Hierarchical Ensemble of Global and Local Classifiers for Face Recognition

Yu Su, Shiguang Shan, *Member, IEEE*, Xilin Chen, *Member, IEEE*, and Wen Gao, *Fellow, IEEE*

***Abstract*—In the literature of psychophysics and neurophysiology, many studies have shown that both global and local features are crucial for face representation and recognition. This paper proposes a novel face recognition method which exploits both global and local discriminative features. In this method, global features are extracted from the whole face images by keeping the low-frequency coefficients of Fourier transform, which we believe encodes the holistic facial information, such as facial contour. For local feature extraction, Gabor wavelets are exploited considering their biological relevance. After that, Fisher's linear discriminant (FLD) is separately applied to the global Fourier features and each local patch of Gabor features. Thus, multiple FLD classifiers are obtained, each embodying different facial evidences for face recognition. Finally, all these classifiers are combined to form a hierarchical ensemble classifier. We evaluate the proposed method using two large-scale face databases: FERET and FRGC version 2.0. Experiments show that the results of our method are impressively better than the best known results with the same evaluation protocol.**

***Index Terms*—Ensemble classifier, face recognition, Fisher's linear discriminant (FLD), Fourier transform, Gabor wavelets, global feature, local feature.**

I. Introduction

FACE recognition from still images and video sequence has been an active research area due to both its scientific challenges and wide range of potential applications such as biometric identity authentication, human-computer interaction, and video surveillance. Within the past two decades, numerous face recognition algorithms have been proposed as reviewed in the literature survey [1]. Even though we human beings can detect and identify faces in a cluttered scene with little effort, building an automated system that accomplishes such objective is very challenging. The challenges mainly come from the large variations in the visual stimulus due to illumination conditions, viewing directions, facial expressions, aging, and disguises such as facial hair, glasses, or cosmetics.

As in any pattern classification task, feature extraction plays a key role in face recognition process. In feature extraction stage, a proper face representation is chosen to make the subsequent face processing not only computationally feasible but also robust to possible intrinsic and extrinsic facial variations. Existing face representations fall into two categories: global-based and local-based. In the global-based face representation, each dimension of the feature vector contains the information embodied in every part (even each pixel) of the face image, thus corresponds to some holistic characteristic of the face. In contrast, for the local-based face representation, each dimension of the feature vector corresponds to merely certain local region in the face, thus only encodes the detailed traits within this specific area.

In the literature of face recognition, there are various face representation methods based on global features, including a great number of subspace-based methods and some spatial-frequency techniques. Subspace-based methods, such as principal component analysis (PCA) [2], Fisher's linear discriminant (FLD) [3] and independent component analysis (ICA) [4], have been widely recognized as the dominant and successful face representation methods. These methods attempt to find a set of basis images from a training set and represent any face as a linear combination of these basis images. Many researchers also propose to extract facial features by using spatial-frequency techniques, such as Fourier transform [5], [6] and discrete cosine transform (DCT) [7], [8]. In these methods, face images are transformed to the frequency domain and only the coefficients in the low-frequency band are reserved for face representation. One of the merits of these methods, compared with the subspace-based methods, is that they do not need a training process to learn the basis images.

While global-based face representations were popular for face recognition, recently, more and more attempts are made to develop face recognition systems based on local features, which are believed more robust to the variations of facial expression, illumination and occlusion etc.

In [9], Penve and Atick proposed local feature analysis (LFA) to encode the local topological structure of face images. LFA is considered as local method as it utilizes a set of kernels to implicitly detect the local structure such as eyes, nose, and mouth.

Manuscript received September 10, 2008; revised March 28, 2009. First published June 23, 2009; current version published July 10, 2009. This work was done at the Institute of Computing Technology, Chinese Academy of Sciences, and supported in part by the NSFC under Contracts 60772071, 60833013, and U0835005; in part by the Hi-Tech Research and Development Program of China under Contract 2007AA01Z163; and in part by 100 Talents Program of CAS, and National Basic Research Program of China (973 Program) under Contract 2009CB320902. Portions of the research in this paper use the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Arun Abraham Ross.

Y. Su is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China (e-mail: ysu@jdl.ac.cn).

S. Shan and X. Chen are with the Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Beijing 100190, China (e-mail: sgshan@jdl.ac.cn; xlchen@jdl.ac.cn).

W. Gao is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China, and also with the Institute of Digital Media, Peking University, Beijing 100871, China (e-mail: wgao@pku.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2009.2021737

Timo *et al.* [10] adopted local binary pattern (LBP), which is originated from the area of texture analysis, for face representation. In their method, LBP operator is first applied and then the resulting LBP "image" is divided into small regions from which histogram features are extracted. The idea of dividing face image is also used in the component-based methods, in which the face image is divided into some blocks by a certain rule. Then, the image blocks may be taken as inputs of classifiers (e.g., SVM [11]) or given to next step for further feature extraction (e.g., PCA [12], [13], FLD [14]).

Among various local features, especially, Gabor wavelets have been recognized as one of the most successful local feature extraction methods for face representation due to their biological relevance. The 2-D Gabor wavelets [15], whose kernels are similar to the 2-D receptive field profiles of the mammalian cortical simple cells, exhibit desirable characteristics of spatial locality and orientation selectivity, and are optimally localized in the space and frequency domains. Typical face recognition methods based on Gabor features include the elastic bunch graph matching (EBGM) [16], Gabor–Fisher classifier (GFC) [17], AdaBoost-based Gabor feature selection [18], and local Gabor binary pattern (LGBP) [19]. Especially, in recent years, Gabor wavelets are often combined with discriminant analysis methods (e.g., FLD) to further enhance the performances of face recognition systems [17], [20]–[23].

Although many successful face representation methods based on global or local features have been proposed, it remains an open problem that what is the most suitable representation for face recognition. However, in the literature of psychophysics and neurophysiology, many studies have shown that both global and local features are crucial for human face perception. Moreover, global and local features play different roles in the process of face perception and recognition. Global features describe the general characteristics of the holistic face and they are often used for coarse representation. Differently, local features reflect and encode more detailed variations within some local facial regions. Thus, it is proper to use local features for finer representation.

Following the above studies, it is natural to expect better performance of face recognition by combining global and local information. The EBGM method [16] had pioneered such an idea, since in EBGM global topological information is modeled as the structure of the graph and local information is encoded as the attribute of the nodes. In addition, Fang *et al.* [24] proposed to combine global PCA features and component-based local features extracted by Haar wavelets. Kim [14] proposed an effective face descriptor by decomposing a face image into several components, extracting FLD features from each component, and finally combining these component FLD features together with the features extracted by using a holistic FLD. Similar idea was proposed in another paper [25], in which the authors experimentally showed that the combined subspace gives smaller Bayesian error than the subspaces of either the global or local features. Lee [26] combined local structures extracted by local feature analysis (LFA) into composite templates which show compromised aspects between kernels of LFA and Eigenfaces. Lin and Tang [27] introduced a multilayer framework for high resolution face recognition. In their method, PCA followed by regularized FLD is exploited to model global appearance and facial organs. Meanwhile, discriminative multiscale texton features and SIFT-activated pictorial structure are used to extract local features such as skin and subtle details.

In this paper, following the same basic belief of combining global and local features, we propose a novel hierarchical ensemble classifier (HEC) for face recognition by combining global Fourier features and local Gabor features. Specifically, in our method, global features are extracted from the whole face images firstly by 2-D discrete Fourier transform. Then, the real and imaginary components of the low frequency band are concatenated to form a single feature vector, called by us global Fourier feature vector (GFFV), for further process. For local feature extraction, Gabor wavelet transform is exploited. Firstly, Gabor wavelets are used to extract local features at every position of the face image. Then, these features are spatially grouped into a number of feature vectors, each corresponding to a local patch of the face image and called by us a local Gabor feature vector (LGFV). After the above processes, a face image can be represented by one GFFV and multiple LGFVs. These feature vectors encode diverse discriminatory information: GFFV contains global discriminatory information and each LGFV embodies discriminatory information within certain local region. In order to make full use of all these diverse information, we propose to train multiple component classifiers by applying FLD on GFFV and each LGFV respectively, and then combine them into one ensemble classifier appropriately. The proposed method is extensively evaluated on the FERET and FRGC databases, and impressive results are achieved. Especially, on FRGC Experiment 4, we have achieved a verification rate of 89% at FAR of 0.1%, while the best known result was 81%.

