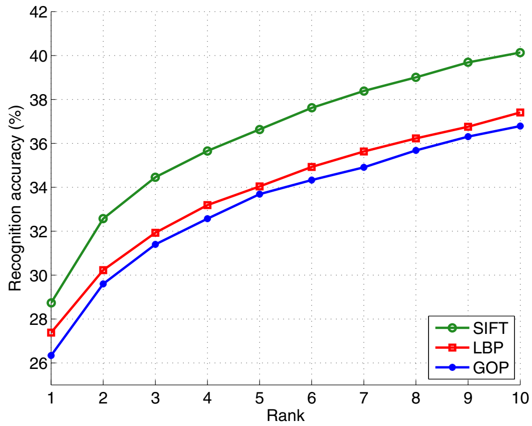

Fig. 5: Cumulative match characteristic (CMC) curves of different local feature representations when PCA is applied.

To evaluate the recognition performance of our approach, each face image was automatically preprocessed by the following steps: 1) rotate each 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 and resize each image into 64×64. Fig. 4 shows several example cropped face images of eight subjects at different ages from the MORPH dataset.

Performance Comparison of Single Feature: We first investigated the discriminative power of different local features by using the conventional PCA and LDA with the nearest neighbor classifier as the classification methods. The comparative results are shown in Figs. 5-6, where the cumulative match characteristic (CMC) curves are used for performance evaluation. We can observe from these two figures that the SIFT feature performs the best for both the PCA and LDA classification methods.

Parameter Determination: Since it is very difficult to

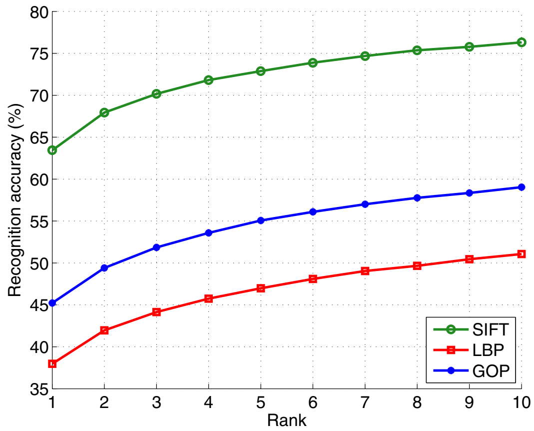

Fig. 6: Cumulative match characteristic (CMC) curves of different local feature representations when LDA is applied.

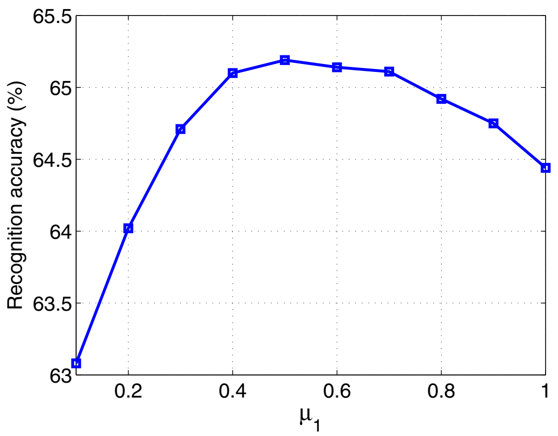

Fig. 7: Rank-1 recognition accuracy versus different values of μ_1, $\mu_2 = 1$ and $\mu_3 = 1$.

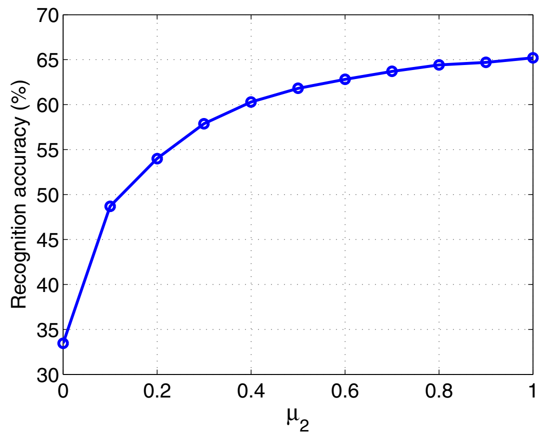

Fig. 8: Rank-1 recognition accuracy versus different values of μ_2, $\mu_1 = 0.5$ and $\mu_3 = 1$.

