Face Recognition by Exploring Information Jointly in Space, Scale and Orientation

Zhen Lei, Shengcai Liao, Matti Pietikäinen, *Senior Member, IEEE*, and Stan Z. Li, *Fellow, IEEE*

Abstract—Information jointly contained in image space, scale and orientation domains can provide rich important clues not seen in either individual of these domains. The position, spatial frequency and orientation selectivity properties are believed to have an important role in visual perception. This paper proposes a novel face representation and recognition approach by exploring information jointly in image space, scale and orientation domains. Specifically, the face image is first decomposed into different scale and orientation responses by convolving multiscale and multiorientation Gabor filters. Second, local binary pattern analysis is used to describe the neighboring relationship not only in image space, but also in different scale and orientation responses. This way, information from different domains is explored to give a good face representation for recognition. Discriminant classification is then performed based upon weighted histogram intersection or conditional mutual information with linear discriminant analysis techniques. Extensive experimental results on FERET, AR, and FRGC ver 2.0 databases show the significant advantages of the proposed method over the existing ones.

Index Terms—Conditional mutual information (CMI), face recognition, Gabor volume based local binary pattern (GV-LBP), Gabor volume representation, local binary pattern (LBP).

I. INTRODUCTION

FACE recognition has attracted much attention due to its potential value for applications and its theoretical challenges. In real world, the face images are usually affected by different expressions, poses, occlusions and illuminations, and the difference of face images from the same person could be larger than those from different ones. Therefore, how to extract robust and discriminant features which make the intraperson faces compact and enlarge the margin among different persons becomes a critical and difficult problem in face recognition.

Up to now, many face representation approaches have been introduced, including subspace based holistic features and local appearance features [1], [2]. Typical holistic features include the well known principal component analysis (PCA) [3], linear dis-

Z. Lei, S. Liao, and S. Z. Li are with the Center for Biometrics and Security Research and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: zlei@cbsr.ia.ac.cn; scliao@cbsr.ia.ac.cn; szli@cbsr.ia.ac.cn).

M. Pietikäinen is with the Machine Vision Group, University of Oulu, FI-90014 Oulun yliopisto, Finland (e-mail: mkp@ee.oulu.fi).

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

Digital Object Identifier 10.1109/TIP.2010.2060207

criminate analysis (LDA) [4], independent component analysis (ICA) [5], etc. PCA provides an optimal linear transformation from the original image space to an orthogonal eigenspace with reduced dimensionality in sense of the least mean square reconstruction error. LDA seeks a linear transformation by maximizing the ratio of between-class variance and within-class variance. ICA is a generalization of PCA, which is sensitive to the high-order relationship among the image pixels. Recently, Wang and Tang [6] unify PCA, LDA and Bayesian methods into the same framework and present a method to find the optimal configuration for LDA. Yan *et al.* [7] reinterpret the subspace learning from the view of graph embedding so that various methods, such as PCA, LDA, ISOMAP [8], LLE [9], LPP [10], NPE [11], MFA [7] etc. can all be interpreted under this framework. Furthermore, in order to handle the nonlinearity in face feature space, the nonlinear kernel techniques (e.g., kernel PCA [12], kernel LDA [13] etc.) are also introduced.

Local appearance features, as opposed to holistic features like PCA and LDA, have certain advantages. They are more stable to local changes such as illumination, expression and inaccurate alignment. Gabor [14], [15] and local binary patterns (LBPs) [16] are two representative features. Gabor wavelets capture the local structure corresponding to specific spatial frequency (scale), spatial locality, and selective orientation which are demonstrated to be discriminative and robust to illumination and expression changes. LBP operator which describes the neighboring changes around the central point, is a simple yet effective way to represent faces. It is invariant to any monotonic gray scale transformation and is, therefore, robust to illumination changes to some extend. Recently, some work has been done to apply LBP on the Gabor responses to obtain a more sufficient and stable representation. Zhang *et al.* [17] propose LBPs descriptor on Gabor magnitude representation and Zhang *et al.* [18], [19] perform LBP on Gabor phase information. The global and local descriptors are presented, respectively, and finally fused for face representation. These combinations of LBP and Gabor features have improved the face recognition performance significantly compared to the individual representation.