The remaining parts of this paper are organized as follows. In Section II, face representation based on global and local features is proposed. In Section III, the construction of the hierarchical ensemble classifier is presented. In Section IV, experiments and analyses are conducted, followed by conclusion and discussion in the last section.

II. Extraction of Global and Local Features for Face Representation

As mentioned previously, global and local facial features play different roles in face perception. Therefore, it is necessary to combine them together smartly. Intuitively, local information is embedded in the detailed local variations of facial appearance, while global information means the holistically structural configuration of facial organs, as well as facial contour. Thus, from the viewpoint of frequency analysis, global features should mainly correspond to the lower frequencies, while local features should be of high frequency and dependent on position and orientation in the face image. Considering that, in this paper, global information is represented as the Fourier coefficients in low frequency band, and local information is encoded as the responses of multiscale and multiorientation Gabor wavelets.

It is known that the Gabor wavelet is a Gaussian modulated Fourier transform. Therefore, it can be tuned to extract global (usually low frequency) features by increasing the bandwidth and the radius of its Gaussian modulator. However, doing like

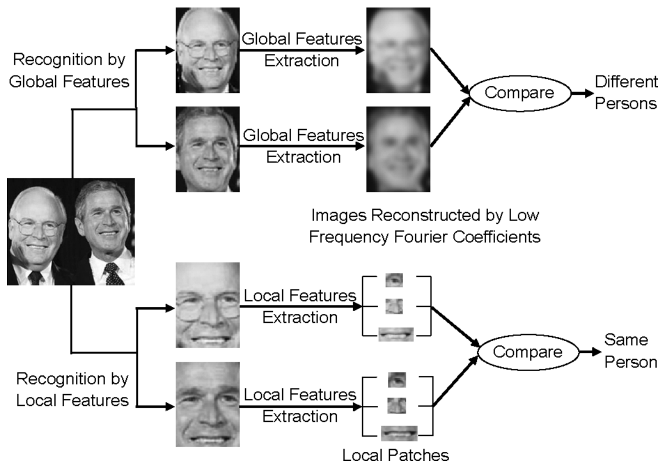

Fig. 1. Illustration of the different roles of global and local features in face recognition. See text for the detailed explanation. (The original faces in this figure are from P. Sinha, and T. Poggio, "Last but not least: 'United' we stand", Perception, 31(1), p.133, 2002.).

this is not as computationally desirable as using Fourier transform directly. Specifically, we hope the global features should be compact and orientation-independent. If we apply multiple Gabor wavelets to achieve orientation-independent, the computational burden of global feature extraction will increase significantly. In addition, the high dimensionality of Gabor features also brings the problem of "curse of dimensionality" and makes the following process much computationally expensive. That is the reason why Fourier transform rather than tuned Gabor wavelets is adopted to extract global features in this paper.

In what follows, we first illuminate the different roles of global and local features. Then, the detailed process of global and local feature extraction is introduced.

A. *Different Roles of Global and Local Features*

In this sub-section, different roles of global and local features are illustrated intuitively using two interesting example images. As shown in Fig. 1, the leftmost two input faces are artificial, whose main components (eyes, nose, and mouth) are actually from an identical person. But they have different facial contours and hairstyles. Therefore, they look holistically very dissimilar in terms of the overall structural configuration, hair and facial contour. Consequently the classifier based on global features will report them as different persons. However, the classifier comparing their local components is apt to reporting them as the same person, since their components are almost the same. The conflicting results of the two classifiers interestingly reflect the above-mentioned different roles of global and local information, which suggests that ideal classifier should be the combination of the two "experts."

B. *Extraction of Global Fourier Features*

In this paper, 2-D Discrete Fourier Transform (DFT) is adopted for global feature extraction. An image can be transformed by 2-D DFT into frequency domain as follows:

$$F(u,v) = \frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)e^{-j2\pi(ux/M+vy/N)} \qquad (1)$$

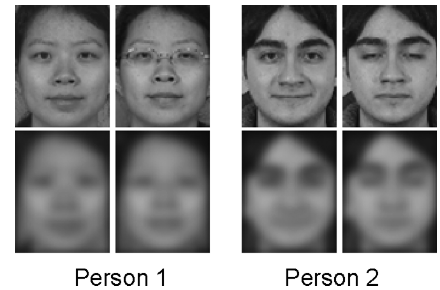

Fig. 2. Reconstruction of the face images by using only low-frequency Fourier coefficients (30% energy reserved).

where $f(x,y)$ represents a 2-D image of size M by N pixels, $0 \le u \le M-1$ and $0 \le v \le N-1$ are frequency variables. When the Fourier transform is applied to a real function, its outputs are complex numbers, that is

$$F(u,v) = R(u,v) + jI(u,v) \qquad (2)$$

where $R(u,v)$ and $I(u,v)$ are the real and imaginary components of $F(u,v)$ respectively. Thus, after Fourier transform, a face image is represented by the real and imaginary components of all the frequencies.

Though all the frequencies contain information about the input image, different bands of frequency play different roles. It is known that generally low frequencies reflect the holistic attributes of the input image. This can be illustrated intuitively by observing the effects of inverse transform with only the frequency band of interest. Fig. 2 gives some examples of inverse transform by using only the low frequency bands (about 30% of all the energy). From Fig. 2, one can safely conclude that the low frequencies indeed mainly contain information about the global structural configuration of the facial organs and the contour of the face. It is also apparent that these low-frequency features are very robust to the detailed local variations in appearance due to facial expressions, noise, and so on. In Section IV, these characteristics are further validated by experiments.

Consequently, in our method, only the Fourier coefficients in low frequency band are reserved as global features. Specifically, for a face image, after Fourier transform, the real and imaginary components in the low frequency band are concatenated into a single feature vector, named global Fourier feature vector (GFFV). As shown in Fig. 3, for both real and imaginary components, only those within low frequency band are reserved, which are illustrated by the white squares in the figure.

C. *Extraction of Local Gabor Features*

In recent years, face descriptors based on Gabor wavelet transform (GWT) have been recognized as one of the most successful face representation methods. Gabor wavelets are in many ways like Fourier transform but have a limited spatial scope. 2-D Gabor wavelets are defined as follows [15]:

$$\psi_{u,v}(z) = \frac{||k_{u,v}||^2}{\sigma^2} e^{(-||k_{u,v}||^2||z||^2/2\sigma^2)}\left[e^{i\vec{k}_{u,v}z} - e^{-\sigma^2/2}\right] \qquad (3)$$

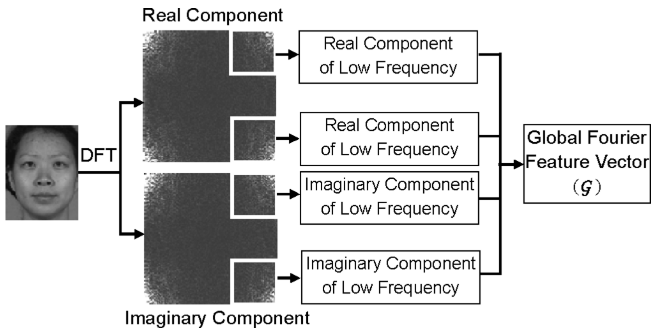

Fig. 3. Global feature extraction by 2-D DFT.

