A Markov Random Field Groupwise Registration Framework for Face Recognition

Shu Liao,
Department of Computer Science and Engineering, the Hong Kong University of Science and Technology, Hong Kong

Dinggang Shen, and
Department of Radiology, BRIC, the University of North Carolina at Chapel Hill, CB# 7513, 130 Mason Farm Rd., Chapel Hill, NC

Albert C.S. Chung
Department of Computer Science and Engineering, the Hong Kong University of Science and Technology, Hong Kong

Shu Liao: liaoshu.cse@gmail.com; Dinggang Shen: dinggang_shen@med.unc.edu; Albert C.S. Chung: achung@cse.ust.hk

Abstract

In this paper, we propose a new framework for tackling face recognition problem. The face recognition problem is formulated as groupwise deformable image registration and feature matching problem. The main contributions of the proposed method lie in the following aspects: (1) Each pixel in a facial image is represented by an anatomical signature obtained from its corresponding most salient scale local region determined by the survival exponential entropy (SEE) information theoretic measure. (2) Based on the anatomical signature calculated from each pixel, a novel Markov random field based groupwise registration framework is proposed to formulate the face recognition problem as a feature guided deformable image registration problem. The similarity between different facial images are measured on the nonlinear Riemannian manifold based on the deformable transformations. (3) The proposed method does not suffer from the generalizability problem which exists commonly in learning based algorithms. The proposed method has been extensively evaluated on four publicly available databases: FERET, CAS-PEAL-R1, FRGC ver 2.0, and the LFW. It is also compared with several state-of-the-art face recognition approaches, and experimental results demonstrate that the proposed method consistently achieves the highest recognition rates among all the methods under comparison.

Index Terms

Face recognition; deformable image registration; groupwise registration; Markov random field; correspondences; anatomical signature

---

1 Introduction

Automatic face recognition (AFR) plays an important role in computer vision. Its application includes, but not limited to financial security, human-computer interaction, and law enforcement. AFR remains an active yet challenging research topic mainly due to three issues. First, facial expressions of the same person can cause large and deformable motions across different facial images [1], [2]. Second, the image appearances can be significantly altered due to illumination changes [3], [4]. Third, facial images taken under different poses also bring additional difficulty in achieving high recognition rates [5], [6].

Many novel methods have been proposed in the literature for AFR, and they can be broadly classified into two main categories: holistic methods [7], [8], [9], [10] and local feature matching methods [11], [12], [13], [14]. Holistic methods use the whole facial regions as input and derive a salient subspace to analyze the similarity between different facial images. Therefore, the core problem of holistic methods is about how to define the principles and optimization criteria to construct the subspace such that the facial images can be projected to the subspace and their similarity can be measured. For instance, Turk and Pentland [7] used the principle component analysis (PCA), which is also known as "eigenface" to project the facial images to the subspace with minimum least square reconstruction error. Belhumeur et al. [8] proposed the use of linear discriminant analysis (LDA) to project facial images to the subspace which simultaneously maximizes the inter-class distances while minimizing the intra-class variations. Bartlett et al. [15] proposed the independent component analysis (ICA) to construct the subspace such that higher order pixel-wise relationship can be captured. In order to analyze facial images in the nonlinear high dimensional feature space, kernel based methods were also proposed [16], [17]. In [18], Yan et al. proposed a general graph embedding framework, where different dimensionality reduction and subspace learning methods such as PCA [7], LDA [8], LPP [19], ISOMAP [20], and LLE [21] can all be reformulated within this framework. Recently, Wright et al. [22] proposed a sparse representation framework for face recognition in the original facial space, and Naseem et al. [23] proposed a linear regression approach for face recognition.

Local feature matching methods extract image appearance features from different local regions of facial images, and the extracted features are combined and served as the input to a classifier. It is shown that local feature matching methods generally are more robust to local illumination changes and expression variations [13], [14], [24]. Two representative features used in local feature matching methods are Gabor wavelet [25] and local binary patterns (LBP) [11]. Gabor wavelet can be viewed as bandpass filters which analyze facial images in different frequency bands, with different orientations and scales. LBP is a powerful yet efficient local image descriptor which is originally proposed for texture classification [26] and has been widely extended to other classification problems by different researchers [11], [27]. In recent years, many local feature matching methods were developed based on Gabor wavelet and LBP. For instance, Zhang et al. [12] extracts LBP features from the Gabor filtered responses for face recognition. In [13] and [14], magnitudes and phases of the Gabor filtered responses are integrated with microscopic local pattern features to conduct face recognition. A comprehensive study about the comparison between holistic methods and local feature matching methods for face recognition can be found in [24].

In recent years, there are new methods proposed to model the facial expression process as diffeomorphic transformations [28], [29], [30] to aim the recognition task. For instance, Guo et al. [29] proposed a generative method to model the dynamic facial expression with the diffeomorphic growth model. Moreover, image registration is also served as a possible solution for pose-invariant face recognition problems [31], [32]. In this paper, we propose a new way to tackle face recognition problem, which is formulated as groupwise deformable image registration and feature matching. The basic principle of the proposed method is to first construct the common group mean facial image space on the Riemannian manifold, and the similarity among different facial images is compared by warping facial images to the common group mean space. The main contributions of the proposed method are summarized as follows. First, instead of using pixel intensity alone, anatomical features are extracted from each pixel position of the facial images from its corresponding most salient scale local regions. A new salient region detector is proposed based on the survival exponential entropy (SEE) theoretical measure. Second, based on the anatomical signature calculated from each pixel position, a feature guided Markov random field (MRF) groupwise registration framework is proposed to construct the group mean facial image space on the Riemannian manifold in hierarchical manner. Finally, the proposed method is an unsupervised learning method. A preliminary version of this work appeared in [33].

The proposed method has been extensively evaluated on four publicly available databases: FERET, CAS-PEAL-R1, FRGC ver 2.0, and LFW. It is also compared with several state-of-the-art face recognition approaches, and experimental results demonstrate that the proposed method consistently achieves the highest recognition rates among other methods under comparison.

The rest of the paper is organized as follows: Section 2 describes the background with respect to groupwise deformable image registration and shows how to formulate the face recognition problem as a groupwise image registration problem. Section 3 introduces the feature guided MRF groupwise registration framework for face recognition. Section 4 gives the experimental results and related analysis. Section 5 concludes the whole paper.

2 Groupwise Image Registration and Its Usage for Face Recognition

In this section, the background knowledge about groupwise deformable image registration is described and the motivation of formulating the face recognition problem as group-wise image registration problem is presented.

2.1 Groupwise Image Registration

In computer vision, the role of image registration is to transform images taken from different times, sensors, viewpoints, or different coordinate systems into a common coordinate system or space, such that comparisons can be made across different images in a common image space.

Given n input images $I_1, \ldots, I_n$, the conventional pairwise registration strategy first selects an image from $I_1, \ldots, I_n$, namely the fixed image I_{fix}, to serve as template. Then, the goal is to transform each image I_i ($i = 1, \ldots, n$), namely the moving image I_{mov}, to the space of fixed

image I_{fix} [34], [35]. The pairwise registration process can be viewed as an iterative optimization problem, as shown in Fig. 1. First of all, image features are extracted from input images. Of course, the most simple feature is the original facial images alone. Then, the optimal transformation φ_{opt} is estimated based on a pre-defined parametric transformation model (in this paper, it is deformable transformation). φ_{opt} is estimated by optimizing the value of a similarity measure function E, which reflects the registration quality at the current iteration. To optimize E, an optimization scheme (i.e., optimizer) is needed. Also, interpolation of the moving image is required in the case that some pixels of the transformed moving image do not fall exactly on the image grid of the fixed image. The registration process thus can be expressed by

$$\phi_{opt} = \arg \min_{\phi} E(\psi \otimes I_{fix}, \phi(\psi \otimes \phi(I_{mov}))), \quad (1)$$

where ψ denotes the feature extraction kernel, $\otimes$ denotes the convolution operation.

However, it is recently observed that [36], [37] explicitly selecting one of the images as the fixed image will lead to bias in registering all the other images to it. The main reason is that the geodesic distance between the fixed image and some of the moving images on the Riemannian manifold represented by deformable transformations may be very large and difficult to register. Therefore, groupwise registration strategies have become widely used [36], [37], [38], [39]. Groupwise registration does not explicitly select an image as the template. Instead, it simultaneously estimates the template $\hat{I}$ (i.e., the group mean) and the transformation φ_i to warp each image I_i ($i = 1, \ldots, n$) to $\hat{I}$ by minimizing

$$J(\hat{I}, \phi_i) = \sum_{i=1}^{n} (\varphi(\hat{I}, \phi_i(I_i))^2 + \lambda Reg(\phi_i)), \quad (2)$$

where $\phi(\cdot)$ is the cost function reflects the matching degree, and $Reg(\cdot)$ is the regularization term enforced on φ_i to prevent unrealistic and unsmooth deformable transformations. λ is the parameter controlling the tradeoff between the matching term and the regularization term. $\hat{I}$ is the Fréchet mean on the non-euclidean Riemannian manifold defined by

$$\hat{I} = \frac{1}{n} \sum_{i=1}^{n} \phi_i(I_i). \quad (3)$$

The general groupwise registration framework is illustrated in Fig. 2.

2.2 Face Recognition as a Groupwise Image Registration Process

Intuitively, faces are topological objects, and motions of faces have close relationship with deformable transformations. For instance, the variation caused by facial expressions, which is one of the major challenges for face recognition, is a physical deformable and topology preserving transformation. It can be well formulated by the registration problem reflected in Equations (1) and (2). Similarly, if robust anatomical features and similarity measure

functions are defined, then the registration process can also be robust against illumination changes.