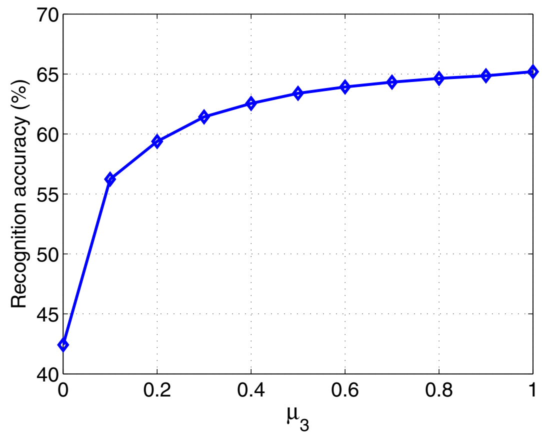

Fig. 9: Rank-1 recognition accuracy versus different values of μ_3, $\mu_1 = 0.5$ and $\mu_2 = 1$.

determine three parameters (μ_1, μ_2, and μ_3) at the same time, we adopted a stepwise selection method to seek these parameters by using a cross-validation strategy. Specifically, we divided the training set into 5 folders, and used 4 folders to learn our MDL model and the remaining one to tune these parameters. We prefix both μ_2, and μ_3 to 1 and try to find the optimal μ_1 value, as shown in Fig. 7. After the determination of $\mu_1 = 0.5$, we prefix $\mu_3 = 1$ and seek the optimal μ_2 value, as shown in Fig. 8. Having determined μ_1 and μ_2, we search the optimal μ_3, as shown in Fig. 9.

Comparison with Existing Discriminative Learning Methods: We compare our MDL method with the following four baselines:

1) Single Feature: each single feature is used for feature representation and LDA is used for discriminative learning. Before applying LDA, each feature was projected into a 300-dimensional PCA subspace.
2) Feature Concatenation: Performing PCA to project each feature into a 100-dimensional feature subspace and concatenating these features of each view together, and then running LDA from the joint representation of the data.
3) Multiview CCA + LDA: CCA is applied to map different features into a common 100-dimensional feature subspace and then LDA is used to extract discriminative feature. PCA is also applied to project each feature into 300 dimensions before CCA is used.
4) Multiview LDA: Only J_1 and J_3 without the covariance matrices are considered in our method to learn the feature space. We also used PCA to project each feature into 300-dimensional subspace.

Table I tabulates the rank-1 recognition accuracy of these methods. We can see from this figure that our method can achieve the best recognition accuracy among all the compared methods.

B. Experiments on the FG-NET Database

In order to verify the effectiveness of our proposed approach, we also conducted age-invariant face recognition experiments on the FG-NET database to compare our discriminative approach with other baselines. There are 1002 images of 82 different subjects (see Fig. 1 for some example images in this dataset). In our experiment, we used all face images for performance evaluation. In order to separate the training (learning) and test phases, we adopted the leave-one-out strategy is our experiments. Specifically, we left one image as the test sample and trained the model by using the remaining 1001 images. We repeated this procedure 1002 times and took the average as the final recognition accuracy. Table III records the rank-1 recognition accuracy of different methods on the FG-NET database. We can observe that our method also outperforms each single feature and the feature concatenation methods.

V. Conclusions and Future Works

In this paper, we have proposed a multiview discriminative learning approach to age-invariant face recognition. Experimental results on two widely used face aging datasets have shown the efficacy of our approach. How to explore more other discriminative age-invariant features and combine them with SIFT, LBP and GOP features to further improve the recognition performance appears to be an interesting direction of our future work.

Acknowledgement

This work is supported by the research grant for the Human Sixth Sense Program at the Advanced Digital Sciences Center (ADSC) from the Agency for Science, Technology and Research (A*STAR) of Singapore.

References

[1] T. Ahonen, A. Hadid, et al. Face description with local binary patterns: application to face recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, 2006.

[2] P. N. Belhumeur, J. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

TABLE I: Rank-1 recognition accuracy of different methods on the MORPH dataset.