Combining information from different domains is usually beneficial for face recognition. Recent biological studies indicate that retinal position, spatial frequency and orientation selectivity properties have an important role in visual perception [20]. Therefore, in this paper, we propose to explore information jointly in space, frequency, and orientation domains to enhance the performance of face recognition.

In previous work [16]–[19], people have studied the neighboring relationship in the spatial domain of a face image by quantizing the difference into binary values. However, the relevant information among different scales and orientations are
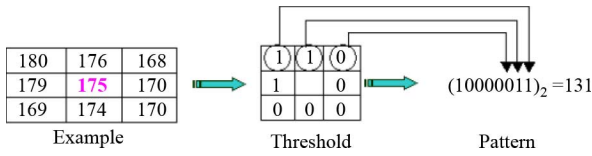
Fig. 1. Calculation of LBP code from 3×3 subwindow.

still not explored. In this paper, we propose a novel face representation method that not only explores the information in the spatial domain, but also among different scales and orientations. The main procedure of the proposed joint information extraction is as follows. First, the multiscale and multiorientation representations are derived by convolving the face image with a Gabor filter bank and formulated as a third-order volume. Second, LBP operator is applied on the three orthogonal planes of Gabor volume, respectively, named GV-LBP-TOP in short. In this way, we encode the neighboring information not only in image space but also among different scales and orientations of Gabor faces.

Zhao and Pietikäinen [21] have proposed a similar method named LBP-TOP and applied it for dynamic texture recognition and facial expression analysis. The difference is that their method is applied in the spatial and temporal domains of the video sequence, whereas ours is conducted on the Gabor face volume to explore the neighboring relationship in spatial, frequency and orientation domains. In order to reduce the computational complexity, we further propose an effective GV-LBP (E-GV-LBP) descriptor that models the neighboring changes around the central point in the joint domains simultaneously for face representation. After that, a statistical uniform pattern mechanism is adopted and local histogram features based upon the uniform patterns are extracted. Discriminant classification is finally performed based upon weighted histogram intersection or conditional mutual information (CMI) with linear discriminant analysis (LDA) techniques.

There are mainly three advantages for the proposed method. First, Gabor feature is applied to the face images to alleviate the variations of facial expression and illumination. Second, the LBP is utilized to model the neighboring relationship jointly in spatial, frequency and orientation domains. In this way, discriminant and robust information, as much as possible, could be explored. The uniform pattern mechanism is then presented to improve the efficacy of the proposed representation. Third, a feature selection and discriminant analysis method is introduced to make the face representation compact and effective for face recognition.

The rest of this paper is organized as follows. Section II briefly reviews the definition of Gabor filters and details the GV-LBP-TOP and E-GV-LBP representations based upon the Gabor faces. Section III describes the details of weighted histogram distance metric and the process of face recognition. Section IV presents the CMI based feature selection mechanism and LDA subspace learning. Experimental results and analysis are demonstrated in Section V and Section VI concludes the paper.

II. GV-LBP-TOP AND E-GV-LBP BASED FACE REPRESENTATION

A. Gabor Faces

Gabor filters, which exhibit desirable characteristics of spatial locality and orientation selectively and are optimally localized in the space and frequency domains, have been extensively and successfully used in face recognition. The Gabor kernels we used are defined as follows:

$$\psi_{\mu,\nu} = \frac{k_{\mu,\nu}^2}{\sigma^2} \exp\left(-\frac{k_{\mu,\nu}^2 z^2}{2\sigma^2}\right) \times \left[\exp(ik_{\mu,\nu}z) - \exp\left(-\frac{\sigma^2}{2}\right)\right] \tag{1}$$

where μ and ν define the orientation and scale of the Gabor kernels, respectively, $z = (x, y)$, and the wave vector $k_{\mu,\nu}$ is defined as:

$$k_{\mu,\nu} = k_\nu e^{i\phi_\mu} \tag{2}$$

where $k_\nu = k_{\max}/f^\nu$, $k_{\max} = \pi/2$, $f = \sqrt{2}$, $\phi_\mu = \pi\mu/8$. The Gabor kernels in (1) are all self-similar since they can be generated from one filter, the mother wavelet, by scaling and rotating via the wave vector $k_{\mu,\nu}$. Hence, a band of Gabor filters is generated by a set of various scales and rotations.

