Front. Comput. Sci., 2018, 12(6): 1173–1191
https://doi.org/10.1007/s11704-017-6275-6

RESEARCH ARTICLE

Fusing magnitude and phase features with multiple face models for robust face recognition

Yan LI[1,2], Shiguang SHAN (✉)[1,2], Ruiping WANG[1,2], Zhen CUI[3], Xilin CHEN[1,2]

1 Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS),
Institute of Computing Technology (ICT), CAS, Beijing 100190, China
2 University of Chinese Academy of Sciences, Beijing 100049, China
3 School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

Abstract High accuracy face recognition is of great importance for a wide variety of real-world applications. Although significant progress has been made in the last decades, fully automatic face recognition systems have not yet approached the goal of surpassing the human vision system, even in controlled conditions. In this paper, we propose an approach for robust face recognition by fusing two complementary features: one is Gabor magnitude of multiple scales and orientations and the other is Fourier phase encoded by spatial pyramid based local phase quantization (SPLPQ). To reduce the high dimensionality of both features, block-wise fisher discriminant analysis (BFDA) is applied and further combined by score-level fusion. Moreover, inspired by the biological cognitive mechanism, multiple face models are exploited to further boost the robustness of the proposed approach. We evaluate the proposed approach on three challenging databases, i.e., FRGC ver2.0, LFW, and CFW-p, that address two face classification scenarios, i.e., verification and identification. Experimental results consistently exhibit the complementarity of the two features and the performance boost gained by the multiple face models. The proposed approach achieved approximately 96% verification rate when FAR was 0.1% on FRGC ver2.0 Exp.4, impressively surpassing all the best known results.

Received May 16, 2016; accepted February 21, 2017

E-mail: yan.li@vipl.ict.ac.cn, sgshan@ict.ac.cn, ruiping.wang@vipl.ict.ac.cn, zhen.cui@vipl.ict.ac.cn, xlchen@ict.ac.cn

Keywords face recognition, fisher discriminant analysis, fusion, Gabor magnitude feature, multiple face models, spatial pyramid based local phase quantization

1 Introduction

Face recognition, as one of the most representative technologies of artificial intelligence, has attracted significant attention over the last decades in many domains including information security, law enforcement, surveillance, and entertainment [1]. Although numerous approaches [2–11] have been proposed and tremendous progress has been made, it remains a challenge for machines to recognize human faces efficiently and accurately under uncontrolled conditions. The main challenges are in the small interpersonal differences caused by similar facial configurations and the significant intrapersonal variations caused by diverse extrinsic imaging factors such as head pose, expression, aging, and illumination.

In general, a typical face recognition system consists of three modules: face detection, face representation, and face classification. In this paper, we focus mainly on the second part, i.e., extracting the internal representation which is considered as the key to high accuracy face recognition. Numerous local descriptors have been proposed for effective face representation. One of the most successful local descriptors is Gabor wavelet transform which is first proposed by Gabor [12] aiming at analyzing signals, and then successfully being extended to the problem of face recognition. Gabor wavelets, 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 therefore achieve higher top-one recognition accuracy [2]. However, the majority of the proposed Gabor-related approaches (e.g., [2–4]) utilize only the magnitude information, and only a small number of approaches utilize the phase information because of the sensitivity of phase to varying positions, which leads to severe problems when matching two faces with a slight misalignment [5]. To investigate the potential of phase information, Ojansivu and Heikkilä [13] proposed a novel descriptor named local phase quantization (LPQ) for texture classification by utilizing the Fourier phase information computed locally in a window for each image pixel. LPQ provides its robustness to image blurring and insensitiveness to uniform illumination changes. Local binary patterns (LBP) [14] is another powerful local descriptor; it consumes less extraction time and has lower-dimensional representation compared with Gabor wavelet transform. This descriptor assigns a label to every pixel of an image by thresholding the 3 × 3 neighborhood pixels with the center pixel value and considering the result as a binary number (binary pattern). Then, the histogram of the labels can be used as feature. Lowe [15] proposed an approach for extracting distinctive invariant features from images that can be used to perform reliable matching among different views of an object or scene, named scale invariant feature transform (SIFT). SIFT features are invariant to image scale and rotation, and can provide robust matching across a substantial range of affine distortion, change in 3D viewpoint, illumination, and addition of noise. Although SIFT is not originally designed for face recognition, it exhibits excellent performance in face recognition applications (e.g., [16–18]). As with SIFT, Histograms of Oriented Gradients (HOG) [19] was originally devised for human detection rather than face recognition. It is based on evaluating well-normalized local histograms of image gradient orientations in a dense grid with the basic idea that local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients or edge directions, even without precise knowledge of the corresponding gradient or edge positions. Several works (e.g., [20, 21]) have proven its successful extension to face recognition.

However, a single descriptor only encodes limited information of the given face [6]; it is reasonable to combine different descriptors for more effective face representation. Recently, approaches fusing diverse descriptors have received considerable attention. Zhang et al. [3] proposed a non-statistics based face representation approach, named local Gabor binary pattern histogram sequence (LGBPHS). In this approach, an input face image is first transformed to obtain multiple Gabor magnitude pictures (GMPs), and then the GMPs are converted to local Gabor binary pattern (LGBP) maps by the LBP operator and further divided into non-overlapping regions in which histograms are computed. Tan and Triggs [7] proposed another effective approach by combining Gabor and LBP features, then further applying kernel discriminative common vector method to nonlinearly extract discriminant feature. Inspired by the fact that human beings recognize faces relying on both global and local facial features, a hierarchical ensemble approach is proposed by Su et al. [4] to simulate the observations in bionic sense by exploiting both global and local features, where the global part is extracted from the whole face image by using Fourier transform, and the local part is extracted from some spatially divided face patches by using Gabor wavelets. Further, from the basic idea of integrating global and local information, Liu et al. [22] proposed a fusion approach by combining Gabor wavelets, multi-resolution LBP, and discrete cosine transform (DCT) in a novel hybrid color space to boost face recognition performance. Chan et al. [8] integrated multiscale LBP (MLBP) and multiscale LPQ (MLPQ) in the form of multiple kernels fusion based on the computationally efficient spectral regression kernel discriminant analysis (KDA) [9]. Recently, Deng et al. [10] proposed a powerful system by emulating biological strategies of human visual system, and the proposed system integrates three parts: dual retinal texture and color features for face representation, an incremental robust discriminant model for high-level face coding, and a hierarchical cue-fusion method for similarity qualification. All the above fusion strategies achieve improved performance compared with single descriptor based approaches.

Intuitively, in fusing different features, the complementarity among them has a key role. In recent years, the fusion of magnitude and phase features in a frequency domain has attracted considerable attention [5, 6]. In this paper, we attempt to fuse magnitude feature extracted by Gabor wavelets transform and phase feature which is locally quantized after Fourier transform. Specifically, we first extract the Gabor magnitude and Fourier phase features from a normalized face image by using Gabor wavelets transform and spatial pyramid based local phase quantization (SPLPQ), respectively. Then, to reduce the high dimensionality of both features and increase the discriminative capability, block-wise fisher discriminant analysis (BFDA) is applied to extract the discriminative low-dimensional features. The BFDA method mainly follows the previous work in [4] and [23], which divides the entire feature set into numerous feature segments and applies fisher discriminant analysis (FDA) on each of them. Finally, score-level fusion is performed to calculate the final similarity. Inspired by the biological cognitive mechanism that the processing performed by the human visual system to judge identity is better characterized as “head recognition” rather than “face recognition” [24–26], we conduct the final fusion of two features in a multiple face models framework. Specifically, three face models in a “zooming” order, i.e., internal, transitional, and external face models, which have the same size and different eye positions are utilized in this paper.

To demonstrate the strength of the proposed approach, we evaluate it on three different large-scale face databases, i.e., Face Recognition Grand Challenge (FRGC) version 2.0 [27] following its standard Exp.4 evaluation protocol, Labeled Faces in the Wild (LFW) [28], and purified Celebrity Faces on the Web (CFW-p) [29] which contains more than 150,000 images of 1,520 subjects. These three databases address two different classification scenarios, i.e., face verification and face identification, and experimental results on them consistently exhibit the complementarity of the two features and the performance boost gained by the multiple face models.

The main contributions of this paper include the following: 1) We demonstrate that the fusion of Gabor magnitude and locally quantized Fourier phase provides a complementary description for robust face recognition; 2) We investigate the significant role of multiple face models for system performance boosting, and offer a guiding suggestion on how to select appropriate face models; 3) The proposed approach achieves the state-of-the-art result on FRGC ver2.0, i.e., approximately 96% verification rate on Exp.4, in other words, the error rate is reduced by approximately 30% compared with the best known results; 4) We present a large-scale benchmark, i.e., CFW-p, on the basis of CFW along with ground truth identity labels and facial landmarks annotated manually, and then further design a challenging face identification protocol for future research.

The rest of this paper is organized as follows. Section 2 describes the extraction methods of magnitude and phase features, i.e., Gabor magnitude and Spatial Pyramid based Local Phase Quantization (SPLPQ). In Section 3, multiple face models and the construction of final fusion system are presented. In Section 4, experiments and analyses are conducted, followed by conclusion and discussion in the last section.

2 Extraction of magnitude and phase features for face representation

It has been verified in previous works [5, 6] that magnitude and phase features in a frequency domain have different yet complementary roles in face perception. More specifically, in the frequency domain, magnitude features always capture the facial structure, whereas phase features can provide a detailed description of the facial texture. Therefore, it is desirable to combine these intelligently. In this paper, we investigate the fusion of two powerful features, i.e., classical Gabor wavelets transform for magnitude feature extraction and SPLPQ for phase feature extraction. In this section, we will first describe the two feature extraction methods in Sections 1 and 2 respectively, and then we demonstrate how to use the Blockwise Fisher Discriminant Analysis (BFDA) to further reduce the high dimensionality of both features in Section 3.

2.1 Magnitude feature extraction by Gabor wavelets transform

Since the pioneering work of Lades et al. [30], local face descriptor based on Gabor wavelets transform has been widely used in face recognition and in recent years proven to be one of the most successful face representations (e.g., [2–5], and [31]). This is mainly due to the fact that Gabor wavelets can well approximate the receptive fields of simple cells in the primary visual cortex of human vision system. The Gabor wavelets are always defined in the form of Gabor kernels [2], given by

$$\varphi_{u,v}(z) = \frac{\left\|k_{u,v}\right\|^2}{\sigma^2} e^{(-\left\|k_{u,v}\right\|^2 \|z\|^2 / 2\sigma^2)} [e^{ik_{u,v}z} - e^{-\sigma^2/2}]. \tag{1}$$

Here, $\varphi_{u,v}(\cdot)$ is the Gabor kernel with orientation u and scale v, z denotes the pixel coordinate, i.e., $z = (x, y)$, $\|\cdot\|$ denotes the norm operator, and the wave vector $k_{u,v}$ is defined as follows:

$$k_{u,v} = k_v e^{i\phi_u}, \tag{2}$$

where $k_v = k_{\max}/f^v$ and $\phi_u = \pi u/8$, $k_{\max}$ is the maximum frequency, and f is the spacing between kernels in the frequency domain.