Method	Rank-1 accuracy (%)
SIFT + LDA	63.4
LBP + LDA	38.0
GOP + LDA	45.2
Feature Concatenation + LDA	52.2
Multiview CCA + LDA	53.5
Multiview LDA	58.4
MDL	**65.2**

TABLE II: Rank-1 recognition accuracy of different methods on the FG-NET dataset.

Method	Rank-1 accuracy (%)
SIFT	34.9
LBP	32.8
GOP	27.3
Feature Concatenation	34.3
MDL	**91.8**

[3] S. Biswas, G. Aggarwal, N. Ramanathan, and R. Chellappa. A non-generative approach for face recognition across aging. In *IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 1–6, 2008.

[4] Y. Fu and T. S. Huang. Human age estimation with regression on discriminative aging manifold. *IEEE Trans. on Multimedia*, 10(4):578–584, 2008.

[5] X. Geng, Z. Zhou, and K. Smith-Miles. Automatic age estimation based on facial aging patterns. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29(12):2234–2240, 2007.

[6] G. Guo, Y. Fu, C. Dyer, and T. Huang. Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Trans. on Image Processing*, 17(7):1178–1188, 2008.

[7] G. Guo, G. Mu, Y. Fu, and T. Huang. Human age estimation using bio-inspired features. In *IEEE International Conference on Computer Vision and Pattern Recognition*, pages 112–119, 2009.

[8] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical report, 2007.

[9] A. Lanitis. Evaluating the performance of face-aging algorithms. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 1–6, 2008.

[10] A. Lanitis, C. Draganova, and C. Christodoulou. Comparing different classifiers for automatic age estimation. *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34(1):621–628, 2004.

[11] A. Lanitis, C. J. Taylor, and T. Cootes. Toward automatic simulation of aging effects on face images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(4):442–455, 2002.

[12] Z. Li, U. Park, and A. Jain. A discriminative model for age invariant face recognition. *IEEE Trans. on Information Forensics and Security*, 6(3):1–1.

[13] H. Ling, S. Soatto, N. Ramanathan, and D. Jacobs. Face verification across age progression using discriminative methods. *IEEE Trans. on Information Forensics and Security*, 5(1):82–91, 2010.

[14] D. Lowe. Distinctive image features from scale-invariant keypoints. *Interntaional Journal of Computer Vision*, 60(2):91–110, 2004.

[15] J. Lu and Y. Tan. A doubly weighted approach for appearance-based subspace learning methods. *IEEE Transactions on Information Forensics and Security*, 5(1):71–81, 2010.

[16] J. Lu and Y. Tan. Regularized locality preserving projections and its extensions for face recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 40(3):958–963, 2010.

[17] J. Lu and Y. Tan. Uncorrelated discriminant nearest feature line analysis for face recognition. *IEEE Signal Processing Letters*, 17(2):185–188, 2010.

[18] J. Lu, Y. Tan, and G. Wang. Discriminative multi-manifold analysis for face recognition from a single training sample per person. In *IEEE International Conference on Computer Vision*, pages 1943–1950, 2011.

[19] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

[20] U. Park, Y. Tong, and A. Jain. Age-invariant face recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 32(5):947–954, 2010.

[21] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *CVPR*, volume 1, pages 947–954, 2005.

[22] P. Phillips, H. Moon, S. Rizvi, and P. Rauss. The feret evaluation methodology for face-recognition algorithms. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.

[23] N. Ramanathan and R. Chellappa. Face verification across age progression. *IEEE Trans. on Image Processing*, 15(11):3349–3361, 2006.

[24] N. Ramanathan and R. Chellappa. Modeling age progression in young faces. In *IEEE International Conference on Computer Vision and Pattern Recognition*, pages 387–394, 2006.

[25] J. Suo, S. Zhu, S. Shan, and X. Chen. A compositional and dynamic model for face aging. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 32(3):385–401, 2010.

[26] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[27] J. Wang, Y. Shang, G. Su, and X. Lin. Age simulation for face recognition. In *International Conference on Pattern Recognition*, pages 913–916, 2006.

[28] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: a literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.