Hierarchical Ensemble of Global and Local Classifiers for Face Recognition

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

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 combines both global and local discriminative features. In this method, global features are extracted from whole face images by Fourier transform and local features are extracted from some spatially partitioned image patches by Gabor wavelet transform. After this, multiple classifiers are obtained by applying Fisher Discriminant Analysis on global Fourier features and local patches of Gabor features. All these classifiers are combined to form a hierarchical ensemble by sum rule. We evaluated the proposed method using Face Recognition Grand Challenge (FRGC) experimental protocols and database known as the largest data sets available. Experimental results on FRGC version 2.0 data set have shown that the proposed method achieves a verification rate of 86%, while the best reported was 76%.

1. Introduction

Face recognition from still and video images has been an active research area due to both its scientific challenge 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 which can be found in the literature surveys [1]. Even though humans can detect and identify faces in a scene with little effort, building an automated system that accomplishes such objectives is very challenging. The challenges mainly come from the large variations in the visual stimulus due to illumination conditions, viewing directions or poses, facial expressions, aging, and disguises such as facial hair, glasses, or cosmetics.

While face representations based on global features, such as Principal Component Analysis (PCA) [2], Linear Discriminant Analysis (LDA) [3], Discrete Fourier Transform (DFT) [4, 5] and Discrete Cosine Transform (DCT) [6], had been popular for face recognition, more recently, there are more and more attempts to develop face recognition systems based on local features. Local features are believed very robust to the variations of facial expression, illumination, and occlusion etc. [7, 8, 9, 10, 11]. Some researchers have compared between global and local features in face recognition. For instance, in [12], B.Heisele et al. reported that component-based system outperforms global system with respect to head pose changes. Much recently, Local Binary Pattern (LBP) [9] and its variant [11] have also achieved very impressive results compared with the methods based on global features. Among the many local features, especially, Gabor wavelets have been recognized as one of the most successful local descriptors for face representation due to their biological relevance and computational properties. The 2D Gabor wavelets [13], whose kernels are similar to the 2D 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 methods based on Gabor features include the Elastic Bunch Graph Matching (EBGM) [8], Gabor Fisher Classifier (GFC) [10] and Local Gabor Binary Pattern (LGBP) [11].

However, in the literature of psychophysics and neurophysiology, many studies [14, 15, 16] have shown that both global and local features are crucial for face perception. Furthermore, global and local features play different roles in the process of face perception and recognition. Global features describe the characteristics of the whole face and they are often used as coarse representation. On the contrary, local features reflect and capture more detailed variations within some local areas in the face. Hence, it is proper to use local features for finer representation.

Following the above studies, it is natural to expect better performance by combining global and local information. In some sense, the well-known Elastic Graph Matching method for face recognition [8] had pioneered such an idea, since global topological information are modeled by the structural of the graph and local features are encoded as the attribute of the nodes. In [17], Fang et al. proposed to combine global features by PCA and component-based local features extracted by Haar wavelets. In [18], Kim et al.

978-1-4244-1631-8/07/$25.00 ©2007 IEEE

proposed an effective face descriptor by decomposing a face image into several components, extracting LDA features from each component, and finally combining these component LDA features together by using a global LDA. In [19], Lee et al. also combined local structures extracted by Local Feature Analysis (LFA) into composite templates which show compromised aspects between kernels of LFA and Eigenfaces. In [20], Kim et al. proposed to combine both global and local features which are obtained by applying Linear Discriminant Analysis (LDA) to either the whole or part of a face image. They experimentally showed that the combined subspace gives smaller Bayesian error than the subspaces of either the global or local features.

In this paper, following the same basic belief to combine global and local features, we propose a novel hierarchical ensemble classifier for face recognition by combining global Fourier features and local Gabor features Specifically, in our method, global features are extracted from whole face image firstly by 2D Discrete Fourier Transform, which is a strong tool to analyze face images in frequency domain [4, 5]. Then, real and imaginary components of low frequency band are concatenated to form a single feature set for further process. For local feature extraction, Gabor wavelet transform is exploited. Firstly, Gabor wavelets are used to extract local features from the whole face image. Then, these features are spatially partitioned into a number of feature sets, each corresponding to a local patch of the face image. After the above processes, a face image can be represented by one Global Fourier Feature Set (GFFS) and multiple Local Gabor Feature Sets (LGFSes). These feature sets contain different discriminative information: GFFS contains global discriminative information and each LGFS contains different local discriminative information. In order to make full use of all these diverse discriminative information, we propose to train multiple component classifiers by applying Fisher Discriminant Analysis (FDA) on GFFS and each LGFS respectively, and then combine them into one ensemble by the weighted sum rule.