As can be observed from the definition, a 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, whereas the frequency information distant from the center of the Gaussian has a negligible effect. Gabor wavelets can assume a variety of forms with different scales and orientations. Figure 1 displays the real and imaginary parts of 40 Gabor wavelets with five scales and eight orientations.

Clearly, 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 a considerably detailed structure of important facial areas such as eyes, nose, and mouth, which are useful for face representation.

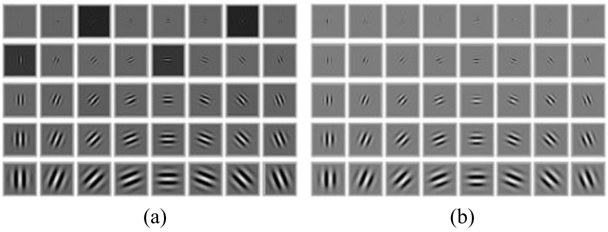


Fig. 1 Visualization of Gabor wavelets. (a) Real and (b) imaginary parts of the Gabor kernels at five scales (i.e., $v \in \{1, 2, \ldots, 5\}$) and eight orientations (i.e., $u \in \{0, 2, \ldots, 7\}$) with the following parameters: $\sigma = 2\pi$, $k_{\max} = \pi$, and $f = \sqrt{2}$. Evidently, the Gabor wavelets exhibit desirable characteristics of spatial frequency, spatial locality, and orientation selectivity

Given the above defined Gabor wavelets, Gabor features are then extracted by convolving them with sub-windows sliding the face image pixel by pixel, given by

$$G_{u,v}(z) = I(z) * \varphi_{u,v}(z). \tag{3}$$

Here, $I(z)$ denotes the input face image, and $*$ denotes the convolution operator. For each Gabor kernel, at every pixel of the face image, a complex number can be generated which contains two Gabor parts (i.e., real part $Re_{u,v}(z)$ and imaginary part $Im_{u,v}(z)$). Based on these two parts, magnitude value $M_{u,v}(z)$ can be computed by

$$M_{u,v}(z) = \sqrt{Re^2_{u,v}(z) + Im^2_{u,v}(z)}. \tag{4}$$

Figure 2 shows the Gabor magnitude feature of a sample face image extracted by convolving the Gabor kernels illustrated in Fig. 1 with size 31 × 31 sub-window sliding the face image pixel by pixel. Now we finish the magnitude feature extraction via Gabor wavelets transform, and next we are going to discuss how to model the face image by utilizing the phase information in the frequency domain.

2.2 Phase feature extraction by SPLPQ

Local phase quantization (LPQ) is a novel descriptor proposed by Ojansivu et al. [13] which is originally devised for texture classification with its excellent properties, i.e., robustness to image blurring and insensitiveness to uniform illumination changes (the design of Gabor wavelets transform cannot ensure such properties). Considering this, LPQ is a complimentary feature with the Gabor magnitude feature discussed in Section 1. In addition, to integrate the scale-invariant property, we further extend the basic LPQ to Spatial Pyramid based LPQ (SPLPQ).

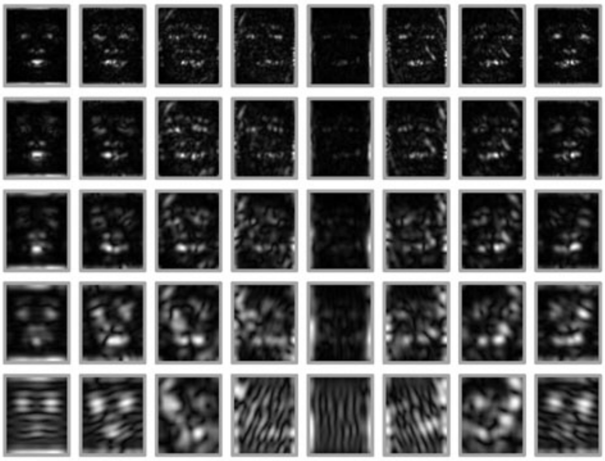


Fig. 2 Gabor magnitude feature of a sample face image extracted by convolving the Gabor kernels illustrated in Fig. 1 with size 31 × 31 sub-window sliding the face image pixel by pixel

2.2.1 Local phase quantization

LPQ utilizes the phase information locally extracted using the short term Fourier transform (STFT) computed over a square $M \times M$ neighborhood N_x at each pixel position x of the face image $f(x)$ [13] defined by

$$F(u, x) = \sum_{y \in N_x} f(x - y) e^{-j2\pi u^{\mathrm{T}} y} = w_u^{\mathrm{T}} f_x, \tag{5}$$

where f_x is a vector containing all the M^2 gray-scale values from N_x, w_u is the basis vector of the STFT at frequency u. As suggested in [13], only four complex coefficients are selected in LPQ, corresponding to 2-D frequencies $u_1 = [a, 0]^{\mathrm{T}}$, $u_2 = [0, a]^{\mathrm{T}}$, $u_3 = [a, a]^{\mathrm{T}}$, and $u_4 = [a, -a]^{\mathrm{T}}$. Let

$$F_x^c = [F(u_1, x), F(u_2, x), F(u_3, x), F(u_4, x)], \tag{6}$$

and

$$F_x = [Re\{F_x^c\}, Im\{F_x^c\}]^{\mathrm{T}}, \tag{7}$$

where $Re\{\cdot\}$ and $Im\{\cdot\}$ are real and imaginary parts of a complex number, respectively. So, the corresponding $8 \times M^2$ transformation matrix is

$$W = [Re\{w_{u_1}, w_{u_2}, w_{u_3}, w_{u_4}\}, Im\{w_{u_1}, w_{u_2}, w_{u_3}, w_{u_4}\}]^{\mathrm{T}}, \tag{8}$$

so that

$$F_x = W f_x. \tag{9}$$

Assuming that $f(x)$ is a result of a first-order Markov process, where the correlation coefficient between adjacent pixel gray-scale values and the variance of each sample are ρ and

σ^2, respectively (assuming $\sigma^2 = 1$ without a loss of generality). Then, the covariance between positions x_i and x_j can be expressed by

$$\sigma_{ij} = \rho^{\|x_i - x_j\|_{L_2}}. \tag{10}$$

Hence, the covariance matrix of all the M samples in N_x can be expressed by

$$C = \begin{bmatrix} 1 & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{21} & 1 & \cdots & \sigma_{2M} \\ \vdots & \vdots & & \vdots \\ \sigma_{M1} & \sigma_{M2} & \cdots & 1 \end{bmatrix}. \tag{11}$$

As a result, the covariance matrix of F_x can be obtained from

$$D = WCW^{\mathrm{T}}. \tag{12}$$

We can easily notice that the coefficients are correlating, as D is not a diagonal matrix when $\rho > 0$. The coefficients should be de-correlated using a whitening transform before quantization, because it can be demonstrated that if the samples to be quantized are statistically independent, information can be maximally preserved in scalar quantization. Whitening transform can be expressed by

$$G_x = V^{\mathrm{T}} F_x, \tag{13}$$

where V is an orthonormal matrix derived from the singular value decomposition of matrix D. G_x is computed for each image position, and the resulting vectors are quantized by a simple scalar quantizer

$$q_j = \begin{cases} 1, & \text{if } g_j \geqslant 0; \\ 0, & \text{otherwise}, \end{cases} \tag{14}$$

where g_j is the jth component of G_x. Eventually, the quantized coefficients can be represented as integer values from zero to 255 using binary coding as in LBP [14]

$$b = \sum_{j=1}^{8} q_j 2^{j-1}. \tag{15}$$

Figure 3 shows the extracted LPQ features of several sample face images.

2.2.2 Spatial pyramid based local phase quantization

In this section, we extend basic LPQ in a spatial pyramid matching (SPM) [32] framework, and the final descriptor is called spatial pyramid based local phase quantization (SPLPQ) which has superior performance compared with the basic LPQ.

SPM was first explored by Grauman and Darrell [33] to determine an approximate correspondence between two feature sets and then extended by Lazebnik et al. [32] to apply to the problem of natural scene category recognition. To some extent, SPM shares a similar idea with multi-resolution histograms [34], which involves sub-sampling an image repeatedly and computing a global histogram of pixel values at each new layer. In other words, multi-resolution histograms work in a way that the image has varying resolutions at which the features are computed, but the histogram resolution stays fixed. Conversely, SPM assumes the opposite approach by repeatedly subdividing the image and computing histograms of local features at increasingly fine resolutions (e.g., Fig. 4 shows a toy example of constructing a three-layer spatial pyramid). This results in a higher-dimensional representation that preserves more information [32].

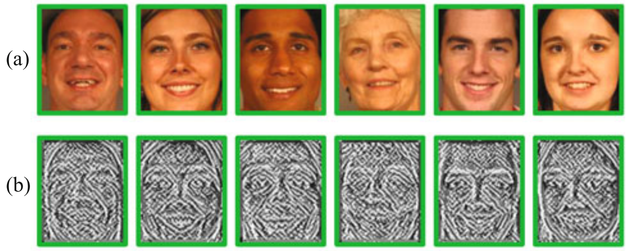


Fig. 3 LPQ features of some sample face images (The LPQ features are extracted with the following parameters: $M = 7$, $a = 1/7$, $\sigma^2 = 1$, and $\rho = 0.9$ which are suggested in [13]) (a) The input face images; (b) the corresponding LPQ features

2.3 Block-wise fisher discriminant analysis and similarity computation

Up to now, we have introduced the extraction methods of the two complimentary features, i.e., Gabor magnitude feature and SPLPQ phase feature. However, it is not reasonable to concatenate them to a single long vector, because by doing that the locality information will not be utilized completely [4]. To overcome this potential weakness, two features are respectively divided into a number of feature vectors correspond to spatially blocks, i.e., block-wise representation, by doing this more locality information can be preserved. Here each block corresponds to a local area of the face image and is of relatively lower dimensionality which means less computing cost for the subsequent processing. In addition, compared with holistic representation, this block-wise representation is more robust to illumination variation. The reason is that the illumination variation within the whole face image is much greater than that within each block. To further reduce the dimensionality of the block-wise features, Block-wise fisher discriminant analysis (BFDA) [4, 23] is applied. Then, the similarities from different local FDA will be fused in the score level.