In this paper, we use Gabor kernels at five scales $\nu \in \{0, 1, 2, 3, 4\}$ and eight orientations $\mu \in \{0, 1, 2, 3, 4, 5, 6, 7\}$ with the parameter $\sigma = 2\pi$ [14] to derive the Gabor representation by convolving face images with corresponding Gabor kernels. For every image pixel we have totally 40 Gabor magnitude and phase coefficients, respectively, that is to say, we can obtain 40 Gabor magnitude and 40 Gabor phase faces from a single input face image.

B. Gabor Volume Based LBP on Three Orthogonal Planes (GV-LBP-TOP)

LBP is introduced as a powerful local descriptor for microfeatures of images [16]. The basic LBP operator labels the pixels of an image by thresholding the 3×3-neighborhood of each pixel with the center value and considering the result as a binary number (or called LBP codes). An illustration of the basic LBP operator is shown in Fig. 1.

Recently, the combination of Gabor and LBP has been demonstrated to be an effective way for face recognition [17]–[19]. In this paper, we propose to explore discriminative information by modeling the neighboring relationship not only in spatial domain, but also among different frequency and orientation properties. Particularly, for a face image, the derived Gabor faces are assembled by the order of different scales and orientations to form a third-order volume as illustrated in Fig. 2, where the three axes X, Y, T denote the different rows, columns of face image and different types of Gabor filters, respectively. It can be seen that the existing methods [17]–[19] essentially applied LBP or LXP operator on XY plane. It is natural and possible to conduct the similar analysis on XT and YT planes to explore more sufficient and discriminative information for face representation. GV-LBP-TOP is originated from this idea.
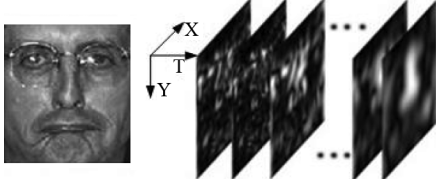
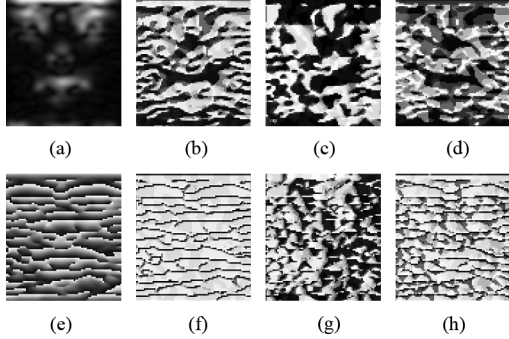Fig. 2. Face image and its corresponding third-order Gabor volume.

Fig. 3. (a) Gabor magnitude and (e) phase faces and (b), (f) their corresponding GV-LBP-XY, (c), (g) GV-LBP-XT, (d), (h) GV-LBP-YT. The first row is the result based upon Gabor magnitude representation and the second row is the one based upon Gabor phase information.

It first applies LBP analysis on the three orthogonal planes (XY, XT, and YT) of Gabor face volume and then combines the description codes together to represent faces.

Fig. 3 illustrates examples of Gabor magnitude and phase faces and their corresponding GV-LBP codes on XY, XT, and YT planes. It is clear to see that the codes from three planes are different and, hence, may supply complementary information helpful for face recognition. After that, three histograms corresponding to GV-LBP-XY, GV-LBP-XT, and GV-LBP-YT codes are computed as

$$H_j(l) = \sum_{x,y} I(f_j(x,y) = l), \quad l = 0, 1, \ldots, L_j - 1 \quad (3)$$

in which $I(\cdot) \in \{0,1\}$ is an indication function of a boolean condition and $f_j(\cdot)$ expresses the GV-LBP codes in jth plane ($j = 0$: XY; 1: XT; 2: YT), and L_j is the number of the jth GV-LBP code.