Though the proposed method shares the same basic idea to combine global and local features, this paper has made the following distinct contributions:

(1) Unlike previous works combining global and local features, this paper exploits Fourier transform and Gabor wavelets as global and local features respectively. We show experimentally their combination achieves impressively results on FRGC database;
(2) The proposed method is a hierarchical method containing two ensemble procedures: one is local ensemble classifier (LEC) integrating all the local classifiers based on LGFSes; the other is the ensemble of the global classifier and the LEC. Such an two-level hierarchical ensemble strategy leads to impressive generalizability for the face recognition system, since, in machine learning, it is well believed that the ensemble of diverse component classifiers generalizes very well to unseen data.
(3) The proposed method is extensively evaluated on the FRGC 2.0 data set, and exciting results are reported. Especially, on FRGC Exp.4, we have achieved a verification rate of 86% at FAR of 0.1%, while the best known result was 76%.

The remaining part of the paper is organized as follows: in section 2, face representation based on global and local features is introduced. In section 3, we present the construction of hierarchical ensemble classifier. Experiments and analysis are conducted in section 4, followed by conclusion and discussion in the last section.

2. Face Representation by Global and Local Features

As mentioned above, global and local facial features play different roles in face perception, and both of them contain discriminative information for face recognition. Therefore it is necessary to combine them together. Intuitively, local information is embedded in the detailed local variations of the facial appearance, while global information means the overall structural configuration of the facial organs, as well as the face contour. Hence, from the viewpoint of frequency analysis, global features should correspond to the lower frequencies, while the higher frequencies contain more detailed local information. So, in this paper, global information is extracted as the lower frequency band of the Fourier transform, and local information is obtained by using the multi-scale and multi-orientation Gabor wavelets. In some sense, Gabor wavelets can enhance edges, which implies that they extract information of high frequencies.

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

2.1. Different Roles of Global and Local Features

In this subsection, different roles of global and local features are illustrated intuitively using two interesting example images. As shown in Fig.1, the leftmost two input images are artificial with the same components (eye, nose, and mouth) but different external contour, hair and clothes. Therefore, they look globally very different in terms of the overall structural configuration, hair and face contour. Consequently, the classifier based on global features will report them as different persons. However, the classifier by comparing their local components is apt to reporting them the same person, since their components are almost the same. The conflicts 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".

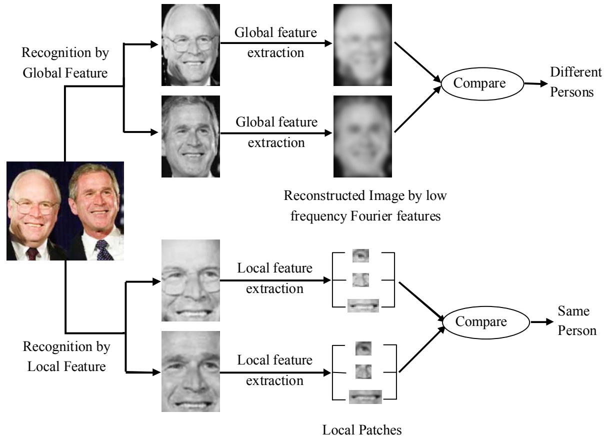


Figure 1: Different roles of global and local features in face perception. See text for detailed explanation.

2.2. Global Fourier Features

2D Discrete Fourier Transform (DFT) is used to extract global facial features. An image can be transformed by 2D DFT into frequency domain as follow:

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

where $f(x,y)$ represents an 2D 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 output is complex, that is

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

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

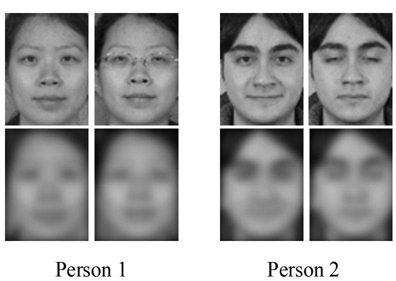


Figure 2: Reconstruction of the input face images by using 30% of the low-frequency Fourier features.

