Combining Local and Global Features for Face Recognition

Xilin Chen[1], Shiguang Shan[1], Yu Su[2], Wenchao Zhang[2], Baochang Zhang[2], Wen Gao[1,2]
[1]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100080, China
[2]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, 150001, China

Abstract: In the literature of psychophysics and neurophysiology, many studies have shown that both global and local features are crucial for face recognition. In this paper, we summarize our work on face recognition by combining both local and global discriminative features. In our work, Fourier transform is exploited to extract global features from the whole face image domain. For local feature, Gabor wavelets and its combination with Local Binary Patterns (LBP) are explored, which are both conducted on some spatially partitioned image patches. These features are then fed into different classifiers based on Fisher Discriminant Analysis respectively, and these classifiers are combined together to make the final decision. The proposed methods are evaluated by using Face Recognition Grand Challenge (FRGC) experimental protocols and database, the largest data sets available. Experimental results on FRGC version 2.0 dataset show that our methods have much higher verification rates than the baseline of FRGC and the best known results under various situations such as illumination changes, expression changes, and time elapses.

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 [Eigenface, Fisherface, Holistic Fourier] had been popular for face recognition, more recently, there are more and more attempts to develop face recognition systems based on local descriptors. Local features are believed very robust to the variations of facial expression, illumination, occlusion and so on [2, 3, 4, 5, 6]. Some researchers have compared between global and local features in face recognition, for instance, in [14], B.Heisele et al. reported that component-based system outperforms global system with respect to head pose changes. Much recently, LBP [4] and its variant [6] have also achieved very impressive results compared with methods based on global features. Among 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 [7, 8], 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) [3], Gabor Fisher Classifier (GFC) [5] and Local Gabor Binary Pattern (LGBP) [6].

However, in the literature of psychophysics and neurophysiology, many studies such as in [9, 10, 11] have shown that both global and local features are crucial for face perception. Global features and local features play different roles in the process of face perception and recognition. Global features can describe the characteristic of the whole face and they are often used as coarse representation. Compared with global features, 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 [3] 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 [13], Fang et al. proposed to combine global features by PCA and component-based local features extracted by Haar wavelets. In [14], Kim et al. 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 [17], 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 [18], 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.

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 images by 2D Discrete Fourier Transform, which is a strong tool to analyze face images in frequency domain [17, 18]. Then, real and imaginary components of low frequency band are concatenated to form a single feature set for further process as in [17]. For local feature extraction, Gabor Wavelet Transform is exploited. Firstly, a face image is spatially partitioned into a number of patches of equal size. Then, Gabor wavelets are used to extract local features within each image patches, forming multiple sets of Gabor features with each feature set corresponding to an image patch. 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 discriminant information: GFFS contains global discriminant information and each LGFS contains different local discriminant information. In order to make full use of all these diverse discriminant 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 weighted sum rule.

2. Global features [22]

2D Discrete Fourier Transform (DFT) is very useful in image processing because there are many things about an image can be revealed by frequency space analysis but are not obvious from the original intensities of the pixels.

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 $R(u,v)$ 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.

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 overall structural configuration of the input image, while those higher frequencies correspond to detailed local variations. This can be illustrated intuitively by observing the effects of inverse transform with part of the frequency band. Fig.1 gives some examples of inverse transform by using only the low frequencies (30% of all the energy). From Fig. 1, one can safely conclude that low frequencies indeed mainly contain globally structural configuration of the facial organs and the contour. And it is also apparent that these low-frequency features are robust to detailed local variations due to facial expressions.

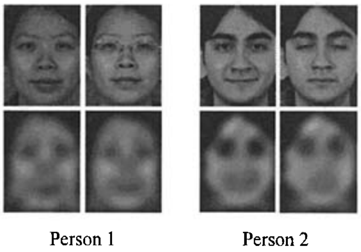


Figure 1: Reconstruction of input face images by using 30% of low-frequency Fourier features.

Consequently, in our method, only the Fourier features in low-frequency band are reserved as global features. Specifically, for a face image, we concatenate its real and imaginary components in low-frequency band into a single feature set, named Global Fourier Feature Set (GFFS).As shown in Fig.2, for both real and imaginary components, only those within low frequency band as denoted by the white squares in the figure are reserved.

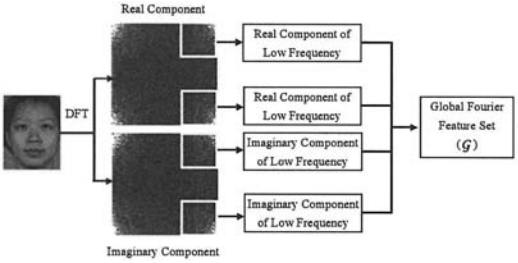

Real Component
DFT
Real Component of Low Frequency
Real Component of Low Frequency
Imaginary Component of Low Frequency
Imaginary Component of Low Frequency
Global Fourier Feature Set (𝒢)
Imaginary Component

Figure 2: Global Fourier features extraction.