The GV-LBP-TOP histogram H is finally derived by concatenating these three histograms $H = [H_1, H_2, H_3]$ to represent the face that incorporates the spatial information and the co-occurrence statistics in Gabor frequency and orientation domains and, thus, is more effective for face representation and recognition.

C. Effective GV-LBP

The aforementioned GV-LBP-TOP is of high computational complexity. The length of the histogram feature vector and the computational cost are threefold compared to those of LGBPHS [17], so it is not very efficient in practical application. To address this problem, we propose an effective formulation of GV-LBP (E-GV-LBP) which encodes the information in spatial, frequency and orientation domains simultaneously and
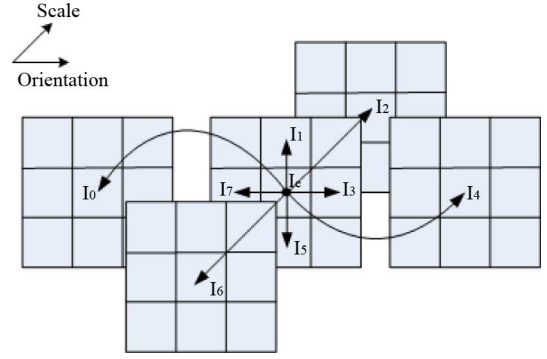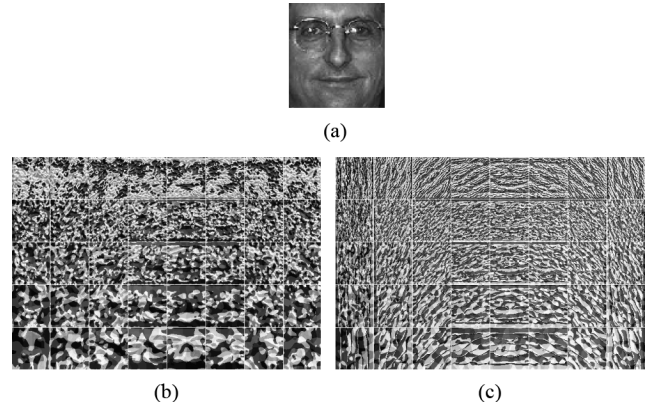

Fig. 4. Formulation of E-GV-LBP.

Fig. 5. (a) One face image and its E-GV-LBP results on (b) Gabor magnitude faces and (c) Gabor phase faces.

reduces the computational cost. Fig. 4 shows the definition of E-GV-LBP coding. For the central point I_c, I_0 and I_4 are the orientation neighboring pixels; I_2 and I_6 are the scale neighboring ones; I_1, I_3, I_5 and I_7 are the neighboring pixels in spatial domains. Like in LBP, all the values of these pixels surrounded are compared to the value of the central pixel, thresholded into 0 or 1 and transformed into a value between 0 and 255 to form the E-GV-LBP value

$$E - GV - LBP = \sum_{p=0}^{7} 2^p S(I_p - I_c) \quad (4)$$

where $S(I_p - I_c)$ is a threshold function defined as

$$S(I_p - I_c) = \begin{cases} 1, & \text{if } I_p - I_c \geq 0 \\ 0, & \text{if } I_p - I_c < 0. \end{cases}$$

Fig. 5 demonstrates the E-GV-LBP codes based upon 40 Gabor magnitude and phase faces for an input face image. The histogram features are then computed based upon the E-GV-LBP codes to provide a more reliable description as