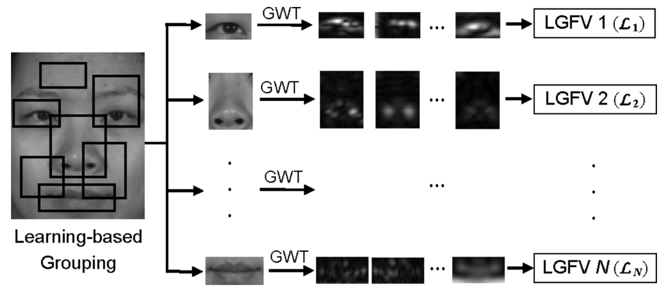

Fig. 5. Procedure of LGFVs construction. In our method, GWT is first applied to the whole face image, and then the resulting Gabor features are spatially grouped to obtain multiple LGFVs.

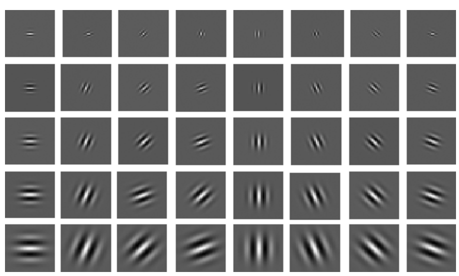

Fig. 4. Real part of the 40 2-D Gabor wavelets with five scales and eight orientations.

where $k_{u,v} = k_v e^{i\varphi_u}$, $k_v = k_{\max}/f^v$ gives the frequency, $\varphi_u = u\pi/8, \varphi_u \in [0, \pi)$ gives the orientation. As can be seen from the definition, Gabor wavelet consists of a planar sinusoid multiplied by a 2-D Gaussian. The Gaussian insures that the convolution is dominated by the region of the image close to the center of the wavelet. That is, when a signal is convolved with a Gabor wavelet, the frequency information near the center of the Gaussian is encoded and frequency information far away from the center of the Gaussian has a negligible effect. Therefore, compared with Fourier transform which extracts the information in the whole face region, Gabor wavelets only focus on some local areas in the face and extract information of specific scale and orientation within these local areas.

Gabor wavelets can take a variety of different forms with different scales and orientations. Fig. 4 shows the real part of the 40 Gabor wavelets with 5 scales and 8 orientations. Evidently, Gabor wavelets with a certain orientation respond to edges and bars along this direction, and Gabor wavelets with a certain scale extract the information in the corresponding frequency band. Thus, Gabor wavelets can extract more details in some important facial areas such as eyes, nose and mouth, which are very useful for face representation.

Given the above-defined Gabor wavelets, Gabor features are then extracted by convolving them with sub-windows sliding the face image pixel by pixel. Thus, if all the Gabor features are concatenated to form a single feature vector (we call it holistic representation, as in GFC [17]), the locality information (or neighboring information) provided by the spatial locations of Gabor features is not completely utilized. However, human faces contain some components with fixed high-level semantics such as eyes, nose and mouth. Consequently, the locality information is very meaningful for face modeling. In this work, in order to reserve more locality information, Gabor features (only magnitudes) are spatially grouped into a number of feature vectors named local Gabor feature vector (LGFV), each of which corresponds to a local patch of the face image. We call it patch-based representation. LGFV is of relatively low dimensionality, which can greatly facilitate the sequent process. In addition, the patch-based representation is more robust to lighting variation than the holistic representation. The reason is that the lighting variation within each patch is much smaller than that within whole face image, thus, can be better modeled by the following subspace learning (e.g., FLD). This point is further validated by the experiment in Section IV.

Naturally, one problem further arises: how to spatially group the Gabor features. Intuitively, these local patches should be located according to the facial features such as eyes, nose, and mouth. However, this requires accurate localization of these facial features, which is still very challenging. In our previous work [28], the patches are artificially and empirically designed. However, in this paper, we propose a patch selection method to automatically determine the positions and sizes of the local patches. Specifically, a feature selection method is adopted to select a number of local patches with high discriminability from a large number of possible local patches. Fig. 5 illustrates the idea of LGFVs construction based on the preselected image patches. In the following subsection, we will introduce the patch selection process in detail.

D. Patch Selection via Greedy Search

In principle, in case of allowing overlapping, the local patch can locate at any position in the image and be of any size. Thus, we will have too many candidate patches to construct LGFVs. Fortunately, their discriminating capacities are different and they are correlative. Therefore, it is feasible to learn only part of them with the largest discriminating capacities and at the same time with as little correlation as possible. By considering the candidate patches as "features", the problem of patch selection can be cast as feature subset selection.

In this paper, the wrapper methodology proposed by Kohavi and John [29] is exploited to address the problem of feature subset selection. In the wrapper methodology, the prediction performances of feature subsets, usually computed on a validation set, are adopted to evaluate the usefulness of them. The measure of prediction performance depends on the task

Patch Selection with Greedy Search

Input: Candidate Patch Set $CPS=\{p_1, p_2, \ldots, p_C\}$ where p_i is the candidate patch and C is the size of CPS, number of selected patches N, Evaluation Function F.

Output: Selected Patch Set (SPS).

Initialize: set $SPS = \varnothing$.

Repeat N times

1. For all $p_i \in CPS$, evaluate the performance of feature subset $SPS \cup \{p_i\}$, i.e., $F(SPS \cup \{p_i\})$.
2. Find the patch with the largest performance increase: $p = \arg\max_{p_i \in CPS} F(SPS \cup \{p_i\})$.
3. Update SPS and CPS: $SPS = SPS \cup \{p\}$, $CPS = CPS / \{p\}$.

Fig. 6. Algorithm of patch selection with greedy search.

to be handled. For example, in pattern classification, the prediction performance of certain feature subset can be set as the classification accuracy of the classifier based on this feature subset.

Generally, the only method for determining the optimal feature subset is the exhaustive search. However, due to its combination nature, the search will quickly become computationally intractable as the number of features increases. Considering that, in this work, the efficient greedy search strategy is adopted to select the desired feature subset. In the greedy search, features are progressively (e.g., one by one) incorporated into a larger and larger subset. For each step, the feature bringing the largest increase of prediction performance is selected and added to the current subset. The detailed process of patch selection by greedy search is formulated in Fig. 6. In this process, the evaluation function F is used to compute the performance of feature subsets on a validation set. In fact, F involves two steps: the first is to learn a predictor for certain feature subset on a training set; the second is to compute the performance of this predictor on a validation set. Furthermore, for the purpose of consistency, these two steps are designed to be the same as the training and testing phase of the proposed face recognition method respectively. More specifically, for certain patch subset, the local ensemble classifier (see Section III) trained based on these patches is considered as the predictor. As for the evaluation, the recognition rate or verification rate is used to measure the performance of certain predictor.

Besides the above patch selection criteria, some other possibilities are also acceptable. For instance, Bicego *et al.* [30] proposed to identify distinctive areas of each individual's face by its comparison to others in the population. This method is in some sense similar to our patch selection, if the discriminability is considered as the measure of saliency. However, there also exists significant difference of this method from ours. Specifically, our method learns a set of patches with high discriminability in the training phase. Then, in the testing phase, all the face images are divided according to these learned patches. Nevertheless, the saliency maps of different face images are not the same. In other words, the saliency map of a certain face image is determined in the testing phase (i.e., in an online manner), whereas the discriminative image patches in our method are determined in the training phase (i.e., in an offline manner).