Though all the frequencies contain information about the input image, different bands of frequency play different roles. We know 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 part of the frequency band. Fig.2 gives some examples of inverse transform by using only the lower frequencies (30% of all the energy). From Fig.2, one can safely conclude that the lower frequencies indeed mainly contain information about the globally structural configuration of the facial organs and the contour of the input face. And 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.

Consequently, in our method, only the Fourier features in the low-frequency band are reserved as global features. Specifically, for a face image, we concatenate its real and imaginary components in the low-frequency band into a single feature set, named Global Fourier Feature Set (GFFS). As shown in Fig.3, for both real and imaginary components, only those within the lower frequency band are reserved, as denoted by the white squares in the figure.

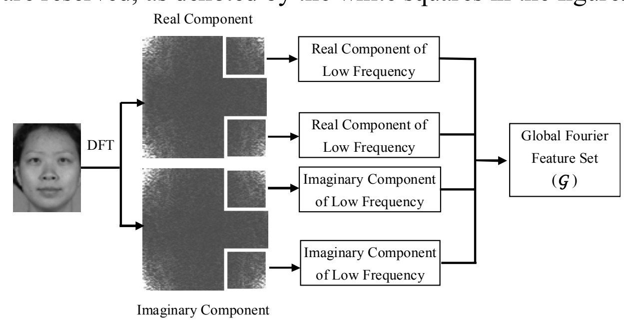


Figure 3: Global Fourier features extraction.

2.3. Local Gabor Features

In recent years, face descriptors based on Gabor wavelets 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. 2D Gabor wavelets are defined as follows [13]:

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

where $k_{u,v}=k_v e^{i\phi_u}$; $k_v=\frac{k_{\max}}{f^v}$ gives the frequency, $\phi_u=\frac{u\pi}{8},\phi_u\in[0,\pi)$ gives the orientation. From the definition, we can see that Gabor wavelet consists of a planar sinusoid multiplied by a two dimensional Gaussian. The sinusoid wave is activated by frequency information in the image. 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 the Gabor wavelet, the frequency information near the center of the Gaussian is captured and frequency information far away from the center of the Gaussian has a negligible effect. Therefore, compared with Fourier transform which extracts the frequency information in the whole face region, Gabor wavelets only focus on some local areas of the face and extract information with multi-frequency and multi-orientation in these local areas.

Gabor wavelets can take a variety of different forms with different scales and orientations. Fig.4 shows 40 Gabor wavelets of 5 scales and 8 orientations. It is obvious that

Gabor wavelets with a certain orientation respond to the edges and bars in this orientation, and Gabor wavelets with a certain scale extract the corresponding frequency information. Hence, Gabor wavelets exhibit desirable characteristics of spatial locality and orientation selectivity. 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.

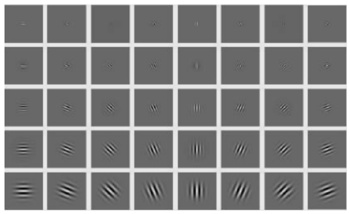


Figure 4: 2D Gabor wavelets of 5 scales and 8 orientations.

As Gabor features are calculated by convolving Gabor wavelets with the whole face image, it covers all the positions of the face image. Thus, the local information provided by the spatial locations of Gabor features is lost when they are integrated to form one single feature vector. In order to reserve more location information, Gabor features are spatially partitioned into a number of feature sets named Local Gabor Feature Set (LGFS), each of which corresponds to a local patch of the face image. In addition, since each LGFS is relatively low dimensional, this can greatly facilitate the sequent feature extraction and pattern classification. Fig.5 illustrates the idea of feature partition and the construction of LGFSes. In Fig.5, N LGFSes, corresponding to N non-overlapping local patches in the face image, are constructed.