Therefore, we are motivated to formulate the face recognition problem as deformable image registration and feature matching problem. The most straightforward solution is to register each query image to each training image in a pairwise manner, and classify the query image to the person which has the resulting smallest energy function value in Equation (1). However, as discussed in Section 2.1, formulate face recognition in this way has two limitations. First, the query image is to transform each individual training image's space for comparison, comparing the energy function value obtained in different image space is problematic. Second, this strategy is sensitive to outliers and registration to a specific training image already introduced bias. To resolve the above two limitations, we propose a novel groupwise registration framework for face recognition, the basic principle of the proposed method is summarized in Algorithm 1.

Algorithm 1

Face Recognition with the Groupwise Registration Framework

Input: Query image I_{new}, n training images $I_1,\ldots, I_n$.

Output: A class label l_{new} assign to I_{new}.

1. Construct the template image by performing groupwise registration among the training images $I_1,\ldots, I_n$ denote the estimated template as , and the deformable transformation from I_i ($i = 1,\ldots, n$) to   as φ_i.

2. Register I_{new} to , denote the registered query image as $\varphi_{new}(I_{new})$, where φ_{new} is the optimal transformation to Warp I_{new} to .

3. Calculate the similarity between $\varphi_{new}(I_{new})$ and each $\varphi_i(I_i)$, set l_{new} as the class label of the transformed training image which is most similar to $\varphi_{new}(I_{new})$.

4. Return l_{new}.

In the next section, we will introduce the feature guided hierarchical MRF groupwise registration framework for face recognition.

3 The Feature Guided Hierarchical MRF Groupwise Registration Framework

There are two main stages for the proposed registration framework. First, anatomical features are extracted from each pixel position at its corresponding most salient scale as the pixel signature. Second, based on the extracted features, the deformation model is formulated as a MRF labeling problem to perform groupwise registration in a hierarchical manner.

3.1 Anatomical Signature Construction for Each Pixel

The first step of the proposed registration framework is to extract salient anatomical features at each pixel location to reflect the structural property around the pixel in facial images.

It is shown in [40] that saliency of features is closely related to the scale from which the features are extracted. Important structures in facial images such as noses, mouths, and eyes have different sizes and shapes. Therefore, anatomical features should be extracted with

different scales of interest from different facial regions to ensure their saliency and representation power.

We measure the saliency based on local image structural complexity. Specifically, the survival exponential entropy [41], [42] was used as a statistical measure for local image structural complexity. It is defined by

$$M_\alpha(X) = \left(\int_{R_+^m} \overline{F}_{|X|}^\alpha(x) dx \right)^{\frac{1}{1-\alpha}}, \quad (4)$$

where $\alpha \geq 0$ is the order of SEE, it is set to 3 in this paper by cross validation. $X \in R^m$ is a m-dimensional random vector. Let $X = [X_1, \ldots, X_m]^T$, $|X|$ denotes the random vector $|X| = [|X_1|, \ldots, |X_m|]^T$. $\overline{F}_{|X|}^\alpha(x)$ is the multivariate survival function defined by

$$\overline{F}_{|X|}^\alpha(x) = P(|X| > x) = P(|X_1| > x_1, \ldots, |X_m| > x_m), \quad (5)$$

where $x \in R_+^m$, and R_+^m is defined as:

$$R_+^m = \{ x \in R^m : x = (x_1, \ldots, x_m), x_i \geq 0, i = 1, \ldots, m \}. \quad (6)$$

In Equation (4), the random vector x is sampled by the following procedure: First, each dimension of x is normalized to the same range [0, 1]. Then each dimension of the normalized random vector is uniformly partitioned into 64 bins.

Comparing to the conventional Shannon entropy, SEE has the following advantages [41], [42]: (1) SEE is always non-negative; (2) SEE has consistent definition in both the continuous and discrete domains; and (3) the probability density function (pdf) used to compute the Shannon entropy may not exist. However, the survival function defined by Equation (5) to calculate SEE always exists.

For each pixel p, we denote the square region centered at it with scale (i.e., the side length) s as $R_s(\vec{p})$. We also denote the histogram of intensity distribution in $R_s(\vec{p})$ as $H_s(\vec{p})$. It is clear that the larger the value of $M_\alpha(H_s(\vec{p}))$, the more complex the image structures contained in $R_s(\vec{p})$. However, it should also be noted that regions exhibit high self-similarity over a range of scales should not be considered as salient, such as regions filled with regular textures. Therefore, the saliency measure $A_s(\vec{p})$ is defined as the SEE value of $H_s(\vec{p})$ weighted by the Jensen Shannon divergence between $H_s(\vec{p})$ and $H_{s-\Delta s}(\vec{p})$ by

$$A_s(\vec{p}) = M_\alpha(H_s(\vec{p})) \cdot \mathrm{JSD}(H_s(\vec{p}), H_{s-\Delta s}(\vec{p})), \quad (7)$$

where $\mathrm{JSD}(\cdot)$ denotes the Jenson Shannon divergence [43], $s - \Delta s$ denotes the predecessor scale with respect to scale s. For each pixel p, the most salient scale S_p from which to extract features is calculated by

$$S_{\overrightarrow{p}} = \arg \max_s A_s(\overrightarrow{p}). \quad (8)$$

In this paper, the best possible scales are defined within a range from 4 to 20 pixels, with scale difference parameter s set to 1. Fig. 3 shows a typical example of the top 10 most salient regions with their corresponding scales determined by the proposed saliency measure operator. It can be observed from Fig. 3 that for images belonging to the same person, the most salient regions and their corresponding scales detected are very similar to each other with high degrees of repeatability, while those from a different person are significantly different. There is no additional processing step for the proposed saliency detector such as non-maximum suppression. Therefore, the effectiveness of the proposed saliency measure operator is illustrated.

Once the most salient scale S_p to extract features from pixel p is determined, anatomical features are extracted from a local square region $R_{S_p}(p)$ centered at p with scale S_p. Because of the robustness of the Gabor wavelet and LBP features against local illumination changes and their superior representation power [24], they are used as the anatomical signature for each pixel p within region $R_{S_p}(p)$. Specifically, 40 Gabor filters with five center frequencies and eight orientations as in [13] are used, and only the magnitude of the Gabor filtered responses are used for anatomical features. For the LBP features, we follow the settings in [11], where the uniform pattern histogram with radius 2 and number of neighboring pixel samples 8 is used as the LBP feature, resulting in 59 dimensional LBP features. Therefore, the final anatomical signature for each pixel is a 99 dimensional feature vector (40 Gabor features and 59 LBP features).

It is worth pointing out that the role of the anatomical signature calculated for each pixel in the proposed method is significantly different from the one in local feature matching methods [12], [13]. In local feature matching methods, the extracted local features from different regions will be concatenated to form the final global feature and directly served as input to the classifiers, while in the proposed method, the anatomical signature for each pixel is used to guided the deformable registration process and the spatial relationship between different pixels is preserved.

3.2 The Hierarchical MRF Groupwise Registration Model

After calculating the anatomical signature for each pixel in Section 3.1, we introduce a hierarchical MRF groupwise registration model to construct the group mean facial image space (i.e., the template).

Given n training images $I_1, \ldots, I_n$, we denote their corresponding feature maps obtained in Section 3.1 as $\overrightarrow{F_1}, \ldots, \overrightarrow{F_n}$. For each pixel p, $\overrightarrow{F_i}(p)$ is the 99 dimensional anatomical signature of p in image I_i ($i = 1, \ldots, n$).

Recall that the general groupwise registration process can be expressed by Equation (2). In this paper, the deformable registration process is formulated as a MRF labeling problem defined by

$$J(\hat{I}, \phi_i) = \sum_{i=1}^{n} \left(\varphi(\hat{I}, \phi_i(I_i)) \right)^2 + \lambda Reg(\phi_i))$$

$$= \sum_{i=1}^{n} \left(\sum_{\vec{p} \in \Omega} D_{\vec{p}}(f_{\vec{p}}^i) + \sum_{(\vec{p}, \vec{q}) \in \mathcal{N}_i} V_{\vec{p}, \vec{q}}(f_{\vec{p}}^i, f_{\vec{q}}^i) \right),$$

where Ω denotes the image domain, and $\mathcal{N}_i$ denotes the neighborhood system defined in Ω. In this paper, the four-connected neighborhood system is adopted. $D_{\vec{p}}(f_{\vec{p}}^i)$ is the data term representing the penalty of assigning pixel p in image I_i the label $f_{\vec{p}}^i$, and $V_{\vec{p}, \vec{q}}(f_{\vec{p}}^i, f_{\vec{q}}^i)$ is the smoothness term which penalizes the cost of label discrepancy between two neighboring voxels p and q in image I_i.

The deformable registration problem is converted to a MRF labeling problem by quantizing the deformable transformation φ_i. Specifically, we define a discrete set of labels $\{L|L = \{f^1, f^2, \ldots, f^m\}\}$, where each label f^j ($j = 1, \ldots, m$) is corresponding to one specific displacement vector $\vec{d_j}$. Therefore, assigning a voxel $p \in I_i$ with label $f_{\vec{p}}^i$ means moving pixel p off the original position with displacement vector $\vec{d}_{f_{\vec{p}}^i}$. In this paper, the displacement vector is represented by a bounded discrete window $W = \{0, \pm\tau, \pm2\tau, \ldots, \pm w\tau\}^U$, where U is the dimension of the window, since we are dealing with 2D facial images, $U = 2$. τ is set to 1, and w is set to 12 in this paper. Therefore, the bounded discrete window is $W = \{0, \pm1, \pm2, \ldots, \pm12\}^2$.