3. Local features

3.1. 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 [8]:

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

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 such as different scales and orientations. Fig.3 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.

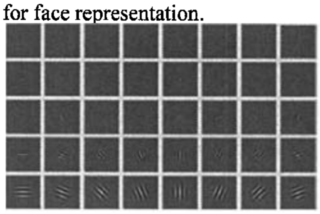


Figure 3: 2D Gabor wavelets of 5 scales and 8 orientations.

As multi-scale and multi-orientation Gabor wavelets are used to convolve with face images in process of feature extraction, the dimension of Gabor features is very high. For example, if we use 40 Gabor wavelets with 5 scales and 8 orientations, the dimension of Gabor features is 40 times of the original dimension of face image. Moreover, these Gabor features cover all the positions of face image. In order to reduce these high dimensional Gabor features and restrict them to cover only some local areas, we propose to spatially partition a face image into a number of patches. From each image patch, multi-scale and multi-orientation Gabor features are concatenated into a Local Gabor Feature Set. As shown in Fig.4, we partition a face image into N non-overlapping patches of equal size and N LGFSes are obtained for further classification.

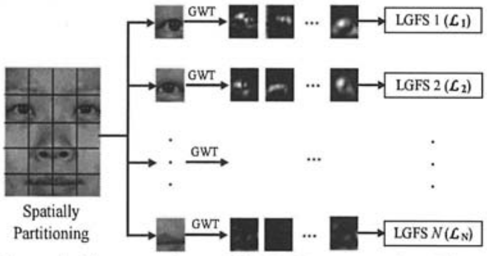


Figure 4: The procedure of N LGFSes extraction. Please note that, actually in our method, Gabor Wavelet Transform (GWT) is firstly applied to the whole face, and then the obtained Gabor features are spatially partitioned to form N LGFSes.

3.2. Local Gabor Binary Patterns (LGBP) [6]

Considering the advantages of the Gabor filters in face recognition, we exploit the multi-resolution and multi-orientation Gabor filters to de-composite the input face images for sequential feature extraction.

The Gabor representation of a face image is derived by convolving the face image with the Gabor filters. Let $f(x,y)$ be the face image, its convolution with a Gabor filter $\psi_{\mu,\nu}(z)$ is defined as follows

$$G_{\psi f}(x,y,\mu,\nu)=f(x,y)*\psi_{\mu,\nu}(z) \tag{4}$$

where $*$ denotes the convolution operator. Five scales $\nu\in\{0,\cdots,4\}$ and eight orientations $\mu\in\{0,\cdots,7\}$ Gabor filters are used. Convolving the image with each of the 40 Gabor filters can then generate the Gabor features. Note that, because the phase information of the transform is time-varying, generally, only its magnitude is explored. Thus, for each Gabor filter, one magnitude value will be computed at each pixel position, which will totally result in 40 Gabor Magnitude Pictures (GMPs).

The magnitude values of the Gabor transform change very slowly with displacement [3], so they can be further encoded. In order to enhance the information in the GMPs, we encode the magnitude values with LBP operator. The original LBP operator [9] labels the pixels of an image by thresholding the 3×3-neighborhood of each pixel $f_p\,(p=0,1,\ldots,7)$ with the center value f_c and considering the result as a binary number [21]

$$S(f_p-f_c)=\begin{cases}1, & f_p\ge f_c\\ 0, & f_p<f_c\end{cases} \tag{5}$$

Then, by assigning a binomial factor 2^p for each $S(f_p-f_c)$, the LBP pattern at the pixel is achieved as

$$LBP=\sum_{p=0}^{7}S(f_p-f_c)2^p \tag{6}$$

which characterizes the spatial structure of the local image texture. The operator *LGBP* denotes the LBP operates on GMP. We denote the transform result at position (x,y) of (μ,ν)-GMP as $G_{\mathrm{lgbp}}(x,y,\mu,\nu)$, which composes the (μ,ν)*-LGBP Map.*

Face recognition under varying imaging conditions such as illumination and expression is a very difficult problem. Usually, the variations will appear more on some specific regions in face image. Therefore, we exploit local feature histogram to summarize the region property of the LGBP patterns by the following procedure: Firstly, each *LGBP Map* is spatially divided into multiple non-overlapping regions. Then, histogram is extracted from each region. Finally, all the histograms estimated from the regions of all the *LGBP Maps* are concatenated into a single histogram sequence to represent the given face image. The above process is formulated as follows:

The histogram $\mathbf{h}$ of an image $f(x,y)$ with gray levels in the range $[0,L-1]$ could be defined as

$$\mathbf{h}_i=\sum_{x,y}I\{f(x,y)=i\},i=0,1,\ldots,L-1 \tag{7}$$

where i is the i-th gray level, $\mathbf{h}_i$ is the number of pixels in the image with gray level i and

$$I\{A\}=\begin{cases}1, & A \text{ is true}\\ 0, & A \text{ is false}\end{cases} \tag{8}$$

Assume each *LGBP Map* is divided into m regions $R_0,R_1,\ldots,R_{m-1}$. The histogram of r-th region of the specific *LGBP Map* (from (μ,ν)-GMP) is computed by

$$\mathrm{H}_{\mu,\nu,r}=(\mathbf{h}_{\mu,\nu,\mathrm{r},0},\mathbf{h}_{\mu,\nu,\mathrm{r},1},\cdots,\mathbf{h}_{\mu,\nu,\mathrm{r},\mathrm{L}\text{-}1}) \tag{9}$$

where

$$\mathbf{h}_{\mu,\nu,\mathrm{r},\mathrm{i}}=\sum_{(x,y)\in\mathrm{R}_\mathrm{r}}I\{G_{\mathrm{lgbp}}(x,y,\mu,\nu)=i\} \tag{10}$$

Finally, all the histogram pieces computed from the regions of all the 40 *LGBP Maps* are concatenated to a histogram sequence, $\Re$, as the final face representation $\Re = (H_{0,0,0}, \cdots, H_{0,0,m-1}, H_{0,1,0}, \cdots, H_{0,1,m-1}, \cdots, H_{7,4,m-1})$.

4. Combining global and local features

After feature extraction, we obtain N+1 feature sets including one GFFS G and N LGFSes L_i (i=1,...,N). Then, N+1 classifiers can be trained by applying FDA on each feature set. As explained above, these feature sets contain different discriminant information for face recognition. Hence, classifiers trained on these feature sets should have large error diversity. Considering that the ensemble-based classifier is generally superior to the single classifier when the predictions of the component classifiers has enough error diversity, we combine 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 C_{L_i} trained on L_i (i=1,...,N) are combined to form a local ensemble classifier C_L, which is formulated as follow:

$$C_L = \sum_{i=1}^{N} w_{L_i} \cdot C_{L_i} \tag{11}$$

where w_{L_i} is the weight of C_{L_i}. In the second layer, the local ensemble classifier C_L is combined with the global classifier C_G trained on G to form the final ensemble classifier C_F, as shown in Eq.12:

$$C_F = w_G C_G + (1 - w_G) C_L \tag{12}$$

where w_G is the weight of C_G.

In each step, sum rule, the most typical combination rule, is exploited to combine classifiers. This hierarchical combination process is shown in Fig.5.

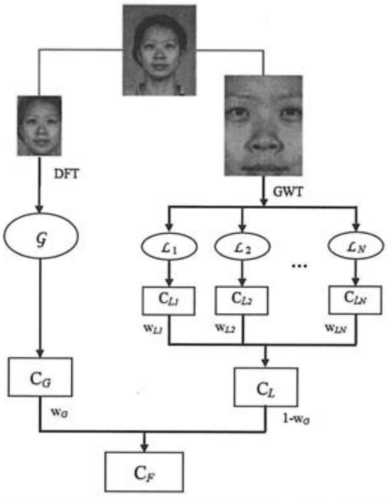


Figure 5: Combination of global and local features.

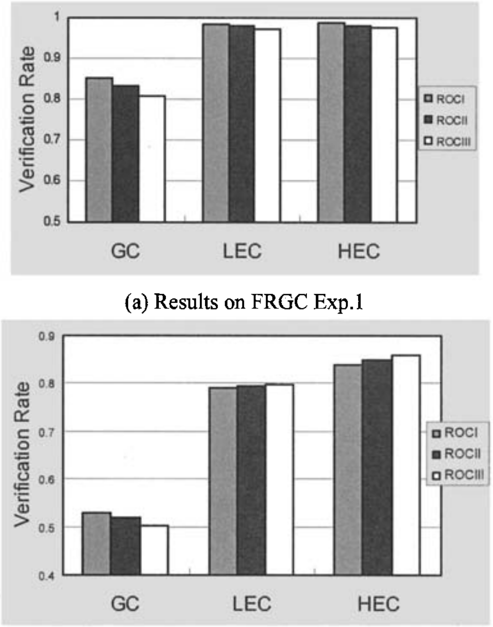
5. Experiments

In order to make full use of both global and local discriminant information and further improve the performance, global classifier and local ensemble classifiers are combined to form a unified ensemble classifier, as formulated in Eq.5. In Eq.5, the weight for global classifier W_G can actually balance the importance of global and local information. This is evidently necessary because we have noticed that the performances of global classifier and local classifier are quite different, as can be seen from Fig.9. And the performance of global classifier is relatively worse than local ensemble classifier. So, it is natural to assign a smaller weight for global classifier.