$$H(l) = \sum_{x,y} I(f(x,y) = l), \quad l = 0, 1, \ldots, L - 1 \quad (5)$$

where $I(\cdot) \in \{0,1\}$ is an indication function of a boolean condition and $f(\cdot)$ denotes the E-GV-LBP codes, and L is the number of the E-GV-LBP codes.
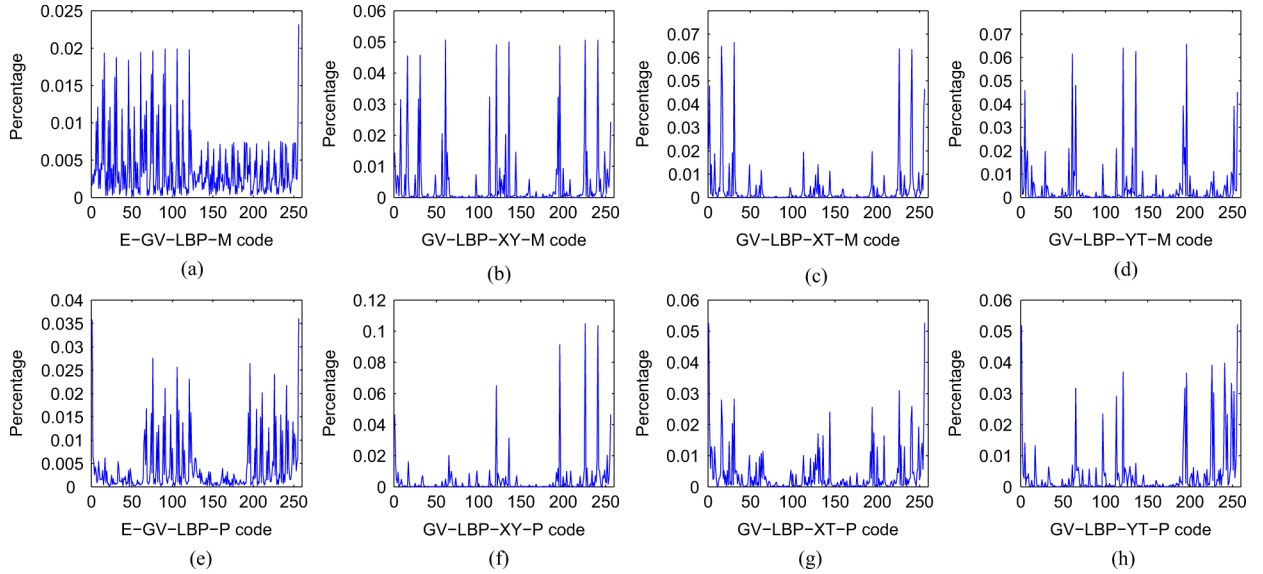
Fig. 6. Distributions of (a). (e) GV-LBP-XY, (b), (f) GV-LBP-XT, (c), (g) GV-LBP-YT, and (d), (h) E-GV-LBP codes on Gabor Magnitude and Phase images. (a)–(d) Distributions on Gabor Magnitude images and (e)–(h) are the ones on Gabor Phase images.

D. Statistical Uniform Pattern

In [16], researchers propose uniform pattern mechanism for LBP code which is robust to noise and improves the recognition performance. In LBP code, the uniform patterns are defined as such code that at most two bitwise transitions from 0 to 1 or vice versa occur when the binary string is considered circular. It is based upon the observation that there are a limited number of transitions or discontinuities in the circular presentation of the 3×3 texture patterns. Therefore, the uniform patterns occupy a vast majority proportion of all LBP patterns in local image texture.

In this paper, we adopt a more general strategy and define the uniform pattern via statistical analysis, according to the occurrence percentage instead of the number of 0–1 and 1–0 transitions for different codings.

Denote a coded face image by $f^i(x, y)$ to indicate the coding value in position (x, y) of the ith image. The occurrence distribution histogram for n face images is computed as

$$H(l) = \sum_{i=1}^{n} \sum_{x,y \in f^i} I(f^i(x, y) = l), \quad l = 0, 1, \dots, 255 \quad (6)$$

where $I(\cdot) \in \{0, 1\}$ is an indication function of a boolean condition. The GV-LBP-XY, GV-LBP-XT, GV-LBP-YT and E-GV-LBP codes distributions, calculated on the training set of FERET database are shown in Fig. 6.

The histogram is then sorted according to the occurrence percentage. In this paper, we define the uniform patterns in an iterative way. In each step, the patterns corresponding to the two smallest occurrence percentage collapse into a single one and then the histogram is resorted. Suppose we originally have K bins, after T iterations, there are $K - T$ labels left. In this work, K equals to 256 and T can be assigned arbitrarily from 0 to 255. Large $K - T$ value will result in huge feature dimension while small value may lead to the loss of discriminative information for recognition. In experiments, we try different numbers of uniform patterns and select the best one based upon the tradeoff between the recognition accuracy and computational cost.