The smoothness term $V_{\vec{p}, \vec{q}}(f_{\vec{p}}^i, f_{\vec{q}}^i)$ is defined by the piece-wise truncated absolute distance by

$$V_{\vec{p}, \vec{q}}(f_{\vec{p}}^i, f_{\vec{q}}^i) = \min\left(\lambda, |\vec{d}_{f_{\vec{p}}^i} - \vec{d}_{f_{\vec{q}}^i}|\right), \quad (10)$$

where λ is a constant representing the maximum penalty.

The data term $D_{\vec{p}}(f_{\vec{p}}^i)$ is defined based on the anatomical signature difference between the warped image I_i and the group mean image $\hat{I}$ by

$$D_{\vec{p}}(f_{\vec{p}}^i) = \|\vec{F}_{\hat{I}}(\vec{p}) - \vec{F}_{I_i}(\vec{p} + \vec{d}_{l_p^i})\|_2^2, \quad (11)$$

where $\vec{F}_{\hat{I}}(p)$ denotes the anatomical signature of voxel p in the group mean image $\hat{I}$, and $\vec{F}_{I_i}(\vec{p} + \vec{d}_{l_p^i})$ denotes the anatomical signature of voxel p in image I_i after p is displaced off the original position with displacement vector $\vec{d}_{l_p^i}$.

The optimization of Equation (9) can be achieved by the greedy iterative algorithm similar to [36] to iteratively refine the group mean $\hat{I}$ and the α-expansion algorithm [44] to solve the MRF labeling problem. The optimization procedure can be summarized by Algorithm 2.

Algorithm 2

Estimate group mean $\hat{I}$ and transformation φ_i of each image I_i.

Input: n input images I_i ($i = 1,\ldots, n$).

Output: Group mean image $\hat{I}$, and transformation φ_i to warp each image I_i, to $\hat{I}$.

1. Initialize $\hat{I} = \frac{1}{n} \sum_{i=1}^{n} I_i$.

2. FOR $i = 1$ to n

3. Perform the α-expansion algorithm to estimate the deformable transformation φ_i to warp I_i to $\hat{I}$ with the MRF labeling framework.

4. END FOR

5. Update $\hat{I} = \frac{1}{n} \sum_{i=1}^{n} \phi_i(I_i)$.

6. Repeat Operations 2 to 5 until $\hat{I}$ converges.

7. Return $\hat{I}$ and φ_i.

With Algorithm 2, we can construct the group mean $\hat{I}$ based on input training images I_i. However, directly perform groupwise registration among all training images may be not sufficient to account for the possible large variations across different training facial images. To this end, we propose a hierarchical groupwise registration strategy. The basic principle is that, facial images with similar appearance are clustered into a group, assuming that they have closer distances in the Riemannian manifold represented by deformable transformations. If a group still contains many facial images which may have large variations across each other, it can be further clustered into different smaller groups. Therefore, a pyramid of groups is formed, and the final group mean (i.e., the template) can be constructed in a hierarchical bottom-up manner.

In this paper, affinity propagation (AP) [45] is adopted for clustering, which can automatically and robustly determine the number of cluster centers. The input of AP is the similarity matrix Q, where $Q(i, j)$ denotes the similarity between the ith facial image and the jth facial image. The survival exponential entropy based normalized mutual information (SEE-NMI) is adopted as the similarity function to calculated $Q(i, j)$. SEE-NMI not only shares the robustness properties of the conventional MI [43] against local illumination changes, but also has the advantages of SEE defined in Section 3.1. The SEE-NMI between two facial images I_i and I_j is defined by

$$\text{SEE}-\text{NMI}(I_i, I_j) = \frac{M_\alpha(\mathscr{H}(I_i)) + M_\alpha(\mathscr{H}(I_j))}{M_\alpha(\mathscr{H}(I_i, I_j))}, \quad (12)$$

where $\mathscr{H}(I_i)$, $\mathscr{H}(I_j)$, and $\mathscr{H}(I_i, I_j)$ denote the marginal intensity distribution of I_i, marginal intensity distribution of I_j, and the joint intensity distribution of I_i and I_j, respectively.

Fig. 4 shows a typical example of the pyramid constructed for 10 facial images to estimate the template in a hierarchical manner. It can be observed from Fig. 4 that images with similar appearance are first clustered into the same group to perform groupwise registration.

The final template (i.e., on the top of the pyramid) can be built by the bottom-up hierarchical groupwise registration strategy with Algorithm 2 as the basic building block to perform group-wise registration within each group in the pyramid.

Fig. 5 shows a typical example of the two facial images belonging to the same person but with different facial expressions before and after they are transformed to the template space. It can be observed that after the groupwise registration process, variations caused by different facial expressions can be satisfactorily recovered, which is evidence from the high degree of similarity between the two images after they are transformed to the template space. Their corresponding deformation fields are also visualized in Figs. 5e and 5f, respectively. Therefore, the motivation and effectiveness of the proposed method is illustrated.

After estimating the template $\hat{I}$ and the deformation field φ_i to transform each training image I_i to $\hat{I}$, we can classify a new facial image I_{new} as described by Operations 3 and 4 in Algorithm 1. We estimate the similarity between the transformed new facial image and each training image based on their corresponding feature maps calculated by the procedure described in Section 3.1 to ensure robustness. Specifically, we first calculate the feature map of I_{new}, and denote it as $\overrightarrow{F_{I_{new}}}$. Then, we register I_{new} to the template $\hat{I}$ with the proposed MRF labeling based registration method, and denoting the resulting deformation field as φ_{new}. Finally, we calculate the euclidean distance between $\varphi_{new}(\overrightarrow{F_{I_{new}}})$ and $\varphi_i(\overrightarrow{F_{I_i}})$ ($i = 1, \ldots, n$), where $\overrightarrow{F_{I_i}}$ is the feature map of training image I_i. The label of I_{new} is determined as the one of the training image which has the most similar transformed feature map to the new facial image.

It should be noted that a possible alternative strategy to construct the group mean space is to first construct the group mean image of each subject, and then perform group-wise registration among the subject-specific group mean images to obtain the final group mean image. This strategy is also similar to the hierarchical registration strategy adopted in this paper, where the final group mean space is estimated in a bottom-up manner from the constructed group means of similar facial images.

It should also be noted that the proposed method requires a training set of facial images to construct the template space. However, the proposed method does not require the subject identity information (i.e., subject labels) during the training stage. Therefore, it is an unsupervised learning method.

4 Experimental Results

Extensive experiments have been conducted to evaluate the proposed method on four face recognition benchmark databases: FERET [46], CAS-PEAL-R1 [47], FRGC ver 2.0 [48], and LFW [56]. The proposed method has also been compared with different state-of-the-art face recognition methods on the three databases. In all experiments, the order α of the survival exponential entropy was set to 3, and the parameter λ controlling the maximum penalty of the smoothness term in Equation (10) was set to 15 by cross validation. The nearest neighbor classifier was adopted in all experiments.

4.1 Experimental Results on the FERET Database

In the FERET database [46], the gallery set (i.e., Fa) contains 1, 196 frontal images of 1, 196 subjects. There are four probe sets in the standard FERET evaluation protocol: The Fb probe set contains 1, 195 images of different facial expression variations, Fc contains 194 images taken under different illumination conditions, Dup I has 722 images taken later in time with aging variations of 243 subjects, and Dup II is a subset of Dup I with 234 images of 75 subjects which were taken at least one year after the corresponding facial images in the gallery set. Sample images of the FERET database are shown in Fig. 6.

For each image, the following preprocessing steps were performed. Each image was cropped and normalized to the resolution of 128×128 based on the manually located eyes positions provided by the FERET database to contain only the facial region. Then, histogram equalization was performed to reduce the possible distortions caused by illumination changes.

In order to evaluate the effectiveness of the proposed hierarchical groupwise MRF based registration strategy, we have also included the results obtained by the pairwise registration strategy in our preliminary work in [33] and the results obtained by two state-of-the-art deformable registration algorithms, namely the diffeomorphic Demons (D-Demons) [49] based on the optical flow equation and the fast free form deformation (FFD) model [50] with B-spline basis functions. Table 1 lists the rank-1 recognition rates of different approaches on all four probe sets of the FERET database. Results obtained by our method without using the hierarchical registration strategy are also reported. For all the other methods under comparison, the recognition rates are directly cited from their corresponding papers, as it is assumed that the recognition rates reported in the corresponding papers are under their optimal parameter settings.

It can be observed from Table 1 that the proposed method consistently achieves the highest recognition rate among all methods under comparison on the four probe sets. Therefore, the discriminant power and robustness of the proposed method against facial expression variations, illumination changes, and aging variations are implied. It can also be observed that the recognition performance of the proposed method significantly degrades without using the hierarchical registration strategy. Therefore, the importance of mitigating the effects of large inter-person face variations at the beginning stage of the registration process is reflected. The hierarchical registration strategy reduces the risk of being stuck at suboptimal solutions for estimating the template space. Moreover, it is shown that the the recognition rates obtained with the groupwise registration scheme are consistently higher than those obtained by the pairwise registration strategy (i.e., even without using the hierarchical registration strategy). Therefore, the importance of estimating the template space in an unbiased manner with the groupwise registration scheme instead of explicitly selecting one facial image as the template is also reflected.

It should be noted that the maximum number of images allowed within each group during hierarchical groupwise registration also affect the performance of the proposed method. If the maximum number of images allowed within each group is too large, then images with large variations may be clustered into the same group and lead to inaccurate registration

results. On the other hand. If this value is too small, the robustness of our method will be reduced as images from the same person maybe clustered into different groups. Fig. 7 shows the rank-1 recognition rates of our method with respect to different maximum numbers of images allowed within each group on the four probe sets. Moreover, to illustrate the importance of incorporating local anatomical features to guide the registration instead of using pixel intensity information alone, in Fig. 7 the rank-1 recognition rates of the proposed method using only the pixel intensity as feature to guide the registration process are also given.