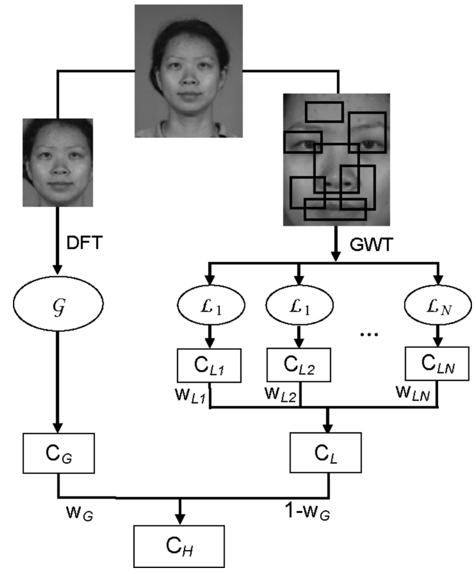

Fig. 7. Paradigm of the hierarchical ensemble classifier.

III. Hierarchical Ensemble Classifier: Combining Global and Local Features

After feature extraction, we obtain $N+1$ feature vectors, that is, one global Fourier feature vector (GFFV) $\mathcal{G}$ and N local Gabor feature vectors (LGFVs) $\mathcal{L}_i$ $(i = 1, \ldots, N)$. Then, $N+1$ classifiers can be trained by applying FLD to each feature vector. These classifiers are named as component classifiers, opposite to the forthcoming ensemble classifier, i.e., the combination of component classifiers. As explained above, these $N+1$ feature vectors contain diverse discriminative information for face recognition. Thus, component classifiers trained on these feature vectors should have certain degree of error diversity. In other words, these component classifiers might agree or disagree with each other when making decision. Considering that the ensemble classifier is generally superior to the single classifier when the predictions of its component classifiers have enough diversity [31], we combine the component classifiers trained on all the feature vectors into a hierarchical ensemble classifier to improve the recognition accuracy. In this process, the weighted sum rule is adopted for classifier combination.

In the remaining part of this section, we will present the detailed process of constructing the hierarchical ensemble classifier and learning the weight of each component classifier.

A. Construction of Hierarchical Ensemble Classifier

As shown in Fig. 7, the proposed hierarchical ensemble method consists of two layers of ensemble: the ensemble of all the local component classifiers, and the ensemble of local classifier and global classifier. In the first layer, local ensemble classifier (LEC) C_L is obtained by combining N local component classifiers (LCC) C_{L_i} $(i = 1, \ldots, N)$, each trained on an LGFV $\mathcal{L}_i$ $(i = 1, \ldots, N)$, with N the number of selected patches. It is formulated as follows:

$$C_L = \sum_{i=1}^{N} w_{L_i} \cdot C_{L_i} \tag{4}$$

where w_{L_i} is the weight of the ith LCC, C_{L_i}. The method to determine the weights is discussed in next sub-section. In the second layer, the LCC C_L obtained in the first layer is combined with the global classifier (GC) C_G trained on the GFFV $\mathcal{G}$ to form the hierarchical ensemble classifier (HEC) C_H, as formulated in (5)

$$C_H = w_G C_G + (1 - w_G) C_L \tag{5}$$

where w_G is the weight of GC C_G.

As mentioned previously, global and local features play different roles in face perception. While global features capture the holistic characteristics of the face, therefore, better for coarse representation; local features encode more details in local face areas, therefore, better for finer representation. Considering that, in our method, the input face image is normalized differently for global and local feature extraction. As shown in Fig. 7, the global Fourier features are extracted from the face image of lower resolution, but covering both external and internal facial features, especially the face contour. On the contrary, the local Gabor features are extracted from the face image of higher resolution, which covers only the internal facial features, e.g., the facial organs. The reason using this strategy lies in the sensitivity of Gabor features to the possible "background" introduced along with the contour, to which the Fourier features are very robust. The effect of different spatial resolutions of face image on the performance of global and local classifiers is analyzed by experiment in Section IV.

As for the component classifier, basically, we have several choices. In this paper, we exploit FLD for its desirable characteristics to well separate within-class and between-class variations simultaneously, which has been demonstrated in some previous works on face recognition [3], [17].

B. Weight Learning for Component Classifiers

In this sub-section, we formulate weight learning as a discriminant analysis problem, in which the weighted sum can be considered as linear projection from N-D to 1-D and the projection coefficients can be treated as weights. The weight learning method consists of three steps. Firstly, the face images in training set are converted into image pairs, which can be divided into two classes named intrapersonal pairs and interpersonal pairs. Secondly, for each image pair, a similarity vector can be obtained. Each similarity in this vector is given by a certain LCC; therefore, the dimension of the similarity vector is N. Each similarity vector can be considered as a sample in the space of N dimensions. Thus, in the last step, two classes of the N-dimensional samples are fed into FLD to get an optimal linear projection from N-D to 1-D. As mentioned above, each coefficient of this linear projection can be considered as the weight of corresponding LCC. The aim of weighting LCCs is to make the similarities of image pairs from two classes (intrapersonal and interpersonal) more discriminable, which is consistent with the aim of FLD. Thus, the projection coefficients computed by FLD is also the optimal weights for discriminating the intrapersonal pairs and the interpersonal pairs. This weight learning process is illustrated in Fig. 8. The weights of GC and LEC can be similarly learned from training set by this method.

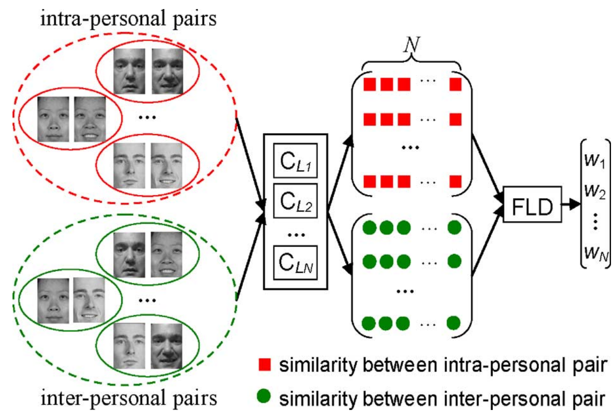

Fig. 8. Illustration of the weight learning process. C_{Li} denotes the ith LCC; w_i denotes the weight of ith LCC; and N denotes the number of LCCs.

TABLE I
SIZES OF TEST SETS IN FRGC EXPERIMENT 1 AND 4. [C] AND [U] MEAN CONTROLLED AND UNCONTROLLED ILLUMINATION CONDITION, RESPECTIVELY

Experiment	Target Set Size	Query Set Size
1	16028 [C]	16028 [C]
4	16028 [C]	8014 [U]

IV. EXPERIMENTS

In this section, the proposed method is tested on some virtual face images and two large-scale face databases: FERET [32] and FRGC version 2.0 [33]. Both databases are publicly released along with standard evaluation protocols. For FERET database, the proposed method is tested on four standard probe sets: fafb, fafc, dup1, and dup2, which are matched against the gallery with 1196 subjects and one image per subject. The images in fafb set are with expression variations, fafc set contains images with lighting variations, and dup1 and dup2 from different times. The readers are referred to [32] for more details. For FRGC database, Experiment.1 and 4 on still images are taken in our testing. Configurations of the two experiments are summarized in Table I. The performance is reported as verification rates (VR) at 0.1% false acceptance rate (FAR). Following the protocol of FRGC, for each experiment, three receiver operator characteristic (ROC) curves are generated. Among them, ROC I is corresponding to the images collected within semesters, ROC II within a year, and ROC III between semesters.

In what follows, we first design experiments to exam the contribution of global and local features on both synthetic and real data. Then, we give the performance of both the ensemble classifier and the component classifiers. We also compare the results of our method with the baseline and the best known results on the FERET and FRGC databases. It is worth pointing out that, to emphasize more the recognition method itself, throughout our experiments, no photometric normalization is performed on images in the FERET and FRGC databases.