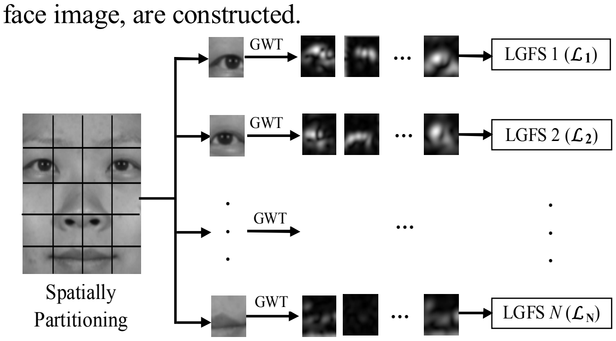


Figure 5: The procedure of LGFS extraction. Please note that, actually in our method, Gabor Wavelet Transform (GWT) is firstly applied to the whole face image, and then the resulting Gabor features are spatially partitioned into N LGFSes.

3. Hierarchical Ensemble Classifier Combining Global and Local Features

After feature extraction, we obtain N+1 feature sets, that is, one GFFS $\mathcal{G}$ and N LGFSes $\mathcal{L}_i$ (i=1,…,N). Then, N+1 classifiers can be trained by applying FDA to each feature set. As explained above, these feature sets contain different discriminant information for face recognition. Hence, the classifiers trained on these feature sets should have large diversity in error. Considering that the ensemble-based classifier is generally superior to the single classifier when the predictions of the component classifiers have enough error diversity, we combine the classifiers trained on each feature set into a hierarchical ensemble to improve the system performance.

The hierarchical ensemble consists of two layers. In the first layer, N Local Component Classifiers (LCCs) C_{L_i} trained on $\mathcal{L}_i$(i=1,…,N) are combined to form a Local Ensemble Classifier (LEC) C_L, which is formulated as follow:

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

where w_{L_i} is the weight of C_{L_i}. In the second layer, LEC C_L is combined with Global Classifier (GC) C_G trained on $\mathcal{G}$ to form the Hierarchical Ensemble Classifier (HEC) C_H, as shown in Eq.5:

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

where w_G is the weight of C_G. As can be seen, in each step, sum rule, the most typical combination rule, is exploited to combine classifiers. The whole hierarchical combination process is shown in Fig.6.

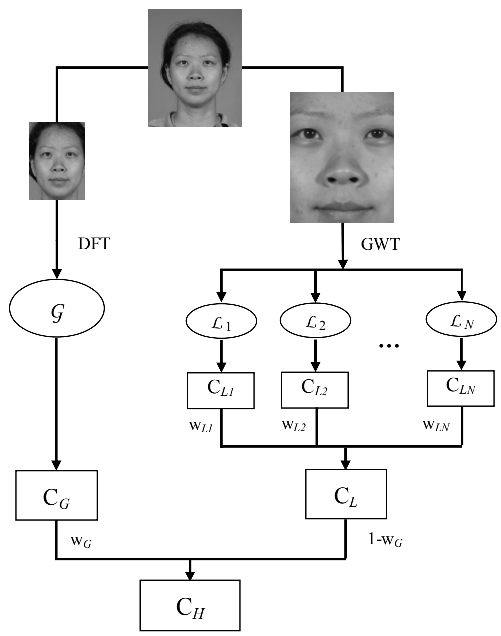


Figure 6: Construction of hierarchical ensemble classifier.

As mentioned above, global features and local features play different roles in face perception. While global features describe the characteristics of the whole face, thus better for coarse representation, local features capture more details in local face areas, thus better for finer representation. Therefore, in our method, global and local features are extracted from normalized face images of different size. As shown in Fig.6, 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 information. On the contrary, the local Gabor features are extracted from the face image of higher resolution, which contains only the internal facial features, i.e. the main facial organs, but excluding the face contour. 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.

4. Experiments

In this section, we validate the proposed method on the FRGC version 2.0 dataset, which is known as the largest face dataset publicly available [21]. Besides the performance of HEC, the performances of the component classifiers in each layer are also given. We also investigate the effect of different weights for the classifier combination. In addition, we compare the performance of our method with the baseline and the best known results.

4.1. FRGC Experimental Protocols

The experiments in FRGC version 2.0 are designed to advance face recognition in general with emphasis on 3D and high resolution still imagery. In our experiments, only the still images are considered. Some example face images in FRGC data set are shown in Fig.7.

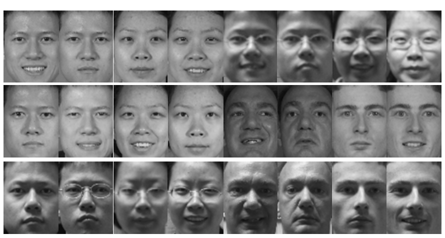


Figure 7: Example face images in the FRGC data set. The top row shows training images: the first four images are controlled and the remaining four are uncontrolled. The middle row displays controlled target images and the bottom row displays uncontrolled query images.