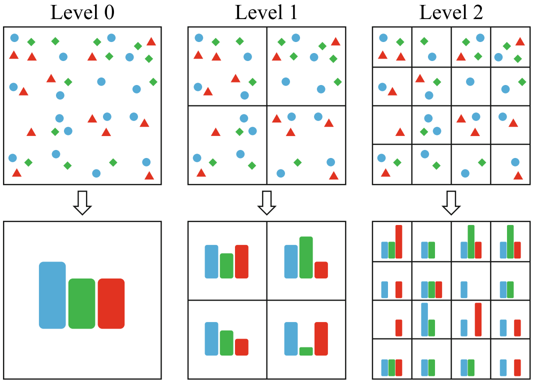


Fig. 4 A toy example of constructing a three-layer spatial pyramid (Assume that the image has three feature types, indicated by circles, triangles and diamonds. First, the image is divided into blocks at three different gridding resolutions. Next, for each layer and each block, we count the features that fall in it in the form of histogram. Finally, all the histograms of the three layers are fused in score level)

BFDA is an extended application of fisher discriminant analysis (FDA) [35] on image blocks. FDA is a linear subspace dimensionality reduction method which tries to shape the scatter in order to make it more reliable for high accuracy classification. This method learns the projection matrix in such a way that the ratio of the between-class scatter and within-class scatter is maximized. Let the between-class scatter matrix be defined as

$$S_B = \sum_{i=1}^{C} N_i(\mu_i - \mu)(\mu_i - \mu)^{\mathrm{T}}, \tag{16}$$

where C is the number of classes, N_i is the number of samples in the ith class, μ_i is the mean representation of the ith class, and μ is the mean representation of all the samples. The within-class scatter matrix can be defined as

$$S_W = \sum_{i=1}^{C} \sum_{f_k \in F_i} (f_k - \mu_i)(f_k - \mu_i)^{\mathrm{T}}, \tag{17}$$

where F_i is the sample set of the ith class, and f_k is the kth sample in a specific class. If S_W is nonsingular, the optimal projection matrix W_{opt} is chosen as the matrix with orthonormal columns which maximizes the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples, i.e.,

$$W_{opt} = \arg\max_{W} \frac{|W^{\mathrm{T}} S_B W|}{|W^{\mathrm{T}} S_W W|} = [w_1\ w_2\ \cdots\ w_m], \tag{18}$$

where $\{w_i | i = 1, 2, \ldots, m\}$ is the set of generalized eigenvectors of S_B and S_W corresponding to the m largest generalized eigenvalues $\{\lambda_i | i = 1, 2, \ldots, m\}$, i.e.,

$$S_B w_i = \lambda_i S_W w_i,\ i = 1, 2, \ldots, m. \tag{19}$$

Note that there are at most $C - 1$ nonzero generalized eigenvalues, and so an upper bound on m is $C - 1$. Now we can reduce the dimensionality of input feature vector f by projecting with W_{opt}^{T} as follows:

$$\widehat{f} = W_{opt}^{\mathrm{T}} \cdot f, \tag{20}$$

where $\widehat{f}$ is the feature after FDA which has lower dimensionality.

However, in practical face recognition problem, the within-class scatter matrix S_W is always singular. In order to overcome the complication of such S_W, we first use principal component analysis (PCA) to reduce the dimensionality of the input feature to a relatively small value and then apply the FDA.

Next we will show how to use BFDA respectively to the above two features, see Figs. 5 and 6 for more intuitive understanding.

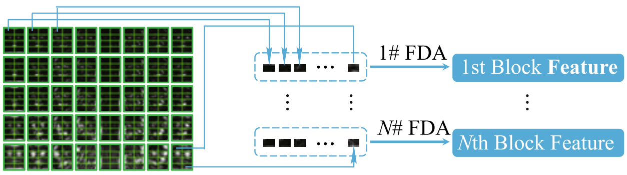


Fig. 5 Illustration of applying BFDA for Gabor magnitude feature

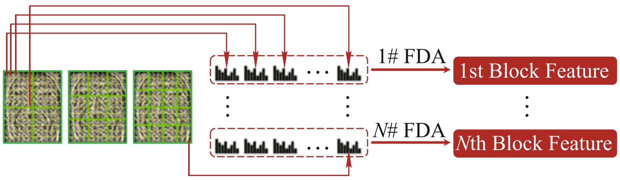


Fig. 6 Illustration of applying BFDA for SPLPQ phase feature

For Gabor magnitude feature, we first convolve the input face image with Gabor kernels of multiple scales and multiple orientations defined in Eq. (1). After the convolution and magnitude value computing, we can get a series of resulting images called Gabor magnitude maps (GMMs), and then we divide each GMM into N non-overlapping blocks. Let us denote

$$G_i = [G^i_{u_0,v_0}, \ldots, G^i_{u_0,v_{s-1}}, G^i_{u_1,v_0}, \ldots, G^i_{u_{o-1},v_{s-1}}], \tag{21}$$

as the concatenated Gabor magnitude feature of the ith block of a specific face image, where $G^k_{u,v}$ is a vector containing all the magnitude values from the kth block of GMM with scale

v and orientation u. Then the similarity between face A and face B with Gabor magnitude feature can be defined as

$$\begin{aligned} S_{AB\_Gabor} &= \sum_{i=1}^{N} w_i \cdot sim(\widehat{G}_{A\_i}, \widehat{G}_{B\_i}) \\ &= \sum_{i=1}^{N} w_i \cdot \frac{\widehat{G}_{A\_i} \cdot \widehat{G}_{B\_i}}{||\widehat{G}_{A\_i}|| \cdot ||\widehat{G}_{B\_i}||}, \end{aligned} \tag{22}$$

where $\widehat{G}_{A\_i}$ and $\widehat{G}_{B\_i}$ represent the features after dimensionality reduction by FDA of the ith block of face A and face B, w_i is the weight of the ith block (in practice, we set equal weight for each block).

For the SPLPQ phase feature, we divide each spatial pyramid layer into different appointed numbers of non-overlapping blocks, and each block is further divided into corresponding number of fixed size sub-blocks. So we can denote

$$H_{li} = [H_{li}^1, H_{li}^2, \ldots, H_{li}^K], \tag{23}$$

as the original concatenated histogram feature of the ith block in the lth layer, where K is the number of sub-blocks on which tuple histograms are computed. In other words, each H_{li} is generated by concatenating K histograms computed on sub-blocks belong to the ith block in the lth layer. Each H_{li}^k has the dimensionality of 256, so the dimensionality of H_{li} is $256 \times K$. Like Gabor magnitude feature, we further denote $\widehat{H}_{li}$ as the feature after FDA, which has lower dimensionality. Now we can define the similarity of face A and face B with SPLPQ phase feature as:

$$\begin{aligned} S_{AB\_SPLPQ} &= \sum_{l=1}^{L} \sum_{i=1}^{N_l} w_{li} \cdot sim(\widehat{H}_{A\_li}, \widehat{H}_{B\_li}) \\ &= \sum_{l=1}^{L} \sum_{i=1}^{N_l} w_{li} \cdot \frac{\widehat{H}_{A\_li} \cdot \widehat{H}_{B\_li}}{||\widehat{H}_{A\_li}|| \cdot ||\widehat{H}_{B\_li}||}, \end{aligned} \tag{24}$$

where L is the number of spatial pyramid layers, N_l denotes the number of blocks in the lth layer, and w_{li} is the weight of the ith block in the lth layer (in practice, we set equal weight for each layer and equal sub-weight for each block). After having the similarities computed based on Gabor magnitude feature and SPLPQ phase feature, we can fuse them in score level as Eq. (25).

$$S_{AB} = w \cdot S_{AB\_Gabor} + (1 - w) \cdot S_{AB\_SPLPQ}, \tag{25}$$

where S_{AB} is the similarity of face A and face B, w is the weight to balance the roles of two complimentary features (without a loss of generality, here we again take the equal weights).

3 Feature fusion in multiple face models framework

In this section we will introduce another helpful technique, i.e., multiple face models, which is designed inspired by a biological cognitive mechanism to further boost the system performance, after that we will give the final fusion approach in the multiple face models framework.

3.1 Multiple face models

Sinha et al. [24–26] pointed out that the human vision system's processing to judge one's identity is better characterized as "head recognition" rather than "face recognition". Figure 7 shows three sample subjects from FRGC ver2.0 and their different types of face models in a "zooming" order from left to right. Apparently, it is difficult for some people to determinate whether two faces are of the same subject based only on the internal images, i.e., images in the second column of Fig. 7. However, one can recognize a face more easily if given the external image, i.e., images in the rightmost column of Fig. 7. Behind this interesting biological cognitive phenomenon, the rationality is that human tend to rely on the contextual information to recognize faces, such as hair style, head contour, jaw and even background [36]. This trend will be enhanced especially when intrinsic information is degraded [10]. Thus, it is smart to add the multiple face models which contain not only intrinsic but also holistic contextual information to the current system. With the above analysis, in the proposed approach, three normalized face models of the same size but different eyes' positions are taken into account to best imitate human vision system, and we call them internal face, transitional face and external face shown in Fig. 8. As can be seen, the internal face model contains only the internal facial organs, such as eyes, mouth, nose and eyebrows which are affected only by the factors related to identity. On the contrary, the external face model is portrait-like and contains some external facial elements such as jaw, head contour and hair. The transitional face image can be regarded as the transition state from the internal face model to the external face model. To sum up, the internal face has the highest facial resolution but the minimum inner facial region, whereas the external face contains larger facial region but relatively low facial resolution.

3.2 Construction of final fusion system

Based on the two features, i.e., Gabor magnitude feature and

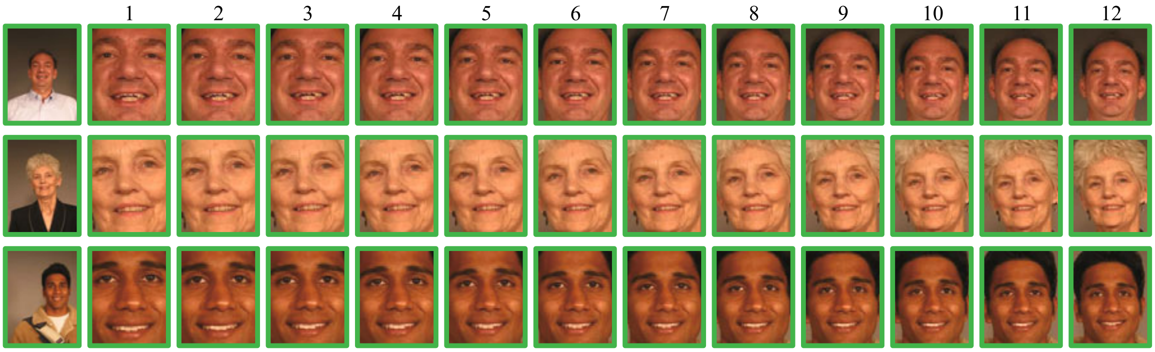


Fig. 7 Three sample subjects from FRGC ver2.0 and their different types of face models, where the leftmost column shows the original images, and the second to the rightmost columns correspond to 12 different face models with the same size but different eyes' positions (The size of all the face models is fixed to 120×96, and the eyes' coordinates are: {(20,36), (75,36)}, {(21,37), (74,37)}, {(22,38), (73,38)}, {(23,40), (72,40)}, {(24,42), (71,42)}, {(25,44), (70,44)}, {(26,46), (69,46)}, {(27,48), (68,48)}, {(28,50), (67,50)}, {(29,52), (66,52)}, {(30,54), (65,54)}, and {(31,55), (64,55)} from model 1 to model 12)

SPLPQ phase feature, and three face models, i.e., internal face, transitional face, and external face, in this subsection we define the final similarity which is the integration of the two features and the three face models. More specially, with a coming image, we first generate the above three face models, and then we extract the two features on each of them. Therefore we will have six similarities, then a simple weighted summation strategy is applied to integrate them to the final similarity defined as

$$\begin{aligned} S_{AB} = w_1 \cdot S_{AB\_Gabor\_I} + w_2 \cdot S_{AB\_SPLPQ\_I} + \\ w_3 \cdot S_{AB\_Gabor\_T} + w_4 \cdot S_{AB\_SPLPQ\_T} + \\ w_5 \cdot S_{AB\_Gabor\_E} + w_6 \cdot S_{AB\_SPLPQ\_E}. \end{aligned} \tag{26}$$

We take one of the elements as an example to explain, e.g., $S_{AB\_Gabor\_I}$ denotes the similarity of face A and face B computed with Gabor magnitude feature on the internal face model, and $\{w_i | i = 1, 2, \ldots, 6\}$ denotes the weight for each similarity (without a loss of generality, we take the equal weights here). Figure 8 shows the whole proposed approach as a summary.