A. Different Roles of Global and Local Features

In order to demonstrate the different roles of global and local features, we generate 10 groups of virtual face images including variations such as facial contour, hairstyle, facial organs and lentigines. Fig. 9 shows one of groups of the virtual face images

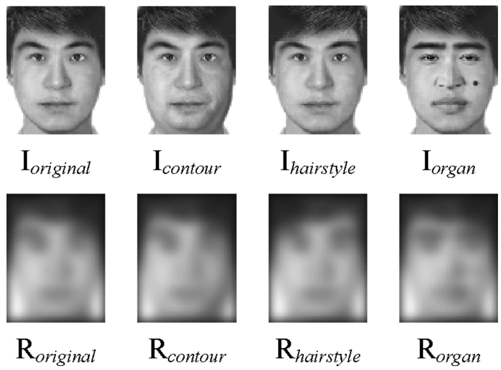

Fig. 9. Example of synthetic face images and their reconstruction only from low-frequency Fourier features. $\mathrm{I}_{\mathrm{original}}$ is the original face image. $\mathrm{I}_{\mathrm{contour}}$, $\mathrm{I}_{\mathrm{hairstyle}}$, and $\mathrm{I}_{\mathrm{organ}}$ are generated by changing the facial contour, hairstyle and facial organs (including lentigines), respectively. The images in bottom row are reconstructed by the low-frequency Fourier features with 30% energy preserved.

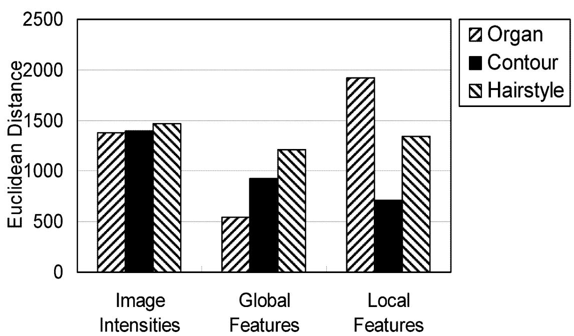

Fig. 10. Demonstration of different roles of global and local feature in face representation. The vertical coordinate represents the Euclidean distance between the original image and the virtually generated images.

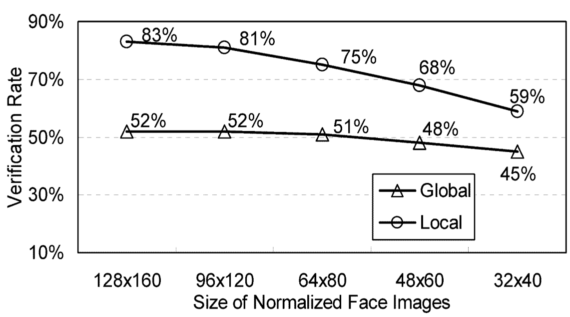

Fig. 11. Accuracy of classifiers based on global and local features extracted from face images of different sizes.

(top row) and their reconstructions only from low-frequency Fourier features (bottom row).

In this experiment, the virtual faces are generated to make the differences of them ($\mathrm{I}_{\mathrm{contour}}$, $\mathrm{I}_{\mathrm{hairstyle}}$ and $\mathrm{I}_{\mathrm{organ}}$) from the original face ($\mathrm{I}_{\mathrm{original}}$) be approximately equal in the sense of Euclidean distance (about 1450-1500 as shown in Fig. 10). However, the differences among their reconstruction images vary widely. Specifically, the Euclidean distance between $\mathrm{R}_{\mathrm{original}}$ and $\mathrm{R}_{\mathrm{organ}}$ is much smaller than the distance between $\mathrm{R}_{\mathrm{original}}$ and $\mathrm{R}_{\mathrm{contour}}$ and the distance between $\mathrm{R}_{\mathrm{original}}$ and $\mathrm{R}_{\mathrm{hairstyle}}$, which is shown in Fig. 10. Note that, Fig. 10 presents the average results of ten different persons.

From Fig. 10, it can be conclude that low-frequency Fourier features are indeed robust to detailed local variations, and mainly reserve large-scale variations such as facial contour and hairstyle. In addition, compared with global Fourier features, the distance of local Gabor features between $\mathrm{I}_{\mathrm{original}}$ and $\mathrm{I}_{\mathrm{organ}}$ is larger than the distance between $\mathrm{I}_{\mathrm{original}}$ and $\mathrm{I}_{\mathrm{contour}}$ and the distance between $\mathrm{I}_{\mathrm{original}}$ and $\mathrm{I}_{\mathrm{hairstyle}}$, which demonstrates that Gabor feature are more sensitive to the detailed local variations.

Different roles of global and local features are further validated on FRGC database with different normalized image sizes. As mentioned previously, global features contain mainly the holistic characteristics of the whole face, whereas local features encode more detailed variations within some local areas in the face. Intuitively, facial details will lose at small size. On the contrary, the global characteristic of the face, such as the configuration and the shape, can be reserved even at small size. Thus, global features should be more robust to the variation of face size, which is also supported by the experiment results on FRGC Experiment 4 with five different normalized sizes: 128×160, 96×120, 64×80, 48×60 and 32×40, as shown in Fig. 11. It can be seen from the figure that, with the decrease of the face size, the accuracies of local classifier drop more quickly than those of global classifier. This observation can further validate the different roles of global and local features: global features encode the holistic characteristics whereas local features capture the details.

B. Experiments About Global and Local Classifier

In this sub-section, we present the detailed experimental setups and report the performances of both global and local classifiers on FERET and FRGC databases. As mentioned in Section III, the input face images are geometrically normalized differently for global and local feature extraction. For both databases, face regions are extracted partially automatically according to the eye centers directly from the FERET and FRGC database. Specifically, for global feature extraction, the scale of normalization is controlled by locating the two eye centers to the coordinates of (19, 31) and (46, 31) respectively, and crop the image size as 64×80. For local feature extraction, the normalized face image is 128×160 pixels while the two eye centers are fixed at (29, 61) and (100, 61) respectively. A pair of example normalized faces can be founded in Fig. 7.

1) Experiments About Global Classifier: Given image size 64×80 as mentioned above, in order to apply the fast Fourier transform (FFT) algorithm, image must be extended to 128×128 pixels. Thus, the full bandwidth available is 64 due to the symmetry of the Fourier coefficients. As explained above, we need only the low-frequency Fourier coefficients. In the experiments, we select the first 16×16 FFT coefficients, which cover about 50% of all the energy, to form the global fourier feature vector (GFFV). Thus, referring to Fig. 3, the dimension of the GFFV is $16 \times 16 \times 4 = 1{,}024$. This feature vector is then processed by FLD to obtain the global classifier, whose accuracies on FRGC Experiment 1 and 4 are shown in Fig. 15.

2) Experiments About Local Classifiers: This sub-section first presents the experimental setup and results of the patch

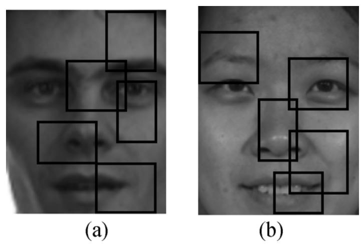

Fig. 12. Top five most discriminating patches learned from FERET (a) and FRGC (b) training set.

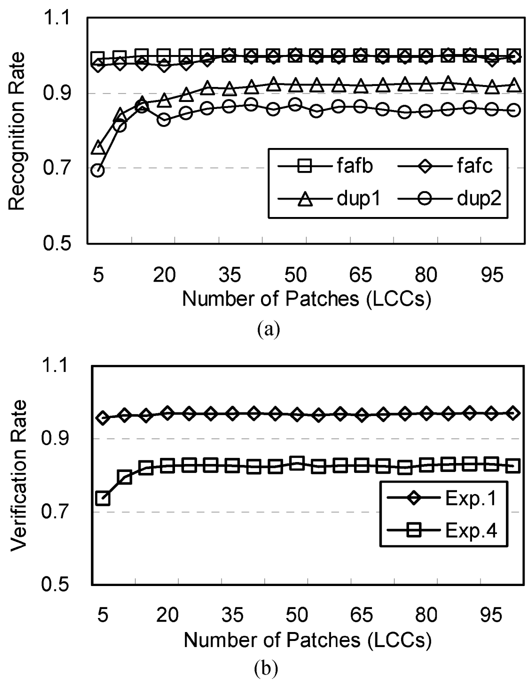

Fig. 13. Performances of LEC with different number of patches (LCCs) on FERET and FRGC databases.