FRGC provides six experimental protocols among which Experiment 1, 2 and 4 are designed for still images. The training set for still images experiments consists of 12,776 images of 222 individuals. We use this training set for training FDA and take Experiment 1 and 4 for evaluations. While Experiment 1 measures performance on 16,028 frontal facial images taken under controlled illumination, Experiment 4 is designed to measure performance on 8,014 uncontrolled query images versus 16,028 controlled target images. Configurations of the two experiments are summarized in Table 1. Note that, Experiment 4 is more challenging because of serious illumination changes, blurring effect, and partial occlusions.

TABLE 1

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

Database sizes in FRGC experiments. [C] and [U] mean controlled and uncontrolled illumination condition, respectively.

The performance is reported as Verification Rates (VR) at 0.1% False Acceptance Rate (FAR). In addition, for each experiment, three Receiving Operator Characteristic (ROC) curves are generated by BEE system. ROC I is corresponding to the images collected within semesters, ROC II within a year, and ROC III between semesters.

4.2. Performances of Global and Local Classifiers

In our experiments, face images are aligned according to the manually located eye positions. For Fourier feature extraction, the face image is normalized to 64 by 80 pixels with the eye center distance being 28 pixels. In order to apply FFT, image must be extended to 128 by 128. So, 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 practice, keeping how many low-frequency features can be determined by checking the percentage of the reserved energy. In this paper, about 50% of the low-frequency energy (bandwidth is 16) is reserved for constructing GC. Thus, referring to Fig.3, the dimension of the GFFS is 16x16x4=1024.

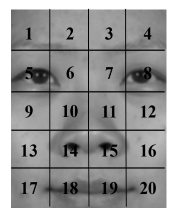
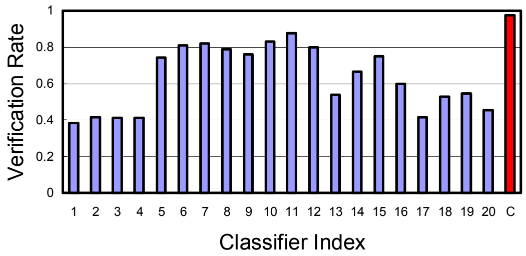
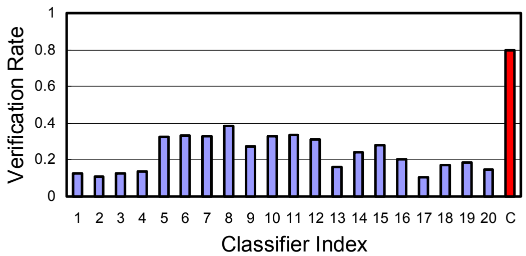


Figure 8: Indexes of the LCC.

For Gabor features, the size of the face image is 128 by 160 pixels with the eye distance being 72 pixels. After Gabor transformation on this "fine" face image, the magnitudes of the resulting Gabor coefficients are spatially partitioned into 20 non-overlapping patches of size 32 by 32 pixels, as shown in Fig.8. Thus, 20 LGFSes can be obtained. In our system, 40 Gabor wavelets (5 scales and 8 orientations) are used with the same parameters in [8], so the dimension of each LGFS is 32x32x5x8=40,960. As it is very difficult for FDA to deal with that high dimensionality, the features are uniformly down-sampled by averaging the Gabor features in an 8x8 grid. So the dimension of each LGFS is reduced from 40,960 to 640 (=4x4x5x8), which is then further processed by FDA. Finally, 20 LCCs are obtained and combined together to form the LEC.