4 Experiment and analysis

In this section, we evaluate the proposed approach on three quite different large-scale face databases, i.e., face recognition grand challenge (FRGC) version 2.0 [27], labeled faces in the wild (LFW) [28], and purified celebrity faces on the Web (CFW-p) [29]. Moreover, these three databases cover two different classification scenarios, i.e., face verification on FRGC ver2.0 and LFW, and face identification on CFW-p.

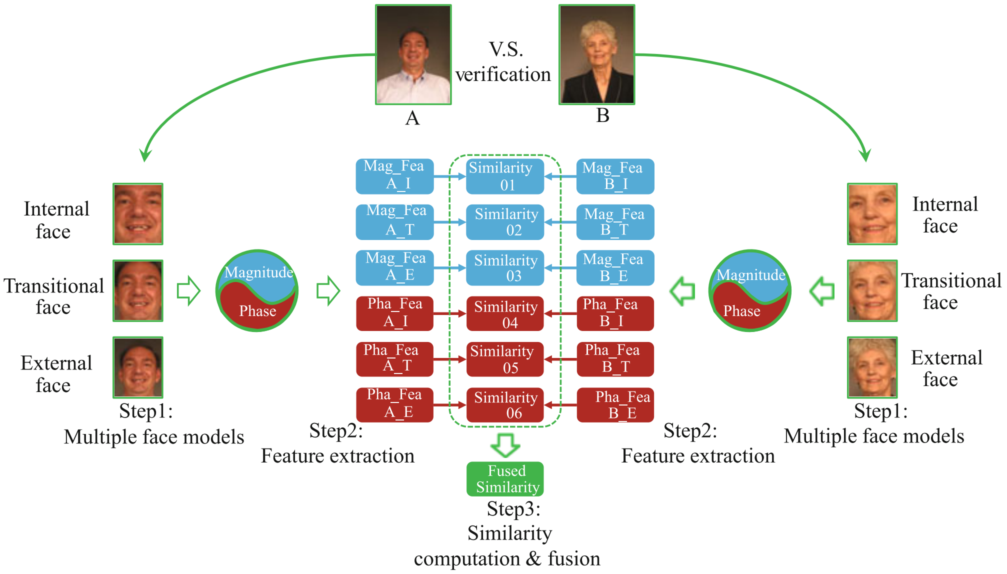


Fig. 8 The proposed approach which fuses magnitude and phase features with multiple face models, where the Gabor magnitude feature part is represented by light blue, and the SPLPQ phase feature part is represented by dark red

4.1 Evaluation of proposed approach on FRGC v2.0 Exp.4

4.1.1 Database and experimental setup

Face recognition grand challenge (FRGC) version 2.0 [27] is a large-scale face recognition evaluation benchmark sponsored by the United States government and collected at the University of Notre Dame. This face database is designed to achieve the goal that reducing the error rate of face recognition systems by an order of magnitude. With the above goal, it presents six challenging experiments along with data corpus of 50,000 recordings divided into training and validation partitions to researchers. The database consists of high resolution still images taken under controlled and uncontrolled conditions. The controlled images taken in a studio setting (two or three studio lights) are full frontal facial images with two facial expressions, i.e., neutral and smiling. The uncontrolled images are taken in varying lighting conditions, e.g., atria, hallways, or outdoors. Each set of uncontrolled images also contains two expressions, i.e., neutral and smiling. Figure 9 shows some examples of FRGC ver2.0.

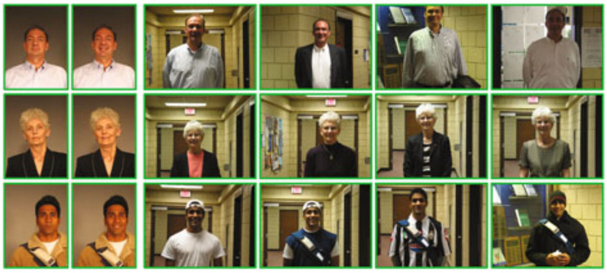


Fig. 9 Example images of three subjects from FRGC ver2.0 Exp.4 (The left two columns show the controlled target images with neutral and smiling expressions, and the other four columns show uncontrolled query images of corresponding subjects with variations caused by expression, out of focus blur, unsymmetrical illumination and large time lapse)

As recognizing faces under uncontrolled conditions, which is considered the security requirement for real-world biometric recognition, has numerous applications and is one of the most challenging problems in the field of face recognition, we choose Exp.4 to evaluate the proposed approach. In Exp.4, training set consists of 12,776 images of 222 subjects, with 6,388 controlled still images and 6,388 uncontrolled still images. The target set consists of 16,028 controlled still images, and the query set consists of 8,014 uncontrolled still images. Participating algorithms therefore produce a $16,028 \times 8,014$ matrix of similarity scores for all possible pairs, i.e., about 128 million pairs of faces.

When testing, the verification performance is reported in the form of receiver operating characteristic (ROC) which indicates the performance level for all possible combinations of correct verification rate (VR) and false acceptance rate (FAR). In particular, the official recommends three ROC curves: ROC I, ROC II, and ROC III, correspond to image pairs collected within semester, within year, and between semesters, respectively. In this paper, the tested approaches are typically compared in terms of the VR at a fixed FAR of 0.1%, which is considered the security requirement for real-world biometric applications.

4.1.2 Technical details

In this work, we used three face models as mentioned in Section 1, i.e., internal, transitional and external face models with eyes located at {(20,47), (75,47)}, {(26,47), (69,47)}, and {(29,51), (66,51)}, respectively. The size of above three face models are fixed to 120 × 96. Moreover, before extracting features, PP [37] which has been proven a robust illumination normalization method is used to eliminate illumination effects.

In the Gabor magnitude feature extraction process, 40 Gabor wavelets with 5 scales (i.e., $v \in \{1, 2, \ldots, 5\}$) and 8 orientations (i.e., $u \in \{0, 1, \ldots, 7\}$) are utilized, where the Gabor kernel's size, the maximum frequency $k_{\max}$, the spacing between kernels in the frequency domain f and the parameter σ are set to 31×31, π, $\sqrt{2}$ and 2π, respectively. 40 Gabor magnitude maps are generated after the convolution with the above Gabor wavelets, after that we use a 4×4 down sampling to preliminarily reduce the huge dimensionality of the original Gabor magnitude feature. As mentioned in Section 3, the proposed approach is based on blocks, so we divide each down sampled GMM into 18 blocks (6 rows × 3 columns), and concatenate the features of the same block into a single vector which followed by FDA to generate lower-dimensional discriminative feature.

In the SPLPQ phase feature extraction process, the correlation coefficient ρ, the sliding window's size and the frequency parameter a are set to 0.9, 7×7 and 1/7, respectively. For the spatial pyramid, we use a three-layer structure which has 3×2, 3×6, and 5×4 blocks (rows × columns) in the 0th, 1st, and 2nd layer, respectively. The sub-block in which we extract histogram has the size of 8×8. Like Gabor magnitude feature, FDA is used to generate lower-dimensional discriminative feature for each block of SPLPQ.

In the dimensionality reduction process, we set the PCA dimensionality to 600 and FDA dimensionality to 221, i.e., one less than the subject amount (222 subjects) of the training set, for all the blocks of both features. Without a loss of generality, in the similarity fusion process, all the weights are set to be equal as mentioned in Sections 2 and 3. To further validate the generality of the proposed approach, we fix the parameters' values for all the two test databases.

4.1.3 Multiple face models selection

Figure 7 shows 12 different face models, from left to right each contains more contextual information and has lower facial resolution. In this part, we elaborate how to select the three face models from the 12 shown in Fig. 7. It is obvious that the neighboring two face models have little appearance difference but large redundancy. However, appearance complementarity is an essential criterion that should be taken into consideration. More specifically, it is better to make the selected face models to have very different scopes of visual facial region and facial resolutions. For this, we group the 12 face models into three sets, i.e., the first set contains the internal face model candidates (from the 1st face model to the 4th face model), the second set contains the transitional face model candidates (from the 5th face model to the 8th face model), and the third set contains the rest face models as candidates of the external face model. More specifically, we list the performance of individual face model in the left part of Table 1. Obviously, combinations within set (i.e., internal, transitional, and external) achieve very limited performance improvement, because components in the combination are too similar to complement with each other. To maximize the complementarity between three face models, we use the strategy of selecting one face model from each of the three sets. Then we evaluate all the 64 (4 × 4 × 4) possible combinations, and get the expected result that the combination of the 1st, 7th and 12th face models achieves almost the best result among all the combinations (for space limitation, we only show the best one among the 64 candidate combinations). This is mainly because in this way the gap between the selected face models is maximized, and this point further ensures the complementarity.

4.1.4 Comparison between LPQ and SPLPQ

This part mainly evaluates the performance of SPLPQ which is an extension of LPQ in the form of spatial pyramid structure. The comparison between LPQ and SPLPQ is shown in Table 2. We decompose the proposed three-layer SPLPQ into three single layers and compare them with SPLPQ. As expected, SPLPQ outperforms single layer LPQ on three different ROC settings. This mainly attributes to the more information preserved by the spatial pyramid structure, e.g., small size block can only characterize local single facial organ (e.g., mouth, nose), whereas larger size block may capture the correspondence between several important facial organs (say, correspondence between mouth and nose).

Table 1 Evaluation of 12 different face models and their combinations on FRGC ver2.0 Exp.4 with two features, i.e., Gabor and SPLPQ

Sing.	Gabor	SPLPQ	Comb.	Gabor	SPLPQ
1	0.9214	0.8553	1,2,3	0.9201	0.8603
2	0.9193	0.8611	1,2,4	0.9255	0.8592
3	0.9221	0.8594	1,3,4	0.9279	0.8661
4	0.9187	0.8686	2,3,4	0.9242	0.8714
5	0.9267	0.8702	5,6,7	0.9297	0.8905
6	0.9232	0.8755	5,6,8	0.9317	0.8933
7	0.9248	0.8818	5,7,8	0.9336	0.8896
8	0.9233	0.8776	6,7,8	0.9305	0.8885
9	0.9188	0.8671	9,10,11	0.9183	0.8702
10	0.9093	0.8599	9,10,12	0.9227	0.8697
11	0.9069	0.8483	9,11,12	0.9209	0.8651
12	0.8989	0.8400	10,11,12	0.9175	0.8689
- -	- - - - -	- - - - -	**1,7,12**	**0.9554**	**0.9350**

Note: For space limitation, the performance reported here are under the ROC I protocol corresponding to image pairs collected within semester

Table 2 Comparison between LPQ and SPLPQ on FRGC ver2.0 Exp.4. ROC I, ROC II and ROC III, corresponding to image pairs collected within semester, within year, and between semesters, respectively

Approach	ROC I	ROC II	ROC III
Layer0:LPQ(3 × 2 blocks)	0.8473	0.8395	0.8296
Layer1:LPQ(3 × 6 blocks)	0.8392	0.8248	0.8079
Layer2:LPQ(5 × 4 blocks)	0.8318	0.8174	0.8008
Three-layer SPLPQ	**0.8818**	**0.8704**	**0.8572**

4.1.5 Evaluation of feature fusion

To validate the merits of the fusion of Gabor magnitude feature and SPLPQ phase feature, we compare the performance between fused approach and non-fused approach which utilizes only one of the two features. We conduct the experiments on the three face models separately, the experimental results can be found in Table 3 and Fig. 10. It is obvious to see that: a) Gabor magnitude feature performs better than SPLPQ in FRGC ver2.0 Exp.4, that's mainly because the images in FRGC ver2.0 are of high resolution, little pose variance, and relatively good alignment. Moreover, these form the exact favorite setting of Gabor magnitude feature (on the contrary, SPLPQ performs better in more challenging settings, please see Section 3); b) The complementarity between two features indeed ensures the system performance increasing, e.g., in ROC I with the transitional face model, Gabor magnitude feature and SPLPQ phase feature have the VRs of 92.48% and 88.18% at 0.1% FAR, respectively, whereas the fused approach achieves 94.14%. For more intuitive interpretation of the complementarity between the two features, we select a couple of testing pairs and show them in Fig. 11. Behind the complementarity, we can easily notice that, compared with Gabor magnitude feature, SPLPQ performs much remarkable on blurred images, especially caused by centrally symmetric blur [13] (e.g., blur caused by motion, out of focus, and atmospheric turbulence).