III. WEIGHTED HISTOGRAM INTERSECTION BASED FACE RECOGNITION

The GV-LBP-TOP or E-GV-LBP histogram is utilized to represent faces, and in face recognition phase, the histogram intersection defined in (7) is used as the dissimilarity to measure different face images

$$d(H^1, H^2) = \sum_i \min(h_i^1, h_i^2) \quad (7)$$

where H^1, H^2 are two histograms and h_i^1, h_i^2 denote the ith bin value. Directly comparing the histograms based upon the whole faces may lose the structure information of faces which is important for face recognition. One possible way is to partition the face image into several blocks. The local histograms are first obtained from different blocks and then concatenated into a histogram sequence to represent the whole face. In this way, we succeed to depict the face image at three levels. The GV-LBP-TOP or E-GV-LBP codes contain information in spatial, frequency and orientation domains at pixel level. Local histogram expresses characteristic at regional level which is robust to alignment errors and finally, they are combined together as a global description for a face image to maintain both its accuracy and robustness. Pervious work has shown that different regions of face make different contributions for the performance of recognition [16], [17], e.g., the areas nearby eyes and nose are more important than others. Therefore, it is sensible to assign different weights onto different blocks when measuring the dissimilarity of two images.

Consequently, the weighted dissimilarity of different histogram sequences can be formulated as

$$D(\mathsf{H}^1, \mathsf{H}^2) = \sum_i w_i d(H_i^1, H_i^2) \quad (8)$$

where H^1, H^2 denote the two histogram sequences and w_i is the weight for the ith local histogram pair $\{H_i^1, H_i^2\}$.

In this paper, we take the similar measure as in [17] to set the weights for different blocks. For each block, we first compute the dissimilarity means m_i, m_e and variances σ_i^2, σ_e^2, for intra (the same person) and extra (different persons) sample pairs, respectively, and then the weight for the block can be computed following the Fisher criterion [22] as

$$w = \frac{(m_i - m_e)^2}{(\sigma_i^2 + \sigma_e^2)} \tag{9}$$

where m_i, σ_i^2 denote the mean and variance of intra sample pairs and m_e, σ_e^2 are those of extra sample ones.

Therefore, if the local histogram features are discriminative, where the means of intra and extra classes are far apart and the variances are small, the corresponding block will be assigned with a large weight. Otherwise, the weight will be small.

IV. CMI AND LDA BASED FACE RECOGNITION

Although the E-GV-LBP representation is more efficient than GV-LBP-TOP, it is still of very huge dimension and there is a lot of redundancy which greatly affects the efficiency in feature matching process. To deal with this problem, we utilize CMI to select the effective and uncorrelated feature set and adopt LDA to learn the discriminative feature space to improve the effectiveness and efficiency.

A. CMI Based Feature Selection

In general, if we slide the local region across the whole image, there will be millions of E-GV-LBP features for a face image. Directly comparing these features lacks efficiency and it is difficult to learn a classifier on them because of the high computational cost and limited memory storage. A straightforward way is to select a subset of original ones to reduce the dimension of features. AdaBoost learning [23] and CMI [24] based methods are two competent ways to play this role. The work in [24] shows CMI based feature selection achieves better results than AdaBoost. Therefore, in this paper, we adopt CMI to select the most discriminative and uncorrelated features to represent face compactly.

Mutual Information (MI) is a basic concept in information theory. It estimates the quantity of information shared between random variables. For two random variables X and Y, their mutual information $I(X;Y)$ is defined as follows:

$$I(X;Y) = H(Y) - H(Y|X) = H(X) - H(X|Y) \tag{10}$$

where $H(\cdot)$ is the entropy of the random variable, which measures the uncertainty of variable. For a discrete random variable X, $H(X)$ is defined as

$$H(X) = -\sum_x p(x) \log p(x) \tag{11}$$

where $p(x)$ represents the marginal probability distribution of X. The conditional entropy $H(Y|X)$ measures the remaining uncertainty of Y, when X is known.