(a) Results on FRGC Experiment 1

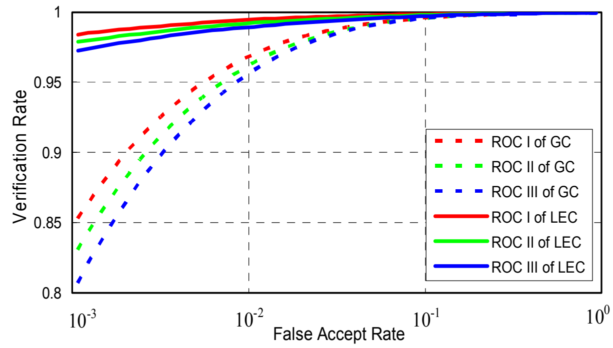


(b) Results on FRGC Experiment 4

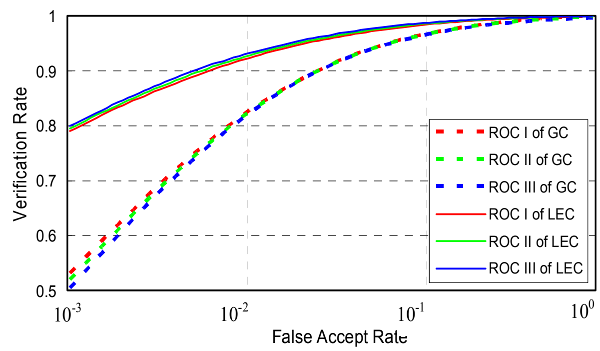
Figure 9: Performances of ROC III for both LCCs and LECs in Experiment 1 and 4. In (a) and (b), 1-20 are the indexes of the LCCs whose positions are shown in Fig.8. C represents LEC.

Figure 9 (a) and (b) give the performances of each of the 20 LCCs, as well as the performance of their ensemble, i.e. LEC. From Fig.9, one can see that LCCs in different positions have quite different discriminant capacities, which is intuitively reasonable. And, it is easy to understand that the LCCs located in the eye, nose and mouth area have relative better discriminant capacities. In addition, since each LCC exploits only part of the information in different facial regions, they may misclassify different patterns; therefore their performances are commonly not good enough. However, attribute to their diversity in prediction error, they should be mutually complementary. Hence, their combination, i.e. LEC, should greatly outperform any of them, which can be clearly observed in Fig.9 (a) and (b). The improvement is especially significant for Experiment 4: the best verification rate of LCCs is 38.5%, while that of LEC is as high as 79.9%. In our experiments, we notice that, in our case, weighting different LCCs has only very trivial effect on the performance of the LEC. Therefore, equal weights are assigned to all the LCCs for simplicity.

In Fig.10, three ROC curves of both GC and LEC are given. It is obvious that local features have much better discriminative ability than global features. Note that, both GC and LEC performs much better than FRGC baseline algorithm (basically PCA).



(a) Results on FRGC Experiments 1



(b) Results on FRGC Experiments 4

Figure 10: ROC curves of GC and LEC on Experiment 1 (a) and Experiment 4 (b).

4.3. Performance of Hierarchical Ensemble Classifier

In order to make full use of both global and local discriminant information and further improve the performance, GC and LEC are combined to form a unified ensemble classifier (HEC), as formulated in Eq.5. In Eq.5, the weight for GC W_G can actually balance the importance of global and local information. This is evidently necessary because we have noticed that the performances of GC and LEC are quite different, as can be seen from Fig.10. And the performance of GC is relatively worse than LEC. So, it is natural to assign a smaller weight for GC.

In this paper, taking FRGC Experiment 4 (ROC III) as example, experiments are conducted to check the influence of W_G on the performance of the HEC. How the performance changes with the varying W_G is shown in Fig.11. From this figure, we know that, at least for FRGC Experiment 4, the best result appears when W_G is about 0.2. Though, this parameter is not necessarily a generalized good setting for any database, at least it illustrates that the local features should be more emphasized than the global features. More importantly, another conclusion we can draw is that the combination of global and local features can further improve the recognition performance. For instance, in Experiment 4 (ROC III), the VR of GC is only

50.7% and VR of LEC is 79.9%. But, their combination with W_G=0.2 achieves a VR of 85.8%, which shows that global and local features are indeed mutually complementary.