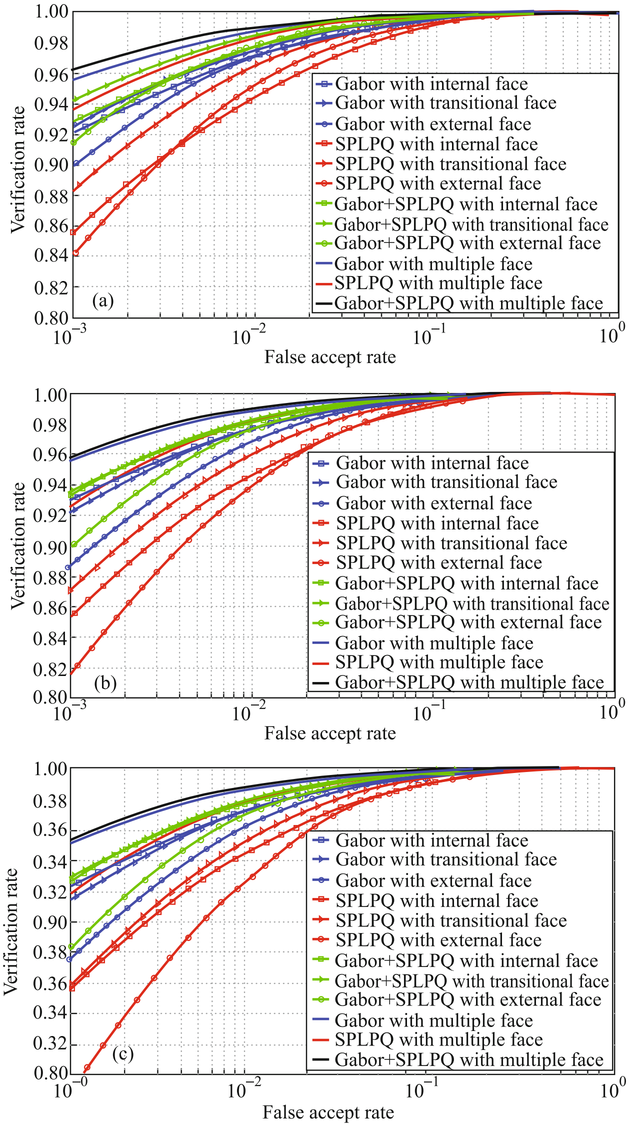


Fig. 10 ROC curves of the proposed approach on FRGC ver2.0 Exp.4. (a) ROC I, (b) ROC II, and (c) ROC III, correspond to image pairs collected within semester, within year, and between semesters, respectively

4.1.6 Evaluation of multiple face models

To validate the merits of multiple face models, we compare the performance between the approaches with single face model and the approach with multiple face models. These experimental results can be found in Table 3 and Fig. 10. As can be seen from the table, the performance enhancement brought by the multiple face models is significant, e.g., in ROC I with Gabor magnitude feature, approaches based on single face model get the VRs of 92.14%, 92.48%, and 89.89% at 0.1% FAR, respectively, whereas the multiple face models based approach achieves 95.54%. Similar phenomenon becomes more significant with the SPLPQ phase feature, e.g., 85.53%, 88.18%, and 84.00% for three single face models in ROC I, and 93.50% for the multiple face models version. Another interesting observation is that approaches based on external face model which contains relative more holistic contextual information perform a little worse compared with the approaches based on internal and transitional face models. Maybe this is caused by the unified background in FRGC ver2.0 which contains little identity relevant information.

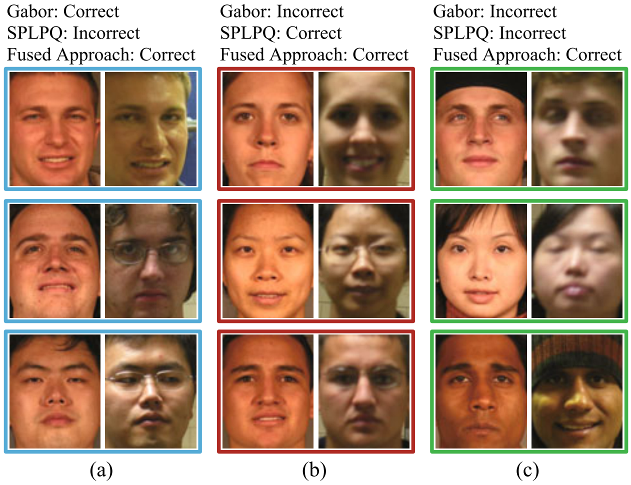


Fig. 11 Illustration of the complementarity between Gabor magnitude feature and SPLPQ phase feature on FRGC ver2.0 Exp.4. (a) and (b) show sample pairs verified correctly by one feature, but verified incorrectly by the other feature. (c) shows sample pairs that neither feature works on. Please note that the two faces in each pair shown here come from the same subject, and all the pairs shown in (a), (b) and (c) are verified correctly by the fused approach

4.1.7 Comparison with other state-of-the-art approaches

After the evaluation of feature fusion and multiple face models, in this part we compare the final fusion approach with the other state-of-the-art approaches. Table 3 summarizes the performance of several representative state-of-the-art approaches that have been proposed since FRGC 2005, where most of the results are cited from [10]. We can roughly group these approaches into two parts, one contains the single feature based approaches, the other one contains the multiple features fusion approaches. The proposed approach achieves the impressive VRs of 96.20%, 95.76%, and 95.32% on ROC I, ROC II, and ROC III, respectively, which outperform the

Table 3 Comparative VRs at 0.1% FAR of state-of-the-art face recognition approaches on FRGC ver2.0 Exp.4. ROC I, ROC II, and ROC III, corresponding to image pairs collected within semester, within year, and between semesters, respectively

Approach	Feature	ROC I	ROC II	ROC III
BEE baseline [27]	Pixel	0.1336	0.1267	0.1186
KCFA [44]	Pixel	N/A	N/A	0.57
R-WWC [45]	Pixel	≈0.35	≈0.35	≈0.35
Extended GCID [38]	Pixel	0.7890	0.7866	0.7826
YQCr [40]	YQCr	0.6447	0.6489	0.6521
HVDA [39]	YQCr	0.7865	0.7850	0.7824
KDCV [7]	LBP	N/A	N/A	0.735
MFM-HFF [46]	Fourier	0.7570	0.7506	0.7433
DIF [47]	Gabor	≈0.72	≈0.74	≈0.76
LEC [4]	Gabor	N/A	N/A	0.83
LGBP+LGXP [5]	LGBP+LGXP	0.836	0.843	0.849
HEC [4]	Gabor+Fourier	N/A	N/A	0.89
KDCV [7]	Gabor+LBP	N/A	N/A	0.836
PS_MLPQ+PS_MLBP+KDA [8]	MLPQ+MLBP	0.8292	0.8434	0.8572
Hybrid RCrQ [22]	Gabor+LBP+DCT	N/A	N/A	0.924
RTF+RCF [10]	RTF+RCF	0.9391	0.9355	0.9312
MultOSS [41]	LBP+SIFT+TPLBP+FPLBP	0.8334	0.8252	0.7769
MLHR [42]	Gabor+SPLPQ	0.9312	0.9174	0.9148
RSSM [43]	Gabor+SPLPQ	0.9456	0.9337	0.9331
VGGFace [48]	DCNN	0.9667	0.9616	0.9599
VGGFace (FRGC Finetune) [48]	DCNN	**0.9791**	**0.9778**	**0.9730**
Gabor (Internal face)	Gabor	0.9214	0.9221	0.9230
Gabor (Transitional face)	Gabor	0.9248	0.9197	0.9135
Gabor (External face)	Gabor	0.8989	0.8872	0.8744
SPLPQ (Internal face)	SPLPQ	0.8553	0.8555	0.8549
SPLPQ (Transitional face)	SPLPQ	0.8818	0.8704	0.8572
SPLPQ (External face)	SPLPQ	0.8400	0.8153	0.7874
Gabor+LPQ (Internal face)	Gabor+SPLPQ	0.9286	0.9286	0.9288
Gabor+LPQ (Transitional face)	Gabor+SPLPQ	0.9414	0.9344	0.9267
Gabor+LPQ (External face)	Gabor+SPLPQ	0.9146	0.8985	0.8819
Gabor (Multiple face models)	Gabor	0.9554	0.9537	0.9518
SPLPQ (Multiple face models)	SPLPQ	0.9350	0.9271	0.9182
Final fusion approach	**Gabor+SPLPQ**	**0.9620**	**0.9576**	**0.9532**

other state-of-the-art approaches. In addition, we can reach the following observations from the comparison: a) No single feature based approach can surpass the VR of 85%, that's because single feature can only encode limited information of the given face. As discussed in Section 2, Gabor magnitude feature captures the structure information of the face, whereas SPLPQ phase feature is more robust to blurred image by efficiently encoding the facial texture. Therefore, it is reasonable to combine different yet complementary features for more effective face representation; b) The utilization of multiple face models, e.g., Deng's approach [10] and the proposed approach, can tremendously boost the system performance. Because holistic contextual information is a necessary supplement to the intrinsic facial information; c) The proposed approach outperforms the other magnitude phase feature fusion approach [5] which is based on local Gabor binary patterns (LGBP) and local Gabor XOR patterns (LGXP); d) It is also important to note that, although only gray-scale images are used, the proposed approach outperforms all the color space based approaches, e.g., [10, 22, 38–40].