Taking FRGC Experiment 4 (ROC III) as example, experiments are conducted to check the influence of the weight for the global classifier on the performance of the final classifier. 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 necessary 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.

In Figure.6, we show three ROC performances of global classifier, local ensemble classifier and unified ensemble classifiers on both Experiment 1 and 4.

We also compare our method with the FRGC baseline algorithm (basically PCA) and the best known results [17, 20] in Experiment 1 and 4, as shown in Fig.7 and Table 1. In [17], Hwang etc. 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 [20], 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 [17] and [20] are the reported best results on FRGC dataset.



(a) Results on FRGC Exp.1

(b) Results on FRGC Exp.4

Figure 6: Three ROC performances of global classifier (GC), local ensemble classifier (LEC), and unified hierarchical ensemble classifiers (HEC) on Experiment 1 (a) and 4 (b).

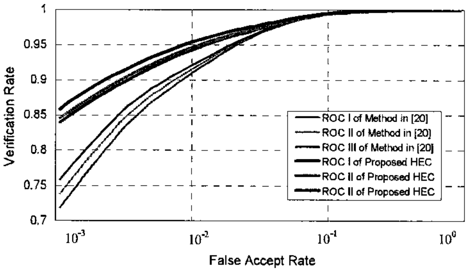


Figure 7: ROC performances comparison between our method and Liu's method in [20] on Experiment 4.

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

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

Methods		Exp.1	Exp.4
FRGC Baseline [19]		66%	12%
Method in [17]		91%	74%
Method in [20]		92%	76%
Our Methods [22]	GC	81%	51%
	LEC	97%	80%
	HEC	98%	86%

6. Conclusions and future work

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 dataset designed for face verification. 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 known best 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, since ensemble learning has been well 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 expressed by Fourier and Gabor respectively. Especially, the local features themselves based on Gabor filtering can achieve excellent performance better than the known best results.

Acknowledgements

The authors would like to thank OMRON for sponsorship of the talk. This research is partially sponsored by Natural Science Foundation of China under contract No. 60332010 and 60533030, the 100 Talents Program of CAS.

References

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

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

[3] L. Wiskott, J.M. Fellous, N. Kruger, C. Malsburg, Face Recognition by Elastic Bunch Graph Matching, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7), pp. 775-779, 1997.

[4] A. Timo, H. Abdenour, and P. Matti, "Face Recognition with Local Binary Patterns", *Euro. Conf. on Computer Vision*, pp. 469-481, 2004.

[5] C. Liu and H. Wechsler, "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition", *IEEE Trans. on Image Processing*, 11(4), pp. 467-476, 2002.

[6] 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", *Int' l Conf. on Computer Vision*, pp.786-791, 2005.

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

[8] J.G. Daugman, "Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two-Dimensional Visual Cortical Filters," *J. Optical Soc. Amer.*, 2(7), pp. 1160-1169, 1985.

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

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

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

[12] B.Heisele, P.Ho, T.Poggio, "Face Recognition with Support Vector Machines: Global versus Component-based Approach", *Int'l Conf. on Computer Vision*, 2, pp. 688-693, 2001

[13] Y.Fang, T.Tan, Y.Wang, "Fusion of Global and Local Features for Face Verification", *Int'l Conf. on Pattern Recognition,* 2, pp.382-385, 2002

[14] 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, 23(7), pp.631-642, 2005.

[15] Y.Lee, K.Lee and S.Pan, "Local and Global Feature Extraction for Face Recognition", *Int'l Conf. on Audio- and Video-Based Biometric Person Authentication,* pp.219-228, 2005.

[16] C.Kim, J.Oh, and C.Choi, "Combined Subspace Method Using Global and Local Features for Face Recognition", *Int'l. J. Conf. on Neural Networks,* 2005

[17] W. Hwang, G. Park and J. Lee, "Multiple Face Model of Hybrid Fourier Feature for Large Face Image Set", *Int'l Conf. on Computer Vision and Pattern Recognition*, pp.1574-1581, 2006.

[18] J.H. Lai, P.C. Yuen and G.C. Feng, "Face Recognition Using Holistic Fourier Invariant

Features", *Pattern Recognition*, 34(1), pp. 95-109, 2001.

[19] 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", *Int'l Conf. on Computer Vision and Pattern Recognition, CVPR*, pp.947-954, 2005.

[20] C. Liu, "Capitalize on dimensionality increasing techniques for improving face recognition performance", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(5), pp.725-737, 2006.

[21] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(7), pp. 971-987, 2002.

[22] Yu Su, Shiguang Shan, Xilin Chen, Wen Gao. Hierarchical Ensemble of Global and Local Classifiers for Face Recognition. Accepted by Int'l Conf. on Computer Vision 2007.