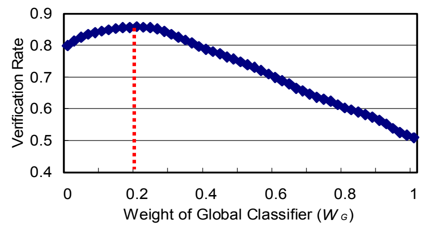
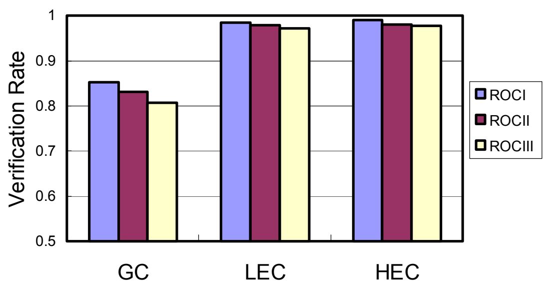
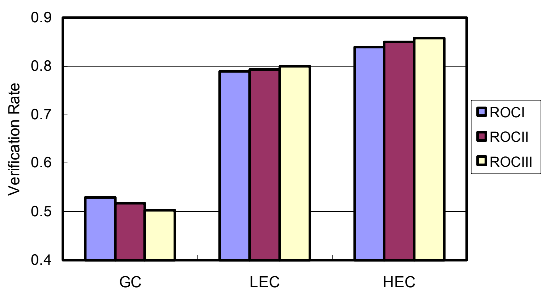


Figure 11: The effect of different W_G on the performance of HEC on Experiment 4 (ROCIII).

In Fig.12, we show three ROC performances of GC, LEC and HEC on both Experiment 1 and 4.



(a) Results on FRGC Exp.1



(b) Results on FRGC Exp.4

Figure12: Three ROC performances of GC, LEC and HEC on Experiment 1 (a) and 4 (b).

We also compare our method with the FRGC baseline algorithm (basically PCA) and the best known results [4, 22] in Experiment 1 and 4, as shown in Fig.13 and Table 2. In [4], 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 by LDA and they are merged. In [22], Liu presented a pattern recognition framework which integrates Gabor image representation, multi-class Kernel Fisher Analysis (KFA) using fractional power polynomial models for improving FRGC performance. So far, the results in [4] and [22] are the reported best results on FRGC data set.

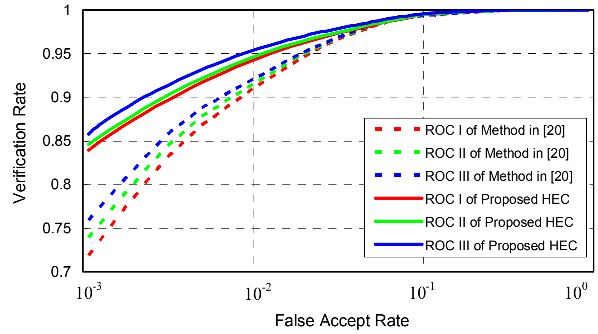


Figure 13: ROC performances comparison between our method and Liu's method in [22] on Experiment 4.

TABLE 2

Method		Exp.1	Exp.4
FRGC Baseline		66%	12%
Method in [4]		91%	74%
Method in [22]		92%	76%
Our Methods	GC	81%	51%
	LEC	97%	80%
	HEC	98%	86%

Performances comparison on Experiment 1 and 4 of FRGC data set (ROC III).

From Table 2, one can see that the proposed method has further improved the verification rates on FRGC especially on Exp.4. Taking ROC III as an example, on Exp.4, a verification rate of 86% is achieved, 10 percents higher than the best known results. We also notice that, the local ensemble classifier itself also outperforms the best known results on both experiments. These comparisons show that the proposed method achieves significant improvement on FRGC Exp.1 and Exp.4, especially attribute to the combination of global and local features expressed by Fourier and Gabor filters respectively.

5. Conclusion and Discussion

We human beings recognize faces relying on both global face features and local details of the facial organs. A hierarchical ensemble of global and local classifiers is proposed to simulate the observations in bionic sense by exploiting both global features and local features. In the proposed method, global features are extracted from whole face images by Fourier transform, and local features are extracted from some spatially partitioned image patches by Gabor wavelet transform. By applying FDA on Fourier features and Gabor feature patches, multiple classifiers are obtained and then combined into a hierarchical ensemble classifier by sum rule. We validate our method on FRGC version 2.0 data set designed for face identification. Experimental results show that the ensemble classifier greatly outperforms its component classifiers which have large error diversity. By the proposed method, we have achieved verification rates of 98% in Experiment 1 and

86% in Experiment 4 respectively. Compared with the baseline and best known results, the proposed method demonstrates significant improvement especially on Experiment 4.