Furthermore, we detailedly compare the proposed method with three classic multi-feature fusion methods [41–43], where [41] utilized the idea of one-shot similarity along with multiple features (i.e., LBP, SIFT, TPLBP, and FPLBP) for face representation (given two vectors, their one-shot similarity score reflects the likelihood of each vector belonging in the same class as the other vector and not in a class defined by a fixed set of "negative" examples.), [42] showed a new multiple feature learning algorithm MLHR with a statistical approach to exploit the structural information of both the labeled and unlabeled data for multimedia content analysis, and [43] proposed a joint learning framework that integrates semi-supervised learning, multi-feature learning and the Riemannian metric (Reimannian metric is used to measure feature significance). Technically, we conduct comparative experiments based on the source codes and defaulted parameters recommended by the authors ([41,42] have released codes, and we carefully implemented [43] by ourselves). More specifically, we directly treat the proposed method as six different features, i.e., combination of two features and three face models, and feed the six feature sets into the comparative methods. For fair comparison, we upgrade the semi-supervised model from [42, 43] to fully supervised version. Comparison results can be found in Table 3, the proposed method exhibits the competitiveness with its straight-forward framework. Compared with the other two comparative methods, [43] performances best with the assumption that the inconsistency can be evaluated by distances between different graphs (here we implement the graph with subspace corresponding to Grassmann manifold and covariance matrix corresponding to Riemannian manifold). If a graph constructed from one feature type is comparatively farther away from those constructed from other features, this feature is possibly inconsistent with other features from the classification perspective.

At last, we compare the proposed method with state-of-the-art deep learning based method. Here, we just consider a classic off-the-shelf deep model, i.e., VGGFace [48], which is an end-to-end convolutional neural network learning framework designed for face recognition. To conduct comparison, we design two versions of VGGFace, i.e., using the official provided model or fine-tuned model with target database. For the first version, we directly utilize the 4,096-dim fc7 feature as final representation, and for the fine-tuned version, we first fine-tune the network with FGRC ver2.0 Exp.4's 12,776 training images of 222 subjects (40 epoch) based on the provided model, and then extract the same 4,096-dim fc7 feature. VVGFace did exhibited its competitiveness in both versions. We believe such powerful performance should mainly thank to the large-scale data for deep network training, e.g., about 2.6 million images spanning more than 2,600 identities are used for VGGFace training.

4.2 Evaluation of proposed approach on LFW

In the last sub-section, we conduct evaluation of the proposed approach on FRGC ver2.0 Exp.4. To further validate the approach effectiveness, especially the roles of feature fusion and multiple face models, we evaluate the proposed approach on two wild face databases, i.e., LFW and CFW-p. Moreover, these two databases correspond to two classification scenarios, where LFW for face verification and CFW-p for face identification. Please kindly note that for better validate the generality of the proposed approach, we just use the exact same parameters with FRGC ver2.0 on the following two databases.

4.2.1 Database and experimental setup

Labeled faces in the wild (LFW) [28] is an image database for unconstrained face verification and it is quite different from FRGC ver2.0. Specifically, FRGC ver2.0 is designed to study the effect of richer, new data types on the face recognition problem, thus including high resolution data, image sequences, and even 3D scans of each subject. Each of these data types is potentially more informative than the simple, moderate resolution images. In contrast, LFW is designed to help study the face recognition problem using previously existing real world images, that is, images are not taken for the special purpose of face recognition by machine. Figure 12 shows some examples of LFW.

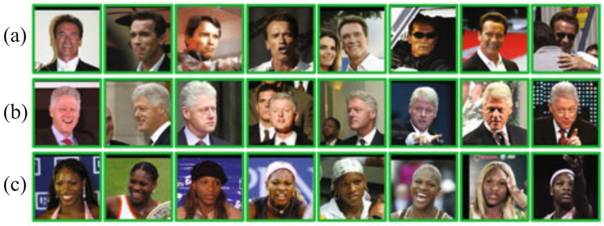


Fig. 12 Example images of three subjects from LFW (Large intra-class variations in lighting, expression, head pose, age, clothing, hairstyles, background, and camera quality can be found here). (a) Arnold Schwarzenegger; (b) Bill Clinton; (c) Serena Williams

Large variations in lighting, head pose, hairstyles, age, expression, background, race, ethnicity, gender, clothing, camera quality, focus, color saturation, and other parameters can be found. LFW contains 13,233 face images of 5,749 subjects collected from the web. Among these, 1,680 subjects have two or more distinct images, the remaining 4,069 subjects have only one image in the database. Two protocols are considered in LFW, the restricted one limits the information available for training to the same/different labels in the training splits; the unrestricted one, on the other hand, allows training methods access to subject identity labels. As the training of FDA needs identity labels, here we conduct the experiment following the unrestricted protocol in form of 10-fold cross validation which splits the data set into 10 subsets with each containing 300 intra-class pairs and 300 inter-class pairs.

4.2.2 Experimental results and analysis

Table 4 summarizes the performance of the proposed approach, several representative approaches [41, 49–51], and multiple feature fusion methods [42, 43] on LFW. Besides, recently deep learning based methods [48, 52–54] start to show their competitiveness on LFW, especially with large-scale labeled outside training data. Thus, we also take such methods into comparison. For fair comparison, all the methods are tested under the unrestricted protocol. As for the usage of labeled outside training data, we mark this information in the second column of Table 4. The performance of listed methods are directly cited from the original literatures or the official LFW homepage. For better validating the method generalizability, we fix all the parameters to stay the same with FRGC ver2.0. Also, we show the proposed approach in similar ways as on FRGC ver2.0, i.e., single feature with single face model, single feature with multiple face models, multiple features with single face model, and multiple features with multiple face models. It can be seen that the proposed approach achieves comparable performance with the other state-of-the-art approaches (not including deep learning based ones). We could find similar observations from Table 4 with the ones on FRGC ver2.0 that the combination of Gabor magnitude feature and SPLPQ phase feature plays a key role for more effective face representation (for interpreting the complementarity of the two features more intuitively, we select a couple of testing pairs and show them in Fig. 13), and multiple face models are reliable engineering techniques for

Table 4 Experimental results of the proposed approach and several typical approaches on LFW, where the performance of listed methods are directly cited from the original literatures or the official LFW homepage

Approach	Labeled outside data	Feature	Result
LBP PLDA, aligned [49]	LFW only	LBP	0.8733 ± 0.0055
Combined multishot, aligned [41]	LFW only	8 features	0.8950 ± 0.0051
MLHR [42]	LFW Only	Gabor+SPLPQ	0.8683 ± 0.0083
RSSM [43]	LFW Only	Gabor+SPLPQ	0.8750 ± 0.0069
Tom-vs-Pete [50]	Columbia University	T-P classifier scores	0.9310 ± 0.0135
HighDimLBP [51]	WDRef	High-Dim LBP	0.9517 ± 0.0113
DeepFace-ensemble [52]	FaceBook Private	DCNN feature	0.9735 ± 0.0025
DeepID [53]	CelebFaces	DCNN feature	0.9745 ± 0.0026
VGGFace [48]	VGG Face	DCNN feature	0.9895 ± - - - - -
FaceNet [54]	Google Private	DCNN feature	0.9963 ± 0.0009
Gabor (Internal face)	LFW Only	Gabor	0.8433 ± 0.0117
Gabor (Transitional face)	LFW Only	Gabor	0.8550 ± 0.0098
Gabor (External face)	LFW Only	Gabor	0.8617 ± 0.0126
SPLPQ (Internal face)	LFW Only	SPLPQ	0.8567 ± 0.0085
SPLPQ (Transitional face)	LFW Only	SPLPQ	0.8600 ± 0.0097
SPLPQ (External face)	LFW Only	SPLPQ	0.8683 ± 0.0072
Gabor+LPQ (Internal face)	LFW Only	Gabor+SPLPQ	0.8667 ± 0.0059
Gabor+LPQ (Transitional face)	LFW Only	Gabor+SPLPQ	0.8717 ± 0.0068
Gabor+LPQ (External face)	LFW Only	Gabor+SPLPQ	0.8783 ± 0.0044
Gabor (Multiple face models)	LFW Only	Gabor	0.8733 ± 0.0053
SPLPQ (Multiple face models)	LFW Only	SPLPQ	0.8817 ± 0.0048
Gabor (Internal face)	Webface	Gabor	0.8933 ± 0.0037
Gabor (Transitional face)	Webface	Gabor	0.8967 ± 0.0062
Gabor (External face)	Webface	Gabor	0.9017 ± 0.0039
SPLPQ (Internal face)	Webface	SPLPQ	0.9000 ± 0.0042
SPLPQ (Transitional face)	Webface	SPLPQ	0.9050 ± 0.0045
SPLPQ (External face)	Webface	SPLPQ	0.9117 ± 0.0071
Gabor+LPQ (Internal face)	Webface	Gabor+SPLPQ	0.9133 ± 0.0029
Gabor+LPQ (Transitional face)	Webface	Gabor+SPLPQ	0.9183 ± 0.0031
Gabor+LPQ (External face)	Webface	Gabor+SPLPQ	0.9233 ± 0.0028
Gabor (Multiple face models)	Webface	Gabor	0.9150 ± 0.0019
SPLPQ (Multiple face models)	Webface	SPLPQ	0.9317 ± 0.0034
Final fusion approach I	**LFW only**	**Gabor+SPLPQ**	**0.8933 ± 0.0047**
Final fusion approach II	**Webface**	**Gabor+SPLPQ**	**0.9450 ± 0.0028**

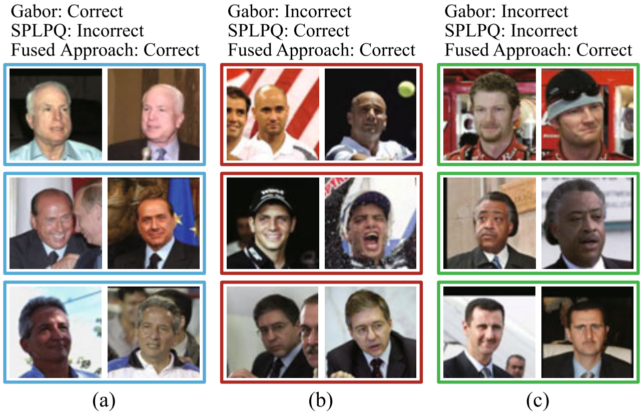


Fig. 13 Illustration of the complementarity between Gabor magnitude feature and SPLPQ phase feature on LFW. (a) and (b) show sample pairs verified correctly by one feature, but verified incorrectly by the other feature. (c) shows sample pairs that neither feature works on. Please note that the two faces in each pair shown here come from the same subject, and all the pairs shown in (a), (b) and (c) are verified correctly by the fused approach

system performance boosting. Besides, we find that in LFW, SPLPQ phase feature performs a little better than Gabor magnitude feature. That is mainly because faces in LFW are not with good alignment compared with FRGC ver2.0, and SPLPQ as a histogram based descriptor is naturally robust to slight spatial misalignment.

As for the deep learning based methods, they did exhibited their competitiveness. Most of such methods require large-scale labeled outside training data, e.g., DeepFace [52] relies on a Facebook private large database with 4.4 million labeled faces from 4,030 people each with 800 to 1,200 faces, FaceNet [54] relies on a Google private database with 100–200 million labeled faces from about 8 million different identities. To compare with such methods, we have to fully take the advantage of outside training data. More specifically, we choose Webface [55] which is regarded as the largest public available database containing 494,414 faces of 10,595 subjects without overlapping with LFW. Technically, we utilize Webface to train PCA and FDA models, and 3,500 dim are kept for final matching. Although the performance still no better than deep learning based methods, significant improvement can be found from 0.8933 (no outside data) to 0.9450 (with Webface). This is mainly because each subject in Webface has about 50 faces which offer a relatively more accurate estimation of within-class scatter compared with LFW itself in which each subject has only two faces in average.