Fig. 7 illustrates that our method achieves its best recognition accuracy when the maximum number of images within each group is around 20. This value was set to 20 in this paper for all experiments. The recognition performance for different databases may be further improved with different values of maximum number of images in each group. In this paper, it is empirically fixed to 20 as the input query images and databases are unknown beforehand in real world applications. The recognition rate for using pixel intensity alone as feature to guide the registration process is significantly lower than the one using salient local anatomical features, especially on the Fc probe set with respect to illumination changes.

The running time required to constructed the template space by the proposed groupwise registration strategy is around 17.6 minutes, and the average time required to recognize each new query image in the recognition phase is around 15 seconds for the FERET database on a computer with Intel Xeon 2.66-GHz CPU. It should be noted that the template construction stage can be completed offline, which does not affect the efficiency in the recognition stage. The computational efficiency of the proposed method can be further improved by parallel processing and code optimization.

4.2 Experimental Results on the CAS-PEAL-R1 Database

The CAS-PEAL-R1 database [47] contains 30, 863 images of 1,040 subjects, among which 595 are males and 445 are females. The CAS-PEAL-R1 standard evaluation protocol contains a gallery set consisting of 1,040 images of 1,040 subjects taken under the normal condition, a training set consisting of 1,200 images of 300 subjects for building the recognition model or tuning the parameters of a model, and six frontal probe sets contain facial images with the following types of variations: expression, accessory, lighting, aging, background, and distance.

For each image, similar preprocessing steps to the FERET database in Section 4.1 were performed, where each image was cropped and normalized to the resolution of 128×128 based on the eyes positions provided by the CAS-PEAL-R1 database, and histogram equalization was performed. Samples images from the CAS-PEAL-R1 database are shown in Fig. 8.

Table 2 lists the rank-1 recognition rates obtained by different approaches on the CAS-PEAL-R1 database. For comparison purpose, recognition rates of using the diffeomorphic Demons [49] (D-Demons) algorithm, fast free form deformation model [50], and the pairwise MRF registration strategy [33] are also included. The recognition rates of all the other methods under comparison are directly cited from their corresponding papers.

Experimental results obtained by our method without using the hierarchical registration strategy are also included. It can be observed from Table 2 that our method achieves the highest overall recognition rates. It is shown in Table 2 that without using the hierarchical registration strategy, the recognition performance of our method degrades. Also, higher recognition rates are obtained by using the group-wise registration scheme than using the pairwise registration strategy.

Similar to Section 4.1, we also compare the proposed method which uses salient anatomical features to guide the registration process to the one using pixel intensity information alone to drive the registration process, and the corresponding rank-1 recognition rates are shown in Fig. 9.

It can be observed from Fig. 9 that the proposed method which uses salient local anatomical feature to drive registration consistently outperforms the intensity guided registration strategy, especially under the illumination change condition. Therefore, the advantage of using feature guided registration strategy is strongly implied.

The running time required to construct the template space is around 54.8 minutes, and the average time required to recognize each new query image in the recognition phase is around 16 seconds for the CAS-PEAL-R1 database on a computer with Intel Xeon 2.66-GHz CPU.

4.3 Experimental Results on the FRGC ver 2.0 Database

The proposed method was also evaluated on the FRGC ver 2.0 database. The FRGC ver 2.0 database [48] is known as one of the largest face image data sets available. Sample images of the FRGC ver 2.0 database are shown in Fig. 10. Similar to the experimental settings in Sections 4.1 and 4.2, each image was normalized and cropped to the size of 128×128, and histogram equalization was performed.

The FRGC ver 2.0 database contains 12, 776 images of 222 subjects in the training set, with 6,360 images were taken under the controlled condition and 6,416 images were taken under the uncontrolled condition. It also contains 16,028 target images taken under the controlled illumination condition. In this paper, the experiment 1 and experiment 4 standard protocols of the FRGC ver 2.0 database were used to evaluate the proposed method. For the experiment 1 protocol, there are 16,028 query images taken under the controlled illumination condition. For the experiment 4 protocol, there are 8,014 query images taken under the uncontrolled illumination condition, which is the most challenging protocol in FRGC.

The recognition performance of different methods on the FRGC ver 2.0 database was measured by the receiving operator characteristics (ROC), which is the face verification rate (FVR) versus the false accept rate (FAR). For both the experiment 1 and experiment 4 protocols, there are three ROC value: ROC 1 corresponding to images collected within semester, ROC 2 corresponding to images collected within year and ROC 3 corresponding to images collected between semesters. Table 3 lists the FVR at FAR of 0.1 percent with respect to the experiment 1 protocol of different methods. The recognition rates of LGBP)

LGXP are directly cited from reference [14], and the recognition rates of LGBPHS, LBP, and the BEE Baseline are directly cited from reference [33].

It can be observed from Table 3 that the proposed method consistently achieves the highest FVR values at FAR = 0.1% for all three ROC values among other methods under comparison. Specifically, comparing to the BEE baseline algorithm (i.e., PCA on large scale data), the proposed method achieves more than 20 percent of the FVR improvement on all the three ROC values. Moreover, the advantage of using the hierarchical registration strategy is illustrated.

Table 4 lists the FVR value at FAR of 0.1 percent with respect to the experiment 4 protocol of different methods, which is the most challenging protocol in FRGC.

It can be observed from Table 4 that our method achieves significantly higher recognition rates than other methods under comparison. Therefore, the robustness of the proposed hierarchical groupwise MRF based registration strategy is demonstrated.

More comparisons between our method and other state-of-the-art face recognition approaches are shown in Table 5 of the ROC 3 values for both the experiment 1 and experiment 4 protocols. One of the best results reported in the literature for the FRGC ver 2.0 experiment 4 protocol is in reference [55], with 92.43 percent recognition rate. The recognition performance of the proposed method can also be improved by integrating more advanced and novel image features such as those used in reference [55].

The running time required to constructed the template space by the proposed groupwise registration strategy is around 32.4 minutes, and the average time required to recognize each new query image in the recognition phase is around 8 seconds for the FRGC ver 2.0 database on a computer with Intel Xeon 2.66-GHz CPU.

4.4 Experimental Results on the LFW Database

Our method was also evaluated on the labeled faces in the wild (LFW) [56] database. It contains 13, 233 facial images of 5,749 different persons. The facial images were taken from unconstrained environments with large variations in pose, facial expression, background, and lighting conditions. In this paper, the 'View 2' set in the LFW database (i.e., the set designed for final testing and benchmarking [56]) was used for evaluation, where our method was evaluated in the 10 fold cross validation manner similar in [57]. We followed the standard image restricted configuration setup in [56], where only the match/non-match information is available for each pair of training images. Moreover, there is no outside training data used other than those provided by the LFW database. Fig. 11 shows some typical examples of the matched and non-matched image pairs from the LFW database.

During the training phase, we first estimate the template space with the proposed hierarchical groupwise registration method based on the training image pairs in the available nine splits (i.e., 'View 2' was evaluated in a 10 fold cross validation manner). Then, based on the registered training image pairs to the template space, we determine the best matching threshold, and apply it to the remaining 10th split. The experiment was repeated 10 times.

Table 6 lists the estimated mean accuracies and the standard error of the mean obtained by our method and those reported by other competing methods on the LFW result website.[1] For fair comparison purpose, most of the competing methods adopted for comparison had the same setting in this paper: They were tested under the strictest LFW protocol by their authors, where the image restricted configuration setup was used and there was no outside training data used other than those provided by LFW. In order to investigate the contribution of the proposed SEE-based salient detector, we also compared the face verification performance of our method using the SEE-based salient detector and the difference of Gaussian (DoG) salient detector used in SIFT [40], which is one of the best detectors stated in [58].

It can be observed from Table 6 that our method achieves the highest verification performance among all the compared methods, which reflects the effectiveness of our method for face verification under the unconstrained environments. Moreover, the verification rate obtained by our method using the SEE-based salient detector is higher than the one using the DoG salient detector, which illustrates the advantage of using the proposed SEE-based salient detector. It is also observed that using the groupwise registration scheme and the hierarchical registration strategy clearly outperforms the pairwise registration strategy.

Fig. 12 shows the verification performance of different methods with the ROC curves, and it can be observed that our method has the best verification performance among all the methods under comparison. It matches with the observations in Table 6.

To investigate how the order α of the survival exponential entropy in Equation (4) influence the recognition performance, the mean face verification rates with different values of α for the LFW database are shown in Fig. 13.

It can be observed from Fig. 11 that when $\alpha = 3$, our method gives satisfactory verification accuracy. When α is too small, the SEE-based saliency measure in Equation (7) may not be effective to capture the salient regions in facial images. On the other hand, when α is too large, the SEE-based saliency measure may be too sensitive and not robust.

The average running time to determine whether a pair of query image is matched or non-matched is around 32 seconds on a computer with Intel Xeon 2.66-GHz CPU.

5 Conclusion

In this paper, we formulate the face recognition problem as a groupwise registration and feature matching problem. A robust salient scale detector based on the survival exponential entropy is proposed to extract the anatomical features from the most salient scales. The deformable transformation space is discretized and represented by the Markov random field labeling framework, which is integrated with the salient anatomical signature of each pixel to drive the registration process. In order to deal with possible large variations between different facial images, a hierarchical groupwise registration strategy is proposed. During the

---

[1]http://vis-www.cs.umass.edu/lfw/results.html.

recognition phase, each query image is transformed to the template space and compared with the existing training images. Our method has been extensively evaluated on four benchmark databases: FERET, CAS-PEAL-R1, FRGC ver 2.0, and the LFW databases. It is also compared with several state-of-the-art face recognition methods. Experimental results demonstrate that our method achieves the highest recognition and verification rates among other methods under comparison, which demonstrates the effectiveness of our method.