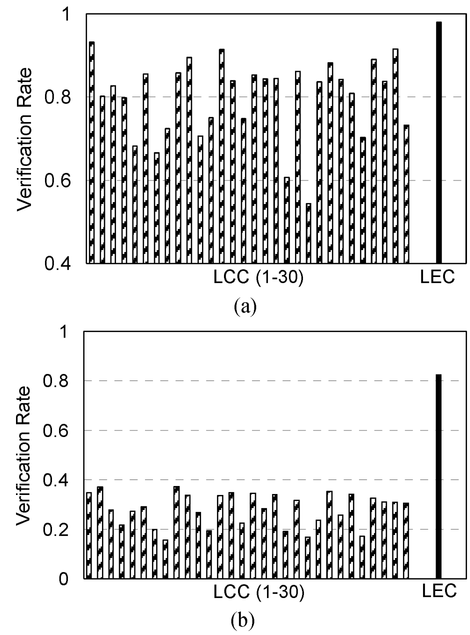

Fig. 14. Performance comparison of LCC and LEC on Experiment 1 (a) and Experiment 4 (b) of FRGC (ROCIII).

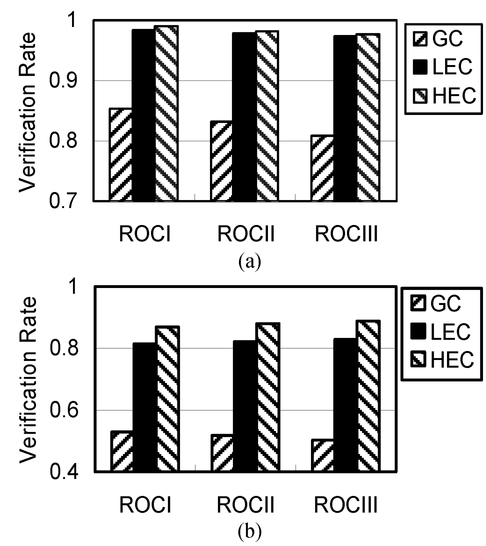

Fig. 15. Three ROC performances of GC, LEC and HEC on FRGC Experiment 1 (a) and Experiment 4 (b).

selection, followed by the description on the experimental design and evaluation results of the local component classifiers and their ensemble.

Patch Selection and the Design of LCC: In the process of patch selection, the size of candidate patches is set to range from [16, 64] $\times$ [16, 64]. To apply the algorithm in Fig. 6, the training sets of the FERET database are randomly divided into two subsets without any overlapping: one for predictor learning and the other for performance evaluation. Similar process is performed on FRGC database. Fig. 12 shows respectively the top five most discriminating patches learned with greedy search from the FERET and FRGC training set.

After patch selection, local Gabor feature vector (LGFV) is extracted from each selected patch by concatenating the magnitudes of all the Gabor convolution results in this patch. In this work, 40 Gabor wavelets (five scales and eight orientations) are used with the same parameters in [16]; therefore, the dimension of each LGFV is $D = \mathrm{w} \times \mathrm{h} \times 5 \times 8$, where w and h are respectively the width and height of the patch. Since the range of w and h is between 16 and 64, the maximum dimensionality D is 163,840 ($64 \times 64 \times 5 \times 8$), which is too high dimensional for FLD. To solve this problem, the Gabor features are uniformly down-sampled by averaging them in an 8×8 grid. The down-sampled features of each LGFV are then further used to train an LCC. Finally, N LCCs are combined to form the LEC.

Since each LCC encodes only part of the discriminating information of a given face, it is expected that the recognition accuracy of LEC can be improved with the increase of the number of LCCs. An experiment is conduced to verify this expectation. Fig. 13 shows how the performance of LEC changes with the number of patches. The figure indicates that, by and large, the performance of LEC increases with more LCCs combined. However, the performance improvement becomes trivial when the number of LCCs exceeds 30. As can be imagined, 30 patches should have covered most of the face region; therefore, adding

TABLE II
INFLUENCE OF EXCHANGING TRAINING SET FOR PATCH SELECTION AND WEIGHT LEARNING ON THE PERFORMANCE OF LEC (WITH 30 PATCHES)

Training set \ Testing set	FRGC (Verification Rate)		FERET (Recognition Rate)	
	Exp.1	Exp.4	fafb	dup1
FRGC training set	97.3%	82.8%	99.7%	90.7%
FERET training set	97.1%	79.7%	99.9%	91.4%

more patches might introduce merely more redundancy rather than more complementary discriminating information. Thus, in the following experiments, only the top 30 LCCs are combined to construct the LEC. By using the patch selection strategy, the performance of LEC increases 3 percents on FRGC Experiment 4 compared with that of our previous work [28], in which the face images are artificially partitioned into 20 nonoverlapping patches.

To show how the training set might influence the patch selection results, we exchange the training sets of FERET and FRGC, and show the experimental results in Table II, from which we can conclude that the accuracies of LEC are basically not sensitive to the training set. The possible reason is that human faces are all very similar in overall configuration; therefore, the results of both weights learning and patch search from a relatively large dataset should have good generalizability.

Comparison of LCCs With Their Ensemble: Another interesting point is how the ensemble of LCCs enhances the performance compared with individual LCC. Intuitively, since each LCC exploits only part of the discriminating information within certain facial regions, their performances are commonly not good enough. This is validated by our experimental results shown in Fig. 14, which shows the verification rates of the 30 LCCs, as well as that of the LEC. As shown in Fig. 14, the performance of LEC is much better than that of the individual LCC (especially on Experiment 4). This large gain can be mainly attributed to the complementarity among the LCCs since they capture the discriminating characteristics within different regions. In addition, as has been proved in machine learning field, by averaging outputs of multiple estimators, ensemble system can achieve a better estimation with less generalization error [34]. Therefore, as an ensemble classifier, LEC is expected to have desirable low generalization error.

Comparison of LEC With Gabor-Fisher Classifier: To further verify the merits of LEC, we also compare its performance with the Gabor–Fisher classifier (GFC) [17], in which FLD is applied to the Gabor features within the whole face image to obtain only one classifier. Since the Gabor features are too high-dimensional, they are down-sampled before applying FLD. As shown in Table III, the proposed LEC outperforms GFC significantly in both experiments. Especially, on Experiment 4, the verification rate of LEC is 23 percents higher than that of GFC. Considering that the query images in Experiment 4 were captured under uncontrolled situation and with severe illumination changes, we can conclude that the proposed LEC seems also more robust to illumination variation.

C. *Experiments About Hierarchical Ensemble Classifier*

In order to make full use of the discriminative information in both the global and the local features and further improve the system performance, GC and LEC are combined to form the hierarchical ensemble classifier (HEC), as formulated in (5). In the combination, the weight for GC W_G can actually balance the importance of global and local information. This is evidently necessary because the performances of GC and LEC are quite different, as can be seen from Fig. 15. Since the accuracy of GC is significantly lower than that of LEC, it is desirable to assign a smaller weight for GC. We use the similar weight learning method proposed in Section III to determine W_G, the weight of GC. By learning on the training set, we get W_G equal to 0.13 for FERET, while W_G equals to 0.16 for FRGC. With the learned weight, experiments are conducted on both FERET and FRGC databases. The results are given in Tables IV and V.

In Table IV, besides the results of the proposed GC, LEC, and HEC, we also give the best results in FERET'96 evaluation [35], the performances of LBP-based method [10] and LGBP-based method [19]. As can be seen from the table, the performances of both LEC and HEC are better than those of the comparison methods. Especially on dup1 and dup2 probe sets, the improvement is very significant, which further verifies the robustness of our method to variations due to illumination, expression and aging, since the images in these two sets cover these variations. From the results, we also observe performance gain by combining global and local classifiers.