4.3 Evaluation of proposed approach on CFW-p

Section 1 shows the evaluation of the proposed approach under face verification scenario. Since face identification is also an important classification scenario that should be taken into account, in this sub-section, we introduce a very large-scale database along with an identification protocol based on it.

4.3.1 Database and experimental setup

Celebrity Faces on the Web (CFW) [29] is a very large-scale database of celebrity face images collected from the web by Microsoft Research Asia, and the released version contains 202,792 faces of 1,583 subjects. Each subject in CFW has more distinctive images (average 128 images for one subject), and these images include more complex, real, and challenging variations (see Fig. 14). However, the officially provided identity labels in CFW cannot be directly used as ground truth for identification, because they are generated by an automatic face annotation system proposed in [29] which would inevitably involve some mistakes. To fix this problem, for every image we invite three volunteers to check whether the claimed label is correct/incorrect or uncertain. Only images with three positive confirmations, i.e., all the three volunteers agree with the correctness of the claimed label, are preserved (153,461 faces of 1,520 subjects are preserved after the above purification, and the purified database is named CFW-p[1)]). When volunteers encounter an unknown celebrity, they are required to look at top Google and Bing image search results to get familiar with the visual appearance of the celebrity. In addition, the three volunteers are also required to annotate five main facial landmarks for each image, i.e., geometric centers of two eyes, tip of nose, and two corners of the mouth.

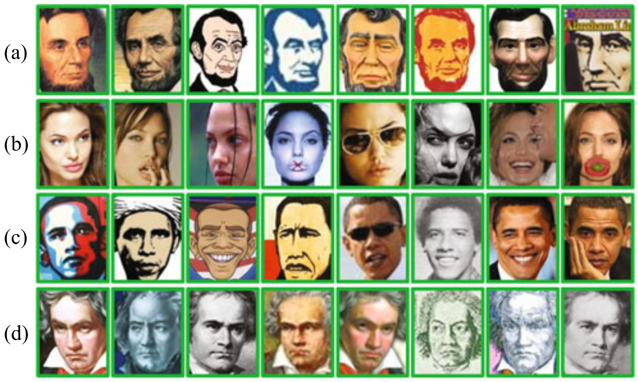


Fig. 14 Example images of five subjects from CFW-p. (Compared with FRGC v2.0, CFW-p is collected from more general web images including an large amount of user edited pictures, e.g., comic portraits, oil paintings, watercolor paintings, sketches, and sculptures) (a) Abraham Lincoln; (b) Angelina Jolie; (c) Barack Obama; (d) Van Beethoven

[1)] We will release the purified CFW, i.e., CFW-p, with 153,461 faces of 1,520 subjects along with corresponding five landmarks' coordinates. Please feel free to send email to the authors

As FRGC ver2.0 is designed for the face verification scenario, we design an identification protocol on CFW-p to further validate the effectiveness of the proposed approach. Specifically, we randomly select 520 subjects to form the training set which contains 54,863 images, and do 10-fold cross validation on the remaining 98,598 images of 1,000 subjects. For each fold, only one image per subject is randomly selected as gallery. As for measurement, we ask each experimenter to report the **estimated mean accuracy** and the **standard error of the mean**. In particular, the **estimated mean accuracy** is given by

$$\hat{\mu} = \frac{1}{10}\sum_{i=1}^{10} p_i, \tag{27}$$

where p_i is the percentage of correct classifications in the ith fold. The **standard error of the mean** is given as

$$S_E = \frac{\hat{\sigma}}{\sqrt{10}}, \tag{28}$$

where $\hat{\sigma}$ is the estimate of the standard deviation, given by

$$\hat{\sigma} = \sqrt{\frac{1}{9}\sum_{i=1}^{10}(p_i - \hat{\mu})^2}. \tag{29}$$

4.3.2 Experimental results and analysis

In this part, we include several representative methods [49–51] in LFW, multiple feature fusion methods [41–43], and deep learning based method [48]. More specifically, we take the source codes of [41, 42] and [48], and employ the defaulted parameters suggested by the authors. For the other comparative methods, we carefully implement them [43, 49–51] by ourselves. Table 3 summarizes the performance of the proposed approach on CFW-p. Like Table 3, we evaluate the proposed approach in several ways for better validation. It can be seen that the proposed approach achieves only about 15% of correct identification accuracy under the challenging protocol, where the probe set is extremely large, and the gallery set is relatively small, i.e., only one image per subject. From Table 3, we can again find the similar observations with FRGC ver2.0 and LFW that the combination of Gabor magnitude feature and SPLPQ phase feature indeed plays a key role for more effective face representation (for more intuitive interpretation, please refer to Fig. 15), and multiple face models again prove themselves to be reliable tools for system performance boosting. Another point to note is that the external face model starts to show its advantage against the internal face model in CFW-p, and the reason behind this observation is that, compared with the unified background in FRGC ver2.0, background in CFW-p contains relative abundant information which has high correlation with identity. Again, deep learning based method achieves the best performance among all the listed methods, especially with target data for fine tuning. However, it may not be fair enough to compare VGGFace with the proposed method. Because, a good deep model usually relies on a large-scale training set (2.6 million

Table 5 Experimental results of the proposed approach on CFW-p (estimated mean accuracy and standard error of the mean)

Approach	Feature	Result
LBP PLDA [49]	LBP	0.0733 ± 0.0025
Tom-vs-Pete [50]	T-P scores	0.1087 ± 0.0019
HighDimLBP [51]	High-Dim LBP	0.1296 ± 0.0017
MultOSS [41]	4 features	0.0798 ± 0.0031
MLHR [42]	Gabor+SPLPQ	0.1361 ± 0.0029
RSSM [43]	Gabor+SPLPQ	0.1437 ± 0.0023
VGGFace [48]	DCNN feature	0.1927 ± 0.0015
VGGFace (CFW Finetune) [48]	DCNN feature	**0.2457 ± 0.0018**
Gabor (Internal face)	Gabor	0.0926 ± 0.0014
Gabor (Transitional face)	Gabor	0.1060 ± 0.0014
Gabor (External face)	Gabor	0.1089 ± 0.0015
SPLPQ (Internal face)	SPLPQ	0.1205 ± 0.0019
SPLPQ (Transitional face)	SPLPQ	0.1358 ± 0.0021
SPLPQ (External face)	SPLPQ	0.1279 ± 0.0018
Gabor+LPQ (Internal face)	Gabor+SPLPQ	0.1212 ± 0.0018
Gabor+LPQ (Transitional face)	Gabor+SPLPQ	0.1386 ± 0.0021
Gabor+LPQ (External face)	Gabor+SPLPQ	0.1390 ± 0.0018
Gabor (Multiple face models)	Gabor	0.1199 ± 0.0017
SPLPQ (Multiple face models)	SPLPQ	0.1475 ± 0.0022
Final fusion approach	**Gabor+SPLPQ**	**0.1476 ± 0.0021**

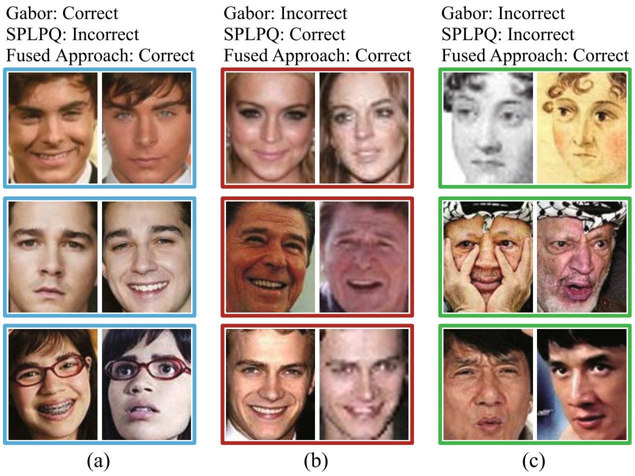


Fig. 15 Illustration of the complementarity between Gabor magnitude feature and SPLPQ phase feature on CFW-p. (a) and (b) show sample pairs recognized correctly by one feature, but recognized incorrectly by the other feature. (c) shows sample pairs that neither feature works on. Please note that the two faces in each pair shown here come from the same subject, where the left one is gallery, the right one is probe, and all the pairs shown in (a), (b) and (c) are recognized correctly by the fused approach

faces) which may has a considerable proportion of overlap with test database.

5 Conclusion and discussion

Inspired by the complementarity between magnitude and phase features and the biological cognitive mechanism of judging identity, a multiple face models based feature fusion approach was proposed to solve the uncontrolled face recognition problem. In the proposed approach, the magnitude feature is extracted by Gabor wavelets transform, and the phase feature is extracted by spatial pyramid based local phase quantization (SPLPQ), the fusion of the two features is embedded into a multiple face models framework which consists of several face images with the same size but very different scopes of visual facial region. To reduce the high dimensionality of the features and increase discriminability, blockwise fisher discriminant analysis (BFDA) is utilized in this paper. The proposed fusion approach is extensively evaluated and compared with previous approaches on FRGC ver2.0, LFW, and CFW-p, and the experimental results indicate that the proposed approach achieves better or comparable result than the best known ones. In particular, on FRGC ver2.0, the proposed approach achieved impressive 96.20%, 95.76% and 95.32% VRs (when FAR=0.1%) under ROC I/II/III of Exp.4 respectively, impressively surpassing all the best known results.

To summarize the proposed approach, we can attribute its favorable performance to the following aspects, which should be valuable to researchers in this area. First, the combination of magnitude and phase features has a key role. Experimental results confirm that they are indeed complementary for distinguishing faces, i.e., Gabor magnitude feature captures the structure information of the face, whereas SPLPQ phase feature is more robust to misalignment and blurred image by efficiently encoding the facial texture. Secondly, multiple face models are a reliable engineering technique for system performance boosting because they characterize the face in different granularity levels, i.e., facial region and resolution. A suggestion for selecting them is always attempt to make the selected face models as separate from each other as possible. Thirdly, spatial pyramid is an effective method to extend the histogram based features by preserving more information.

For future work, the current implementation of multiple face models is not efficient, i.e., there is significant redundancy between each face model. That is, the time consumption of the proposed approach is three times as great as the approaches based on a single face model. The external face model contains the maximum scope of visual facial region yet has the lowest facial resolution, whereas the internal face model is with the reverse setting. Therefore, we believe that there exists a smart approach to only use one integrated face model with the maximum scope of visual facial region such as the external face model, and at the same time has the highest facial resolution such as the internal face model. Further, we will test statistic weighting methods to fuse different features.

Acknowledgements This work was partially supported by the National Basic Research Program of China (2015CB351802), the National Natural Science Foundation of China (Grant Nos. 61390511, 61222211, 61379083, 61271445), the Strategic Priority Research Program of the CAS (XDB02070004), and the Youth Innovation Promotion Association CAS (2015085).