References

1. Tian T, Kanade T, Cohn J. Recognizing Action Units for Facial Expression Analysis. IEEE Trans Pattern Analysis and Machine Intelligence. Feb; 2001 23(2):97–115.

2. Fasel B, Luettin J. Automatic Facial Expression Analysis: A Survey. Pattern Recognition. 2003; 36(1):259–275.

3. Shashua A, Riklin-Raviv T. The Quotient Image: Class Based Re-Rendering and Recognition with Varying Illuminations. IEEE Trans Pattern Analysis and Machine Intelligence. Feb; 2001 23(2): 129–139.

4. Xie X, Lam K. Face Recognition under Varying Illumination Based on a 2D Face Shape Model. Pattern Recognition. 2005; 38(2):221–230.

5. Pan Z, Healey G, Prasad M, Tromberg B. Face Recognition in Hyperspectral Images. IEEE Trans Pattern Analysis and Machine Intelligence. Dec; 2003 25(12):1552–1560.

6. Blanz V, Vetter T. Face Recognition Based on Fitting a 3D Morphable Model. IEEE Trans Pattern Analysis and Machine Intelligence. Sep; 2003 25(9):1063–1074.

7. Turk M, Pentland A. Eigenfaces for Recognition. J Cognitive Neuroscience. 1991; 3:71–86.

8. Belhumeur P, Hespanha J, Kriegman D. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Trans Pattern Analysis and Machine Intelligence. Jul; 1997 19(7): 711–720.

9. Zhu, M.; Martinez, A. Selecting Principal Components in a Two-Stage LDA Algorithm. Proc. IEEE Conf. Computer Vision and Pattern Recognition; 2006. p. 132-137.

10. Zhu M, Martinez A. Subclass Discriminant Analysis. IEEE Trans Pattern Analysis and Machine Intelligence. Aug; 2006 28(8):1274–1286.

11. Ahonen T, Hadid A, Pietikainen M. Face Description with Local Binary Patterns: Application to Face Recognition. IEEE Trans Pattern Analysis and Machine Intelligence. Dec; 2006 28(12): 2037–2041.

12. Zhang, W.; Shan, S.; Gao, W.; Chen, X.; Zhang, H. Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A Novel Non-Statistical Model for Face Representation and Recognition. Proc. IEEE Conf. Computer Vision; 2005. p. 786-791.

13. Lei Z, Liao S, Pietikainen M, Li S. Face Recognition by Exploring Information Jointly in Space, Scale and Orientation. IEEE Trans Image Processing. Jan; 2011 20(1):247–256.

14. Xie S, Shan S, Chen X, Chen J. Fusing Local Binary Patterns of Gabor Magnitude and Phase for Face Recognition. IEEE Trans Image Processing. May; 2010 19(5):1349–1361.

15. Bartlett M, Movellan J, Sejnowski T. Face Recognition by Independent Component Analysis. IEEE Trans Neural Network. Nov; 2002 13(6):1450–1464.

16. Scholkopf B, Smola A, Muller K. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. Neural Computing. 1999; 10:1299–1319.

17. Liu C. Capitalize on Dimensionality Increasing Techniques for Improving Face Recognition Grand Challenge Performance. IEEE Trans Pattern Analysis and Machine Intelligence. May; 2006 28(5): 725–737.

18. Yan S, Xu D, Zhang B, Zhang H, Yang Q, Lin S. Graph Embedding and Extensions: A General Framework for Dimensionality Reduction. IEEE Trans Pattern Analysis and Machine Intelligence. Jan; 2007 29(1):40–51.

19. He X, Yan S, Hu Y, Niyogi P, Zhang H. Face Recognition Using Laplacianfaces. IEEE Trans Pattern Analysis and Machine Intelligence. Mar; 2005 27(3):328–340.

20. Tenenbaum J, Silva V, Langford J. A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science. 2000; 290(22):2319–2323. [PubMed: 11125149]

21. Roweis S, Saul L. Nonlinear Dimensionality Reduction by Locally Linear Embedding. Science. 2000; 290(22):2323–2326. [PubMed: 11125150]

22. Wright J, Yang A, Ganesh A, Sastry S, Ma Y. Robust Face Recognition via Sparse Representation. IEEE Trans Pattern Analysis and Machine Intelligence. Feb; 2009 31(2):210–227.

23. Naseem I, Togneri R, Bennamoun M. Linear Regression for Face Recognition. IEEE Trans Pattern Analysis and Machine Intelligence. Nov; 2010 32(11):2106–2112.

24. Zou J, Ji Q, Nagy G. A Comparative Study of Local Matching Approach for Face Recognition. IEEE Trans Image Processing. Oct; 2007 16(10):2617–2628.

25. Wiskott L, Fellous J, Kruger N, Malsburg C. Face Recognition by Elastic Bunch Graph Matching. IEEE Trans Pattern Analysis and Machine Intelligence. Jul; 1997 19(7):775–779.

26. Ojala T, Pietikainen M, Maenpaa T. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Trans Pattern Analysis and Machine Intelligence. Jul; 2002 24(7):971–987.

27. Guo Y, Zhao G, Pietikainen M. Discriminative Features for Texture Description. Pattern Recognition. 2012; 45(10):3834–3843.

28. Yousefi, S.; Minh, P.; Kehtarnavaz, N.; Yan, C. Facial Expression Recognition Based on Diffeomorphic Matching. Proc. Int'l Conf. Image Processing; 2010. p. 4549-4552.

29. Guo, Y.; Zhao, G.; Pietikainen, M. Dynamic Facial Expression Recognition Using Longitudinal Facial Expression Atlases. Proc. European Conf. Computer Vision; 2012. p. 631-644.

30. Koelstra S, Pantic M, Patras I. A Dynamic Texture-Based Approach to Recognition of Facial Actions and Their Temporal Models. IEEE Trans Pattern Analysis and Machine Intelligence. Nov; 2010 32(11):1940–1954.

31. Lee M, Ranganath S. Pose-Invariant Face Recognition Using a 3D Deformable Model. Pattern Recognition. 2003; 36(8):1835–1846.

32. Asthana, A.; Marks, T.; Jones, M.; Tieu, K.; Rohith, M. Fully Automatic Pose-Invariant Face Recognition via 3D Pose Normalization. Proc. IEEE Conf. Computer Vision; 2011. p. 937-944.

33. Liao, S.; Chung, A. A Novel Markov Random Field Based Deformable Model for Face Recognition. Proc. IEEE Conf. Computer Vision and Pattern Recognition; 2010. p. 2675-2682.

34. Zhan, Y.; Feldman, M.; Tomaszeweski, J.; Davatzikos, C.; Shen, D. Registering Histological and MR Images of Prostate for Image-Based Cancer Detection. Proc. Int'l Conf. Medical Image Computing and Computer-Assisted Intervention; 2006. p. 620-628.

35. Shen D, Wong W, Ip H. Affine-Invariant Image Retrieval by Correspondence Matching of Shapes. Image and Vision Computing. 1999; 17:489–499.

36. Joshi S, Davis B, Jomier M, Gerig G. Unbiased Diffeomorphic Atlas Construction for Computational Anatomy. Neuro-Image. 2004; 23:151–160.

37. Jia H, Wu G, Wang Q, Shen D. ABSORB: Atlas Building by Self-Organized Registration and Bundling. NeuroImage. 2010; 51:1057–1070. [PubMed: 20226255]

38. Cootes T, Twining C, Petrovic V, Babalola K, Taylor C. Computing Accurate Correspondences across Groups of Images. IEEE Trans Pattern Analysis and Machine Intelligence. Nov; 2010 32(11):1994–2005.

39. Balci, S.; Golland, P.; Shenton, M.; Wells, W. Free-Form B-Spline Deformation Model for Groupwise Registration. Proc. MICCAI 2007 Statistical Registration Workshop: Pair-Wise and Group-Wise Alignment and Atlas Formation; 2007. p. 23-30.

40. Lowe D. Distinctive Image Features from Scale-Invariant Key-points. Int'l J Computer Vision. 2004; 60(2):91–110.

41. Zografos K, Nadarajah S. Survival Exponential Entropies. IEEE Trans Information Theory. Mar; 2005 51(3):1239–1246.

42. Liao, S.; Chung, A. Multi-Modal Image Registration Using the Generalized Survival Exponential Entropy. Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention; 2006. p. 964-971.

43. Thomas, M.; Thomas, J. Elements of Information Theory. Wiley-Interscience; 1991.

44. Yuri B, Olga V, Ramin Z. Fast Approximate Energy Minimization via Graph Cuts. IEEE Trans Pattern Analysis and Machine Intelligence. Nov; 2001 23(11):1222–1239.

45. Frey B, Dueck D. Clustering by Passing Messages between Data Points. Science. 2007; 315:972–976. [PubMed: 17218491]

46. Phillips P, Moon H, Rizvi A, Rauss P. The FERET Evaluation Methodology for Face Recognition Algorithms. IEEE Trans Pattern Analysis and Machine Intelligence. Oct; 2000 22(10):1090–1104.

47. Gao W, Cao B, Shan S, Chen X, Zhou D, Zhang X, Zhao D. The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations. IEEE Trans Systems Man, and Cybernetics—Part A: Systems and Humans. Jan; 2008 38(1):149–161.