TABLE III
PERFORMANCE COMPARISON OF GFC AND THE PROPOSED LEC ON FRGC EXPERIMENT 1 AND 4 (ROC III)

Method \ Testing set	Verification Rate at 0.1% FAR	
	Exp. 1	Exp. 4
GFC	92.5%	59.8%
Proposed LEC	97.3%	82.8%

TABLE IV
PERFORMANCE COMPARISONS ON FOUR STANDARD PROBE SETS OF FERET. THE RESULTS OF OTHER METHODS ARE DIRECTLY CITED FROM THE CORRESPONDING PAPERS

Method \ Testing set		Rank-1 Recognition Rate			
		fafb	fafc	dup1	dup2
FERET'96 [35]		96%	82%	59%	52%
Method in [10]		97%	79%	66%	64%
Method in [19]		98%	97%	74%	71%
Our Methods	GC	96%	77%	53%	23%
	LEC	99%	99%	91%	86%
	HEC	**99%**	**99%**	**92%**	**88%**

TABLE V
PERFORMANCES COMPARISON ON EXPERIMENT 1 AND 4 OF THE FRGC DATABASE (ROC III). THE RESULTS IN [6], [36], [37] ARE CITED DIRECTLY FROM THE CORRESPONDING PAPERS

Method \ Testing set		Verification Rate at 0.1% FAR	
		Exp. 1	Exp. 4
FRGC Baseline [33]		66%	12%
Method in [6]		91%	74%
Method in [36]		92%	76%
Method in [37]		N/A	81%
Our Methods	GC	81%	51%
	LEC	97%	83%
	HEC	**98**%	**89%**

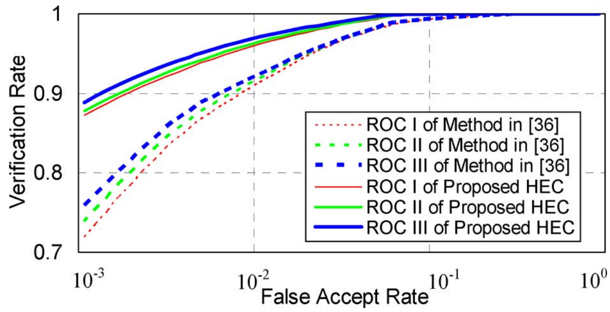

Fig. 16. ROC performances comparison between our method and Liu's method in [36] on Exp. 4. The results of Liu's method are taken directly from his paper.

Similarly, in Table V, we also compare our method with the FRGC baseline algorithm (basically PCA) and the best known results reported in [6], [36], and [37] on both Experiment 1 and 4. In [6], Hwang *et al.* proposed a Fourier-based face recognition system, in which Fourier features with different frequency bands and face models are projected into some linear discriminant subspaces and merged together. In [36], the author presented a pattern recognition framework which integrates Gabor image representation and multiclass kernel Fisher analysis (KFA) with fractional power polynomial models. In [37], the authors proposed to combine complementary features from three different representation levels on color images in YIQ color space. In each level, features are extracted by using discrete cosine transform (DCT) and enhanced Fisher model (EFM).

From Table V, one can see that the proposed methods (both LEC and HEC) significantly outperform the comparison methods on both Experiment 1 and 4 (ROC III). Especially on Experiment 4, the proposed method (HEC) achieves 89% verification rate at $\mathrm{FAR} = 0.1\%$, which is 8 percents higher than the best known results. As the query images in Experiment 4 cover illumination, blur and partial occlusion, we can also conclude that the proposed method is more robust to these extrinsic variations.

Another observation from Table V is that the combination of global and local features leads to significant performance gain. On Experiment 4, by combining GC with verification rate 51% and LEC with verification rate 83%, the verification rate of HEC increases to 89%. This large gain further validates their complementarity and the necessity to combine them.

Besides the results in Table V, we also report in Fig. 15 more results of the ROC I, II, and III on Experiment 1 and 4. Fig. 16 gives the complete comparison on the three ROC curves of our method (HEC) with those in [36]. From both figures, one can reach consistent observation as can be drawn from Table V.

V. Conclusion and Discussion

Inspired by the fact that we human beings recognize faces relying on both global and local facial features, a hierarchical ensemble method is proposed to simulate the observations in bionic sense by exploiting both global and local features. In the proposed method, the global features are extracted from the whole face images by using Fourier transform, and the local features are emphasized on some spatially divided face patches by using Gabor wavelets. The position and size of the patches are learned from a training data via greedy search. The hierarchical ensemble classifier is formed by weighted sum of the component classifiers, which are all FLDs on either global or local features. Experimental results on both FERET and FRGC version 2.0 databases show that the ensemble classifier outperforms other competitors. Compared with the baseline and best known results, the proposed method demonstrates significant improvement especially on FRGC Experiment 4.

To summarize the proposed method, we would like to attribute its favorable performance to the following aspects, which should be valuable to researchers in this area.

First, the combination of global and local features plays a key role. Experimental results show that they are indeed complementary for distinguishing faces. Although the local features seem significantly better than global features for face recognition, the accuracy can still be improved by carefully combining them.

Second, ensemble is also a key contributor to improve generalizability. In machine learning area, ensemble learning has been widely recognized as a successful method to avoid overfitting. In the proposed method, ensemble is applied in two stages: the combination of the local classifiers and the combination of the global and local classifiers. Experimental results show that both ensemble procedures improve impressively the performance compared with the component classifiers.

Finally, we would especially remark the "Gabor + FLD" method for face recognition. To our knowledge, Gabor wavelet and FLD have been recognized as two valuable pearls for face recognition. In previous work combining them together, many possibly discriminating Gabor features had to be down-sampled heavily due to the small sample size problem of FLD. In this paper, with the "divide and conquer" strategy, the Gabor features can be much less down-sampled when applying FLD to the patches. This should also be one of the sources of the performance gain.

As to the future work, though the global features are not as effective as the local ones, we still believe that the global features should play more important role in human perception consistent recognition. Therefore, it is one of our future efforts to study better methods for global feature extraction. In addition, patch-based method provides a nature way to deal with occlusion, and we will adopt the idea of switching off some occluded patches, for instance as presented in one of our previous works [38], to extend this work to solve partially occluded face recognition problem.

Acknowledgment

The authors would like to thank Prof. A. Ross and the anonymous reviewers, whose comments helped to improve the paper greatly.

REFERENCES

[1] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM. Computi. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.

[2] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[3] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 7, pp. 711–720, Jul. 1997.

[4] M. Bartlett, J. Movellan, and T. Sejnowski, "Face recognition by independent component analysis," *IEEE Trans. Neural Netw.*, vol. 13, pp. 1450–1464, 2002.

[5] J. Lai, P. Yuen, and G. Feng, "Face recognition using holistic fourier invariant features," *Pattern Recognit.*, vol. 34, no. 1, pp. 95–109, 2001.

[6] W. Hwang, G. Park, J. Lee, and S. Kee, "Multiple face model of hybrid fourier feature for large face image set," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2006, pp. 1574–1581.

[7] Z. Hafed and M. Levine, "Face recognition using the discrete cosine transform," *Int. J. Comput. Vis.*, vol. 43, no. 3, pp. 167–188, 2001.

[8] M. Savvides, J. Heo, R. Abiantun, C. Xie, and B. Kumar, "Class dependent kernel discrete cosine transform features for enhanced holistic face recognition in FRGC-II," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing*, 2006, vol. 2, pp. 185–188.

[9] A. Penev and J. Atick, "Local feature analysis: A general statistical theory for object representation," *Network: Comput. Neural Syst.*, vol. 7, pp. 477–500, 1996.