References

1. Zhao W Y, Chellappa R, Phillips P J, Rosenfeld A. Face recognition: a literature survey. ACM Computing Surveys, 2003, 35(4): 399–458
2. Liu C J, Wechsler H. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Transactions on Image Processing, 2002, 11(4): 467–476
3. Zhang W C, Shan S G, Gao W, Chen X L, Zhang H M. Local gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition. In: Proceedings of the 10th IEEE International Conference on Computer Vision. 2005, 786–791
4. Su Y, Shan S G, Chen X L, Gao W. Hierarchical ensemble of global and local classifiers for face recognition. IEEE Transactions on Image Processing, 2009, 18(8): 1885–1896
5. Xie S F, Shan S G, Chen X L, Chen J. Fusing local patterns of gabor magnitude and phase for face recognition. IEEE Transactions on Image Processing, 2010, 19(5): 1349–1361
6. Li Y, Shan S G, Zhang H H, Lao S H, Chen X L. Fusing magnitude and phase features for robust face recognition. In: Proceedings of Asian Conference on Computer Vision. 2013, 601–612
7. Tan X Y, Triggs B. Fusing gabor and lbp feature sets for kernel-based face recognition. In: Proceedings of International Conference on Automatic Face and Gesture Recognition. 2007, 235–249
8. Chan C H, Kittler J, Tahir M A. Kernel fusion of multiple histogram descriptors for robust face recognition. In: Proceedings of Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition and Structural, Syntactic, and Statistical Pattern Recognition. 2010, 718–727
9. Cai D, He X F, Han J W. Efficient kernel discriminant analysis via spectral regression. In: Proceedings of the 7th IEEE International Conference on Data Mining. 2007, 427–432
10. Deng W H, Hu J N, Guo J, Cai W, Feng D G. Emulating biological strategies for uncontrolled face recognition. Pattern Recognition, 2010, 43(6): 2210–2223
11. Deng W H, Hu J N, Lu J W, Guo J. Transform-invariant PCA: a unified approach to fully automatic facealignment, representation, and recognition. IEEE Transactions on Pattern Analysis and Machine In-

telligence, 2014, 36(6): 1275–1284

12. Gabor D. Theory of communication. Part 1: the analysis of information. Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering, 1946, 93(26): 429–441
13. Ojansivu V, Heikkilä J. Blur insensitive texture classification using local phase quantization. In: Proceedings of International Conference on Image and Signal Processing. 2008, 236–243
14. Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions. Pattern Recognition, 1996, 29(1): 51–59
15. Lowe D G. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 2004, 60(2): 91–110
16. Bicego M, Lagorio A, Grosso E, Tistarelli M. On the use of sift features for face authentication. In: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop. 2006
17. Luo J, Ma Y, Takikawa E, Lao S, Kawade M, Lu B L. Person-specific sift features for face recognition. In: Proceedings of International Conference on Acoustics, Speech and Signal Processing. 2007
18. Mian A S, Bennamoun M, Owens R. An efficient multimodal 2D-3D hybrid approach to automatic face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(11): 1927–1943
19. Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005, 886–893
20. Cao Z M, Yin Q, Tang X O, Sun J. Face recognition with learning-based descriptor. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2010, 2707–2714
21. Albiol A, Monzo D, Martin A, Sastre J, Albiol A. Face recognition using HOG-EBGM. Pattern Recognition Letters, 2008, 29(10): 1537–1543
22. Liu Z M, Liu C J. Robust face recognition using color information. In: Proceedings of International Conference on Biometrics. 2009, 122–131
23. Shan S G, Zhang W C, Su Y, Chen X L, Gao W. Ensemble of piecewise FDA based on spatial histograms of local (Gabor) binary patterns for face recognition. In: Proceedings of International Conference on Pattern Recognition. 2006, 606–609
24. Sinha P, Poggio T. I think I know that face. Nature, 1996, 384(6608): 404
25. Davies G. Perceiving and Remembering Faces. London: Academic Press, 1981
26. Ellis H D. Aspects of Face Processing. Boston: Martinus Nijhoff Publishers, 1986
27. Phillips P J, Flynn P J, Scruggs T, Bowyer K W, Chang J, Hoffman K, Marques J, Min J, Worek W. Overview of the face recognition grand challenge. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005, 947–954
28. Huang G B, Mattar M, Berg T, Learned-Miller E. Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical Report 07-49, 2007
29. Zhang X, Zhang L, Wang X J, Shum H Y. Finding celebrities in billions of Web images. IEEE Transactions on Multimedia, 2012, 14(4): 995–1007
30. Lades M, Vorbruggen J C, Buhmann J, Lange J, Malsburg V D C, Wurtz R P, Konen W. Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions on Computers, 1993, 42(3): 300–311
31. Zhang B C, Shan S G, Chen X L, Gao W. Histogram of Gabor phase patterns (HGPP): a novel object representation approach for face recognition. IEEE Transactions on Image Processing, 2007, 16(1): 57–68
32. Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2006, 2169–2178
33. Grauman K, Darrell T. The pyramid match kernel: discriminative classification with sets of image features. In: Proceedings of International Conference on Computer Vision. 2005, 1458–1465
34. Hadjidemetriou E, Grossberg M D, Nayar S K. Multiresolution histograms and their use for recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(7): 831–847
35. Fisher R A. The use of multiple measurements in taxonomic problems. Annals of Human Genetics, 1936, 7(2): 179–188
36. Kumar N, Berg A C, Belhumeur P N, Nayar S K. Attribute and simile classifiers for face verification. In: Proceedings of International Conference on Computer Vision. 2009, 365–372
37. Tan X Y, Triggs B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In: Proceedings of International Conference on Automatic Face and Gesture Recognition. 2007, 168–182
38. Yang J, Liu C J. Color image discriminant models and algorithms for face recognition. IEEE Transactions on Neural Networks, 2008, 19(12): 2088–2098
39. Yang J, Liu C J. Horizontal and vertical 2DPCA-based discriminant analysis for face verification on a large-scale database. IEEE Transactions on Information Forensics and Security, 2007, 2(4): 781–792
40. Shih P, Liu C J. Improving the face recognition grand challenge baseline performance using color configurations across color spaces. In: Proceedings of International Conference on Image Processing. 2006, 1001–1004
41. Taigman Y, Wolf L, Hassner T. Multiple one-shots for utilizing class label information. In: Proceedings of British Machine Vision Conference. 2009, 1–12
42. Yang Y, Song J K, Huang Z, Ma Z G, Sebe N, Hauptmann A G. Multi-feature fusion via hierarchical regression for multimedia analysis. IEEE Transaction on Multimedia, 2013, 15(3): 572–581
43. Ma Z G, Yang Y, Sebe N, Hauptmann A G. Multiple features but few labels? a symbiotic solution exemplified for video analysis. In: Proceedings of the 22nd ACM International Conference on Multimedia. 2014, 77–86
44. Kumar B, Savvides M, Xie C Y. Correlation pattern recognition for face recognition. Proceedings of the IEEE, 2006, 94(11): 1963–1976
45. Liu C J. The bayes decision rule induced similarity measures. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1086–1090
46. Hwang W, Park G, Lee J, Kee S C. Multiple face model of hybrid fourier feature for large face image set. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2006, 1574–1581
47. Liu C J. Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(5): 725–737
48. Parkhi O M, Vedaldi A, Zisserman A. Deep face recognition. In: Proceedings of British Machine Vision Conference. 2015, 1–12
49. Li P, Fu Y, Mohammed U, Elder J H, Prince S J. Probabilistic models for inference about identity. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(1): 144–157

50. Berg T, Belhumeur P N. Tom-vs-Pete classifiers and identity-preserving alignment for face verification. In: Proceedings of British Machine Vision Conference. 2012

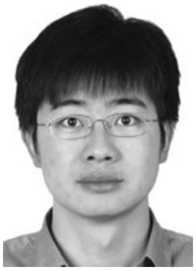
51. Chen D, Cao X D, Wen F, Sun J. Blessing of dimensionality: high-dimensional feature and its efficient compression for face verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013, 3025–3032

52. Taigman Y, Yang M, Ranzato M, Wolf L. Deepface: closing the gap to human-level performance in face verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014, 1701–1708

53. Sun Y, Wang X G, Tang X O. Deep learning face representation from predicting 10,000 classes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014, 1891–1898

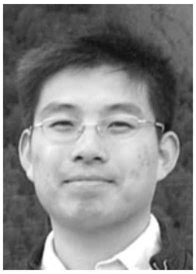
54. Schroff F, Kalenichenko D, Philbin J. Facenet: a unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015, 815–823

55. Yi D, Lei Z, Liao S C, Li S Z. Learning face representation from scratch. 2014, arXiv preprint arXiv:1411.7923v1



Yan Li received the BS degree in computer science and technology from Nankai University, China in 2010. He is currently pursuing the PhD degree with the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), China. He also spent nine months working as a Research Scholar with Lane Department of Computer Science and Electrical Engineering in the Benjamin M.Statler College of Engineering, and Mineral Resources at West Virginia University, USA. His research interests include computer vision, pattern recognition, image processing, and in particular, image and video face recognition, face retrieval, and binary code learning.

Shiguang Shan received MS degree in computer science from the Harbin Institute of Technology, China in 1999, and PhD degree in computer science from the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), China, in 2004. He joined ICT, CAS in 2002 and has been a professor since 2010. His research interests cover computer vision, pattern recognition, and machine learning. He especially focuses on face recognition related research topics. He has published more than 200 papers in refereed journals and proceedings in the areas of computer vision and pattern recognition. He is a recipient of the China's State Natural Science Award in 2015, and the China's State S&T Progress Award in 2005 for his research.



Ruiping Wang received the BS degree in applied mathematics from Beijing Jiaotong University, China in 2003, and the PhD degree in computer science from the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), China in 2010. He was a postdoctoral researcher with the Department of Automation, Tsinghua University, China from 2010 to 2012. He also spent one year working as a research associate with the Institute for Advanced Computer Studies, at the University of Maryland, College Park, from Nov. 2010 to Oct. 2011. He has been with the faculty of the ICT, CAS since July 2012, where he is currently an associate professor. His research interests include computer vision, pattern recognition, and machine learning.

Zhen Cui received the BS, MS, and PhD degrees from Shandong Normal University, Sun Yat-sen University, and Institute of Computing Technology (ICT), Chinese Academy of Sciences, China in 2004, 2006, and 2014, respectively. He was a research fellow in the Department of Electrical and Computer Engineering at National University of Singapore (NUS), Singapore from 2014 to 2015. He also spent half a year as a Research Assistant on Nanyang Technological University (NTU) from Jun. 2012 to Dec. 2012. Currently, he is an associate professor of Southeast University, China. His research interests cover computer vision, pattern recognition and machine learning, especially focusing on deep learning, manifold learning, sparse coding, face detection/alignment/recognition, object tracking, image super resolution, emotion analysis, etc.

Xilin Chen received the BS, MS, and PhD degrees in computer science from the Harbin Institute of Technology, Harbin, China in 1988, 1991, and 1994, respectively. He was a professor with the Harbin Institute of Technology from 1999 to 2005. He has been a professor with the Institute of Computing Technology, Chinese Academy of Sciences (CAS), China. Since August 2004. He has published one book and over 200 papers in refereed journals and proceedings in the areas of computer vision, pattern recognition, image processing, and multimodal interfaces. He served as an Organizing Committee / Program Committee Member for more than 50 conferences. He is a fellow of IEEE, and a fellow of China Computer Federation (CCF).