48. Phillips, P.; Flynn, P.; Scruggs, T.; Bowyer, K.; Chang, J.; Hoffman, K.; Marques, J.; Min, J.; Worek, W. Overview of the Face Recognition Grand Challenge. Proc. IEEE Conf. Computer Vision and Pattern Recognition; 2005. p. 947-954.

49. Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic Demons: Efficient Non-Parametric Image Registration. NeuroImage. 2009; 45:61–72.

50. Rueckert D, Sonoda L, Hayes C, Hill D, Leach M, Hawkes D. Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images. IEEE Trans Medical Imaging. Aug; 1999 18(8):712–721.

51. Ravela, S.; Manmatha, R. Retrieving Images by Appearance. Proc. IEEE Conf. Computer Vision; 1998. p. 608-613.

52. Hwang, W.; Park, G.; Lee, J. Multiple Face Model of Hybrid Fourier Feature for Large Face Image Set. Proc. IEEE Conf. Computer Vision and Pattern Recognition; 2006. p. 1574-1581.

53. Su Y, Shan S, Chen X, Gao W. Hierarchical Ensemble of Global and Local Classifiers for Face Recognition. IEEE Trans Image Processing. Aug; 2009 18(8):1885–1896.

54. Tan X, Triggs B. Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions. IEEE Trans Image Processing. Jun; 2010 19(6):1635–1650.

55. Liu Z, Liu C. Fusion of Color, Local Spatial and Global Frequency Information for Face Recognition. Pattern Recognition. 2010; 43(8):2882–2890.

56. Huang, G.; Ramesh, M.; Berg, T.; Learned-Miller, E. Technical Report 07–49. Univ. of Massachusetts; Amherst: Oct. 2007 Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments.

57. Hussain, S.; Napoleon, T.; Jurie, F. Face Recognition Using Local Quantized Patterns. Proc. British Machine Vision Conf; 2012. p. 1-11.

58. Mikolajczyk K, Schmid C. A Performance Evaluation of Local Descriptors. IEEE Trans Pattern Analysis and Machine Intelligence. Oct; 2005 27(10):1615–1630.

59. Nowak, E.; Jurie, F. Learning Visual Similarity Measures for Comparing Never Seen Objects. Proc. IEEE Conf. Computer Vision and Pattern Recognition; 2007. p. 1-8.

60. Wolf, L.; Hassner, T.; Taigman, Y. Descriptor Based Methods in the Wild. Proc. Faces in Real-Life Images Workshop in European Conf. Computer Vision; 2008.

61. Seo H, Milanfar P. Face Verification Using the Lark Representation. IEEE Trans Information Forensics and Security. Dec; 2011 6(4):1275–1286.

62. Pinto, N.; DiCarlo, J.; Cox, D. How Far Can You Get with a Modern Face Recognition Test Set Using Only Simple Features?. Proc. IEEE Conf. Computer Vision and Pattern Recognition; 2009. p. 2591-2598.
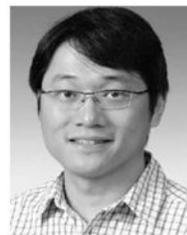
Biographies

Shu Liao received the BEng degree (First Class Honors and Academic Achievement Awards) in computer engineering and the MPhil and PhD degrees in computer science and engineering from The Hong Kong University of Science and Technology in 2005, 2007, and 2010, respectively. He is currently a postdoctoral research fellow in the Department of Radiology, Biomedical Research Imaging Center, University of North Carolina at Chapel Hill. His research interests include medical image analysis, texture analysis, and face recognition.

Dinggang Shen is a professor of Radiology, Biomedical Research Imaging Center (BRIC), Computer Science, and Biomedical Engineering, University of North Carolina at Chapel Hill (UNC-CH). He is currently directing the Image Display, Enhancement, and Analysis (IDEA) Lab in the Department of Radiology, and also the medical image analysis core in the BRIC. He was a tenure-track assistant professor at the University of Pennsylvania (UPen), and a faculty member at the Johns Hopkins University. His research interests include medical image analysis, computer vision, and pattern recognition. He has published more than 300 papers in the international journals and conference proceedings. He serves as an editorial board member for four international journals.


Albert C.S. Chung received the BEng degree (First Class Honors) in computer engineering from The University of Hong Kong in 1995 and the MPhil degree in computer science from The Hong Kong University of Science and Technology in 1998. He joined the Medical