[10] A. Timo, H. Abdenour, and P. Matti, "Face recognition with local binary patterns," in *Proc. Eur. Conf. Computer Vision*, 2004, pp. 469–481.

[11] B. Heisele, P. Ho, and T. Poggio, "Face recognition with support vector machines: Global versus component-based approach," in *Proc. IEEE Int. Conf. Computer Vision*, 2001, vol. 2, pp. 688–693.

[12] S. Chen and Y. Zhu, "Subpattern-based principal component analysis," *Pattern Recognit.*, vol. 37, no. 5, pp. 1081–1083, 2004.

[13] R. Gottumukkal and V. Asari, "An improved face recognition technique based on modular PCA approach," *Pattern Recognit. Lett.*, vol. 25, no. 4, pp. 429–436, 2004.

[14] T. Kim, H. Kim, W. Hwang, and J. Kittler, "Component-based LDA face description for image retrieval and MPEG-7 standardisation," *Image Vis. Comput.*, vol. 23, no. 7, pp. 631–642, 2005.

[15] J. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Amer.*, vol. 2, no. 7, pp. 1160–1169, 1985.

[16] L. Wiskott, J. Fellous, N. Kruger, and C. Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 775–779, Jul. 1997.

[17] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Trans. Image Process.*, vol. 11, pp. 467–476, 2002.

[18] P. Yang, S. Shan, W. Gao, and S. Li, "Face recognition using Ada-Boosted Gabor features," in *Proc. Int. Conf. Automatic Face and Gesture Recognition*, 2004, pp. 356–361.

[19] W. Zhang, S. Shan, W. Gao, and X. Chen, "Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," in *Proc. IEEE Int. Conf. Computer Vision*, 2005, pp. 786–791.

[20] D. Liu, K. Lam, and L. Shen, "Optimal sampling of Gabor features for face recognition," *Pattern Recognit. Lett.*, vol. 25, no. 2, pp. 267–276, 2004.

[21] L. Shen, L. Bai, and M. Fairhurst, "Gabor wavelets and general discriminant analysis for face identification and verification," *Image Vis. Comput.*, vol. 25, no. 5, pp. 553–563, 2007.

[22] W. Fan, Y. Wang, W. Liu, and T. Tan, "Combining null space-based Gabor features for face recognition," in *Proc. IEEE Int. Conf. Pattern Recognition*, 2006, pp. 330–333.

[23] S. Li, W. Liu, D. Lin, and X. Tang, "Nonparametric subspace analysis for face recognition," in *Proc. IEEE Int. Conf. Pattern Recognition*, 2005, vol. 2, pp. 961–966.

[24] Y. Fang, T. Tan, and Y. Wang, "Fusion of global and local features for face verification," in *Proc. IEEE Int. Conf. Pattern Recognition*, 2002, vol. 2, pp. 382–385.

[25] C. Kim, J. Oh, and C. Choi, "Combined subspace method using global and local features for face recognition," in *Proc. IEEE Int. Joint Conf. Neural Networks*, 2005, vol. 4, pp. 2030–2035.

[26] Y. Lee, K. Lee, and S. Pan, "Local and global feature extraction for face recognition," in *Proc. Int. Conf. Audio- and Video-Based Biometric Person Authentication*, 2005, pp. 219–228.

[27] D. Lin and X. Tang, "Recognize high resolution faces: From macrocosm to microcosm," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2006, vol. 2, pp. 1355–1362.

[28] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical ensemble of global and local classifiers for face recognition," in *Proc. IEEE Int. Conf. Computer Vision*, 2007, 978-1-4244-1631-8/07.

[29] R. Kohavi and G. John, "Wrappers for feature selection," *Artif. Intell.*, vol. 97, pp. 273–324, 1997.

[30] M. Bicego, G. Brelstaff, L. Brodo, E. Grosso, A. Lagorio, and M. Tistarelli, "Distinctiveness of faces: A computational approach," *ACM Trans. Appl. Percept.*, vol. 5, no. 2, pp. 1–18, 2008.

[31] G. Brown, J. Wyatt, R. Harris, and X. Yao, "Diversity creation methods: A survey and categorization," *Inf. Fusion*, vol. 6, no. 1, pp. 5–20, 2005.

[32] P. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image Vis. Comput.*, vol. 16, no. 5, pp. 295–306, 1998.

[33] 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 *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2005, vol. 1, pp. 947–954.

[34] N. Ueda and R. Nakano, "Generalization error of ensemble estimators," in *Proc. IEEE Int. Conf. Neural Networks*, 1996, vol. 1, pp. 90–95.

[35] P. Phillips, H. Moon, P. Rauss, and S. Rizvi, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, 2000.

[36] C. Liu, "Capitalize on dimensionality increasing techniques for improving face recognition performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 5, pp. 725–737, 2006.

[37] Z. Liu and C. Liu, "Fusion of the complementary discrete cosine features in the YIQ color space for face recognition," *Comput. Vis. Image Understand.*, vol. 111, no. 3, pp. 249–262, 2008.

[38] W. Zhang, S. Shan, X. Chen, and W. Gao, "Local Gabor binary patterns based on Kullback-Leibler divergence for partially occluded face," *IEEE Signal Process. Lett.*, vol. 14, no. 11, pp. 875–878, 2007.

Yu Su received the B.S. and M.S. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 2003 and 2005, respectively, where he is currently pursuing the Ph.D. degree.

His research interests include pattern recognition, image processing, and especially focus on face recognition.

Shiguang Shan (M'04) received the M.S. degree in computer science from the Harbin Institute of Technology, Harbin, China, in 1999, and the Ph.D. degree in computer science from the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing, in 2004.

He has been with ICT, CAS, since 2002, and has been an Associate Professor since 2005. He is also the Vice Director of the ICT–ISVISION Joint Research and Development Laboratory for Face Recognition, ICT, CAS. His research interests cover image analysis, pattern recognition, and computer vision. He is focusing especially on face recognition related research topics. He received the China's State Scientific and Technological Progress Awards in 2005 for his work on face recognition technologies.

Xilin Chen (M'00) received the B.S., M.S., and Ph.D. degrees in computer science from the Harbin Institute of Technology (HIT), Harbin, China, in 1988, 1991, and 1994, respectively.

He was a Professor with the HIT from 1999 to 2005 and was a Visiting Scholar with Carnegie Mellon University, Pittsburgh, PA, from 2001 to 2004. He has been with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, since August 2004. His research interests include image processing, pattern recognition, computer vision, and multimodal interface.

Dr. Chen is a member of the IEEE Computer Society. He has received several awards, including the China's State Scientific and Technological Progress Award in 2000, 2003, and 2005 for his research work. He has served as a program committee member for more than 20 international and national conferences.

Wen Gao (M'92–SM'05–F'09) received the M.S. degree in computer science from the Harbin Institute of Technology (HIT), Harbin, China, in 1985, and the Ph.D. degree in electronics engineering from the University of Tokyo, Tokyo, Japan, in 1991.

He was a Professor in computer science at the HIT from 1991 to 1995 and a Professor in computer science at the Institute of Computing Technology of Chinese Academy of Sciences from 1996 to 2005. He is currently a Professor at the School of Electronics Engineering and Computer Science, Peking University, China. He has been leading research efforts to develop systems and technologies for video coding, face recognition, sign language recognition and synthesis, and multimedia retrieval. He has published four books and over 500 technical articles in refereed journals and proceedings in the areas of signal processing, image and video communication, computer vision, multimodal interface, pattern recognition, and bioinformatics.

Prof. Gao earned many awards, including five national awards by research achievements and activities. He has served academic society by being general Co-Chair of IEEE ICME07 and the Head of Chinese delegation to the Moving Picture Expert Group (MPEG) of International Standard Organization (ISO) since 1997. He is also the Chairman of the working group responsible for setting a national Audio Video coding Standard (AVS) for China.