The success of the proposed method comes from several aspects. First of all, we should mention the ensemble process in the method. Ensemble learning has been widely recognized as an important method with excellent generalizability. In our method, ensemble lies in two stages: the combination of local classifiers, and the combination of the global and local classifiers. Both ensemble procedures improve impressively the performance of the component classifiers. Another critical success point is of course the usage of both global and local features extracted by Fourier and Gabor respectively. Especially, the local features themselves based on Gabor filtering can achieve excellent performance better than the best known results.

Though we have shown that the Fourier-based global features are not as efficient as the Gabor-based local features, we must point out that it is still too early to conclude that global features are less important than local features. Therefore, it is one of our future efforts to study better representation methods for global feature. In our opinion, more overall structure information should be considered.

6. Acknowledgements

This paper is partially supported by National Natural Science Foundation of China under contract No.60332010 and No.60673091, Hi-Tech Research and Development Program of China under contract No.2006AA01Z122 and No.2007AA01Z163, 100 Talents Program of CAS, Program of New Century Excellent Talents in University (NCET-04-0320), and ISVISION Technology Co. Ltd.

7. References

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

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

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

[4] W. Hwang, G. Park, J. Lee and S.C.Kee, Multiple Face Model of Hybrid Fourier Feature for Large Face Image Set, *International Conference on Computer Vision and Pattern Recognition,* pp.1574-1581, 2006.

[5] J. H. Lai, P. C. Yuen and G. C. Feng, Face Recognition Using Holistic Fourier Invariant Features, *Pattern Recognition*, vol. 34, no.1, pp. 95-109, 2001.

[6] 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, *International Conference on Acoustics, Speech and Signal Processing*, 2006.

[7] P. S. Penev, J. J. Atick, Local Feature Analysis: a General Statistical Theory for Object Representation, *Network: Computation in Neural Systems*, vol. 7, pp.477-500, 1996.

[8] L. Wiskott, J. M. Fellous, N. Kruger, C. Malsburg, Face Recognition by Elastic Bunch Graph Matching, *IEEE Transaction on. Pattern Analysis and Machine Intelligence*, vol.19, no. 7, pp.775-779, 1997.

[9] A. Timo, H. Abdenour and P. Matti, Face Recognition with Local Binary Patterns, *European Conference on Computer Vision,* pp.469-481, 2004.

[10] C. Liu and H. Wechsler, Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition, *IEEE Transaction on Image Processing*, vol. 11, no. 4, pp. 467-476, 2002.

[11] 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, *International Conference on Computer Vision*, pp.786-791, 2005.

[12] B. Heisele, P. Ho, T. Poggio, Face Recognition with Support Vector Machines: Global versus Component-based Approach, *International Conference on Computer Vision,* vol.2, pp.688-693, 2001.

[13] J. G. Daugman, Two-Dimensional Spectral Analysis of Cortical Receptive Field Profile, *Vision Research*, vol. 20, pp. 847-856, 1980.

[14] V. Bruce, *Recognizing Faces*, London:Erlbaum, 1988.

[15] G. Davies, H. Ellis, and E.J. Shepherd, *Perceiving and Remembering Faces*, New York: Academic, 1981.

[16] H. Ellis, M. Jeeves, F. Newcombe, and A. Young, *Aspects of Face Processing*, Dordrecht: Nijhoff, 1986.

[17] Y. Fang, T. Tan and Y. Wang, Fusion of Global and Local Features for Face Verification, *International Conference on Pattern Recognition,*vol.2, pp.382-385, 2002.

[18] T. Kim, H. Kim, W. Hwang and J. Kittler, Component-based LDA Face Description for Image Retrieval and MPEG-7 Standardisation, *Image and Vision Computing*, vol.23, no.7, pp.631-642, 2005.

[19] Y. Lee, K. Lee and S. Pan, Local and Global Feature Extraction for Face Recognition, *International Conference on Audio- and Video-Based Biometric Person Authentication,* pp.219-228, 2005.

[20] C. Kim, J. Oh, and C. Choi, Combined Subspace Method Using Global and Local Features for Face Recognition, *International Joint Conference on Neural Networks*, 2005.

[21] P. J. Phillips, P. J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, Overview of the Face Recognition Grand Challenge, *International Conference on Computer Vision and Pattern Recognition,* pp.947-954, 2005.

[22] C. Liu, Capitalize on dimensionality increasing techniques for improving face recognition performance, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.28, No. 5, pp.725-737, 2006.