Vision Laboratory, University of Oxford, United Kingdom, as a doctoral research student with a Croucher Foundation scholarship and graduated in 2001. He was a visiting scientist at the Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, in 2001. He is currently an associate professor with the Department of Computer Science and Engineering, The Hong Kong University of Science and Technology. His research interests include medical image analysis, image processing, and computer vision. He received the 2002 British Machine Vision Association Sullivan Thesis Award for the best doctoral thesis submitted to a United Kingdom university in the field of computer or natural vision.
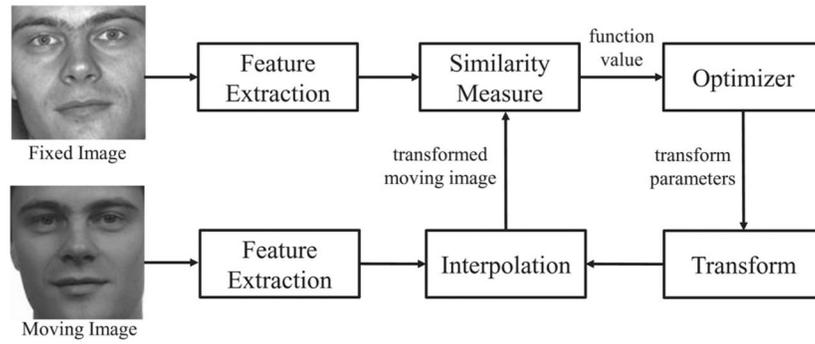
Fig. 1.
The schematic illustration of the pairwise registration process.
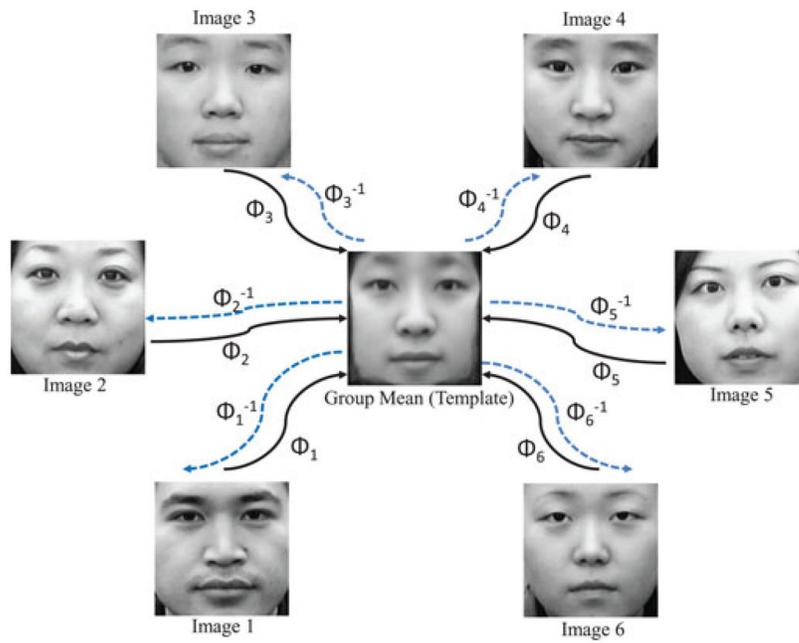
Fig. 2.
The schematic illustration of groupwise registration. The template (group mean) is estimated which has the smallest geodesic distances among the six input images on the Riemannian manifold. Each image I_i can be transformed to the template space with transformation φ_i (black solid arrows). The template can also be warped to each individual image's space by the backward transformation ϕ_i^{-1} (blue dashed arrows).
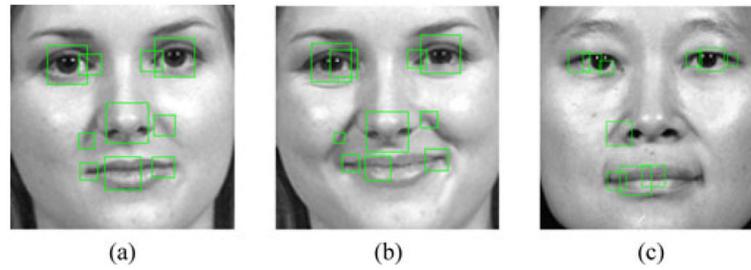
Fig. 3.
Typical example of the top 10 salient regions detected by the proposed saliency measure operator on three facial images of the FERET database. (a) and (b) are facial images belonging to the same person, while (c) is a facial image of a different person. The most salient regions are highlighted by the green squares, and the side length of each square denotes the most salient scale of the corresponding region to extract features.
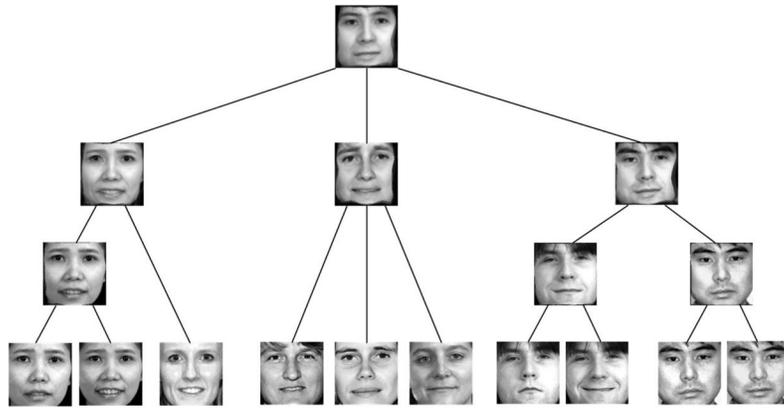
Fig. 4.
A typical example of the pyramid constructed for 10 facial images from the FERET database to build the template with hierarchical groupwise registration. The original facial images are at the bottom level. Note that all the images will be compared to the same template space (i.e., the top level of the pyramid).
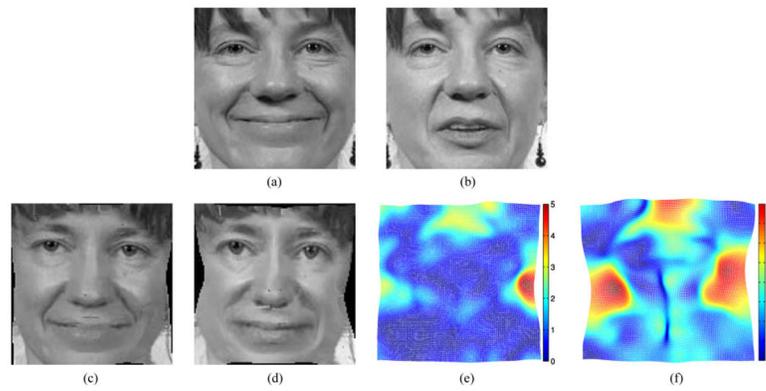
Fig. 5.
(a) and (b) are two original facial images belonging to the same person but with different facial expressions from the FERET database. (c) and (d) are the corresponding results after transforming (a) and (b) to the template space, respectively. (e) and (f) are the corresponding deformation fields to transform (a) and (b) to the template space, respectively. It can be observed that the variations caused by facial expressions have been recovered.
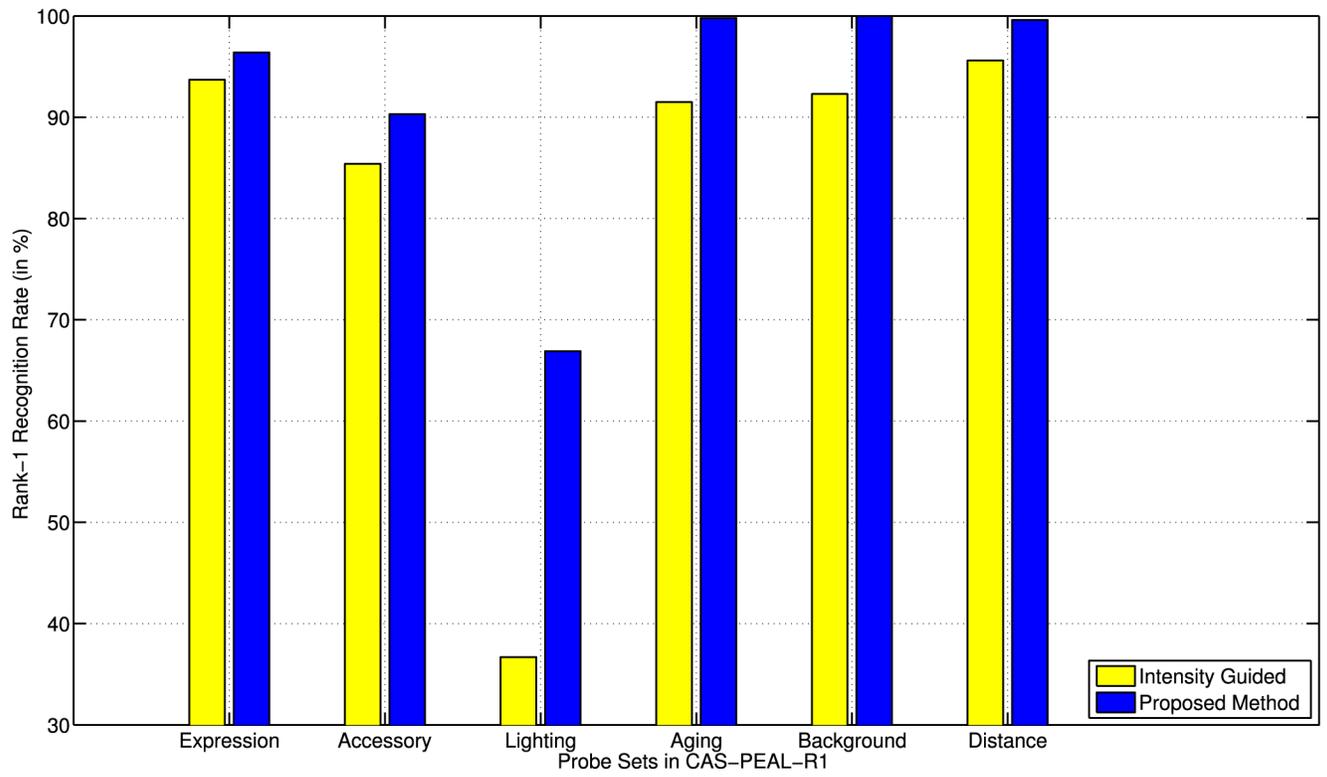
Fig. 6.
Sample images from the FERET database.
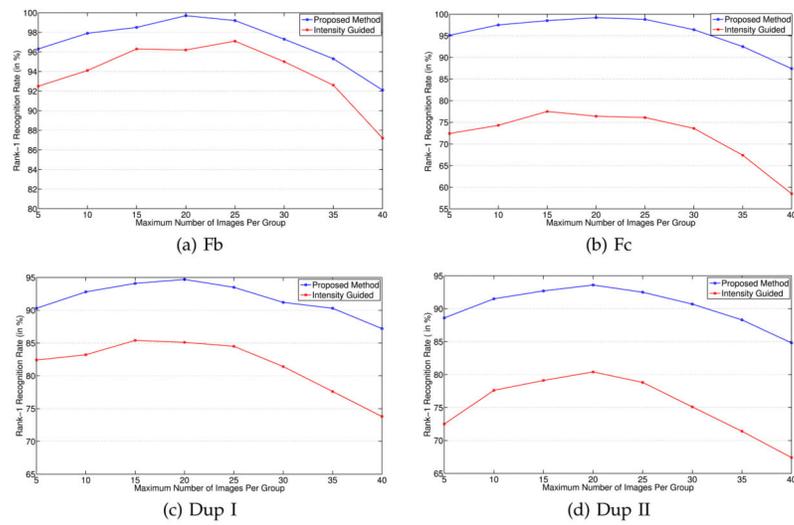
Fig. 7.
Rank-1 recognition rates of our method with respect to the maximum number of images within each group on the FERET database. For comparison purpose, the rank-1 recognition rates of using only the pixel intensity as feature to guide registration are also included.
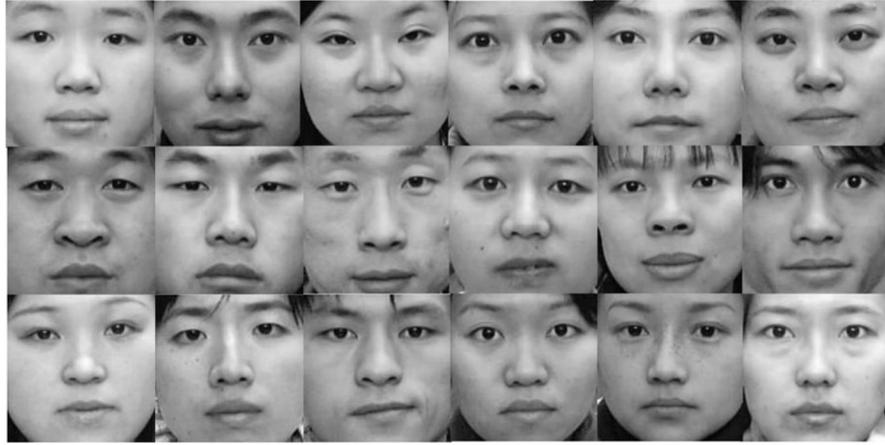
Fig. 8.
Sample images from the CAS-PEAL-R1 database.

Fig. 9.
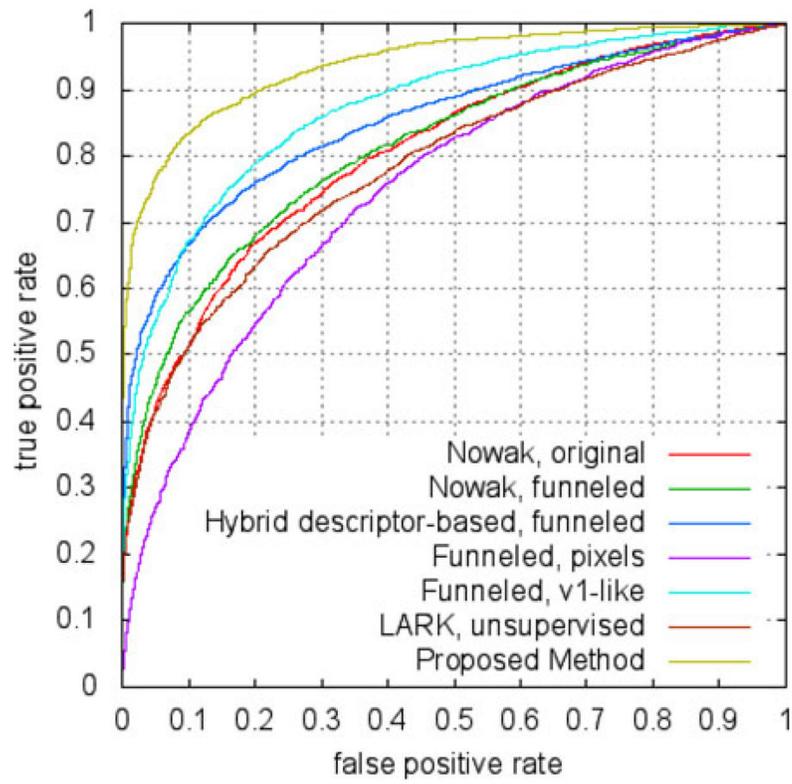Rank-1 recognition rates of the six probe sets on the CAS-PEAL-R1 database by using the proposed method and using pixel intensity guided registration strategy.
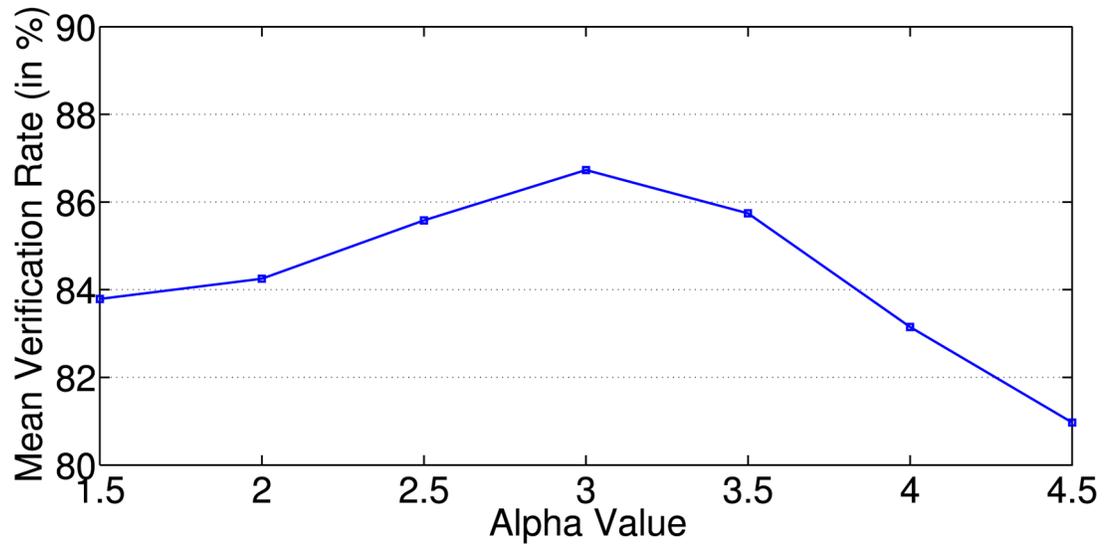
Fig. 10.
Sample images from the FRGC ver 2.0 database.

Fig. 11.
Sample matched and non-matched image pairs from the LFW database.

Fig. 12.
The ROC curve comparisons between the proposed method and other state-of-the-art methods on the LFW database.

Fig. 13.
The mean face verification rates with different values of α in Equation (4) on the LFW database.

TABLE 1

The Rank-1 Recognition Rates (in Percent) of Different Approaches on Four Probe Sets of the FERET Database

Methods	Fb	Fc	Dup I	Dup II
1. H-Groupwise MRF	**99.7**	**99.2**	**94.7**	**93.6**
2. Groupwise MRF	98.5	98.8	87.7	86.2
3. Pairwise MRF [33]	98.2	98.8	83.2	79.4
4. D-Demons [49]	95.8	78.4	71.5	68.3
5. FFD [50]	93.2	72.3	66.4	67.8
6. E-GV-LBP [13]	98.4	99.0	82.0	81.6
7. LGBP + LGXP [14]	99.0	99.0	94.0	93.0
8. LGBPHS [12]	98.0	97.0	74.0	71.0
9. LBP [11]	97.0	79.0	66.0	64.0
10. PCA [7]	85.0	65.0	44.0	22.0
11. UMD LDA [51]	96.2	58.8	47.2	20.9
12. USC EBGM [51]	95.0	82.0	59.1	52.1
13. Bayesian, MAP [46]	82.0	37.0	52.0	32.0

H-groupwise MRF and groupwise MRF denote our method with and without using the hierarchical registration strategy, respectively, and pairwise MRF denotes the pairwise registration strategy in our preliminary work in [33]. Rows 4 and 5 show the recognition rates obtained by using the diffeomorphic Demons [49] and fast free form deformation (FFD) [50] pairwise registration strategies, respectively. For each column, the highest recognition rate is bolded.

TABLE 2

The Rank-1 Recognition Rates (in Percent) of Different Approaches on Six Probe Sets of the CAS-PEAL-R1 Database

Methods	Expression	Accessory	Lighting	Aging	Background	Distance	All
1. H-Groupwise MRF	**96.4**	**90.3**	**66.9**	99.8	**100.0**	99.6	**85.4**
2. Groupwise MRF	94.8	90.1	63.7	99.2	98.8	99.5	83.6
3. Pairwise MRF [33]	94.3	88.4	62.7	98.6	98.3	99.2	82.8
4. D-Demons [49]	90.2	84.3	48.4	96.5	94.3	96.2	75.2
5. FFD [50]	85.4	82.8	43.7	93.3	94.5	93.7	72.1
3. LGBPHS [47]	95.2	86.8	51.0	**100.0**	98.7	98.9	78.7
4. Gabor + PCA + LDA [47]	90.6	82.8	44.8	98.5	98.0	**100.0**	74.4
5. PCA + LDA [47]	71.3	61.0	21.8	72.7	94.4	93.5	54.8
6. PCA [47]	53.7	37.1	8.2	50.0	80.5	74.2	36.6

H-groupwise MRF and groupwise MRF denote the proposed method with and without using the hierarchical registration strategy, respectively. The last column reports the recognition rate of the union of the six probe sets (i.e., All) for each method. For each column, the highest recognition rate is bolded.

TABLE 3

The FVR Value of Different Approaches at FAR = 0.1% in Experiment 1 of the FRGC Ver 2.0 Database, Where H-Groupwise MRF and Groupwise MRF Denote Our Method with and without Using the Hierarchical Registration Strategy, Respectively

	FVR at FAR = 0.1% (in %)		
Methods	ROC 1	ROC 2	ROC 3
1. H-Groupwise MRF	**99.3**	**98.6**	**98.2**
2. Groupwise MRF	97.9	96.3	94.9
3. Pairwise MRF [33]	97.5	95.9	92.6
4. D-Demons [49]	91.4	88.7	83.5
5. FFD [50]	87.3	85.0	79.5
6. LGBP + LGXP [14]	98.7	98.1	97.5
7. LGBPHS [33]	92.8	91.2	87.9
8. LBP [33]	86.1	83.3	79.2
9. BEE Baseline [33]	77.6	75.2	70.6

For each column, the highest FVR value is bolded.

TABLE 4

The FVR Value of Different Approaches at FAR = 0.1% in Experiment 4 of the FRGC Ver 2.0 Database, Where H-Groupwise MRF and Groupwise MRF Denote Our Method with and without Using the Hierarchical Registration Strategy, Respectively

Methods	FVR at FAR = 0.1% (in %)		
	ROC 1	ROC 2	ROC 3
1. H-Groupwise MRF	**89.7**	**90.8**	**90.3**
2. Groupwise MRF	79.5	78.7	78.8
3. Pairwise MRF [33]	74.2	73.4	71.8
4. D-Demons [49]	36.5	31.4	27.2
5. FFD [50]	28.1	27.4	24.8
3. LGBP + LGXP [14]	83.6	84.3	84.9
4. LGBPHS [33]	31.1	28.8	23.9
5. LBP [33]	26.8	22.6	19.1
6. BEE Baseline [33]	17.1	15.1	13.5

For each column, the highest FVR value is bolded.

TABLE 5

Comparisons with More State-of-the-Art Methods on the FRGC Ver 2.0 Database with Respect to FVR at FAR = 0.1% (in %)

Methods	Exp 1 ROC 3	Exp 4 ROC 3
1. H-Groupwise MRF	**98.2**	90.3
2. Liu's method [17]	92.0	76.0
3. Hwang et al's method [52]	91.5	74.3
4. Su et al's method [53]	98.0	89.0
5. Tan et al's method [54]	N/A	88.1
6. Lei et al's method [13]	N/A	89.9
7. Liu et al's method [55]	N/A	**92.43**

For each column, the highest recognition rate is bolded.

TABLE 6

The Estimated Mean Accuracies and the Standard Deviations for View 2 of the LFW Database

Methods	Accuracy ($\mu \pm S_E$)
1. SEE-H-Groupwise MRF	**0.8673 ± 0.0043**
2. DoG-H-Groupwise MRF	0.8442 ± 0.0047
3. Groupwise MRF	0.8104 ± 0.0052
4. Pairwise MRF [33]	0.7886 ± 0.0049
5. D-Demons [49]	0.6947 ± 0.0035
6. FFD [50]	0.6173 ± 0.0047
7. Nowak, original [59]	0.7245 ± 0.0040
8. Nowak, funneled [59]	0.7393 ± 0.0049
9. Hybrid descriptor-based [60]	0.7847 ± 0.0051
10. LARK, unsupervised [61]	0.7223 ± 0.0049
11. Pixels/MKL, funneled [62]	0.6822 ± 0.0041
12. V1-like/MKL, funneled [62]	0.7935 ± 0.0055

SEE-H-Groupwise MRF and DOG-H-Groupwise MRF denote our method using the proposed SEE and the difference of Gaussian (DoG) salient detector in [40], respectively. Groupwise MRF denote our method without using the hierarchical registration strategy. The highest verification rate is bolded.