In the context of feature selection, the main goal is to select a small subset of features $X_{v(1)}, X_{v(2)}, \ldots, X_{v(N)}$ that carries as much information as possible for classification. In our problem, when we have selected k features, our purpose is to select the next feature $X_{v(k+1)}$ to maximize the CMI $I(Y; X_{v(k+1)}|X_{v(1)}, X_{v(2)}, \ldots, X_{v(k)})$, where Y denotes the class label variable. In practice, due to the various feature values and small sample size, it is difficult to estimate the joint distribution of X directly. To address this problem, we take the following strategy. First, following the idea in [25], intra and inter personal spaces are constructed to increase the number of samples, where the image pairs from the same person form the intrasamples and the image pairs from different persons form the intersamples. Second, we quantize the feature values in intra and inter personal spaces into binary values by a predefined threshold to make the variable distribution more reliable and meanwhile simplify the computation process.

Given a set of training samples with class labels $\{(g_1, y_1), (g_2, y_2), \ldots, (g_M, y_M)\}$, where M is the sample number, and $g_i = [x_1^i, x_2^i, \ldots, x_d^i]$ is the ith intra or inter sample with d dimension features. $y_i \in \{1, -1\}$ denotes the intra and inter classes. Each dimension of feature is converted into a binary-value variable as

$$\tilde{x}_n^i = \begin{cases} 0, & x_n^i < t_n \\ 1, & x_n^i \geq t_n \end{cases}$$

where t_n is the predefined threshold which is decided by minimizing the classification error in this work as

$$t_n = \arg\min_{t_n} \left\{ \sum_{y_i=1} p(x_n^i \geq t_n) + \sum_{y_i=-1} p(x_n^i < t_n) \right\} \tag{12}$$

where p denotes the probability distribution.

For one feature, if the difference of the image pair is less than a threshold, the feature value is set to 0, otherwise it is set to 1. The feature selection principle is to find the feature that maximizes the CMI $I(Y; X_{v(k+1)}|X_{v(1)}, X_{v(2)}, \ldots, X_{v(k)})$ which is computed as

$$\begin{aligned}
&I(Y; X_{v(k+1)}|X_{v(1)}, X_{v(2)}, \ldots, X_{v(k)}) \\
&= H(Y|X_{v(1)}, \ldots, X_{v(k)}) \\
&\quad - H(Y|X_{v(k+1)}, X_{v(1)}, \ldots, X_{v(k)}) \\
&= H(Y, X_{v(1)}, \ldots, X_{v(k)}) - H(X_{v(1)}, \ldots, X_{v(k)}) \\
&\quad - H(Y, X_{v(k+1)}, X_{v(1)}, \ldots, X_{v(k)}) \\
&\quad + H(X_{v(k+1)}, X_{v(1)}, \ldots, X_{v(k)}). \tag{13}
\end{aligned}$$

However, the joint probabilities of multivariables shown previously are still difficult to estimate, especially when k is large because there are 2^k cases for k binary variables' joint distribution. A suboptimized way is to compute a series of CMIs based upon one selected feature in sequence [(14), where $v(s)$ denotes the selected feature]. The feature whose corresponding minimum of CMIs is the largest is selected as a new one. In this way, we only need to at most estimate the joint probability of ternary variables which is feasible to be calculated in practice. The whole process of CMI based feature selection is illustrated in Fig. 7

$$I(Y; X_{v(k+1)}|X_{v(s)})$$

Input: Given training samples $\{(g_1, y_1), \cdots, (g_M, y_M)\}$, where $g_i = [x_1^i, x_2^i, \cdots, x_d^i]$ is the i-th sample with d dimension original features and $y_i \in \{1, -1\}$ is the intra and inter sample label;

$w(\cdot), v(\cdot)$ denote the feature index to be selected and the feature index in the selected feature set.

For $t = 1, \ldots, T$:

  For $k = w(1), \ldots, w(d - t + 1)$:

    For $s = v(1), \ldots, v(t - 1)$:

      a) Compute the conditional mutual information of label variable and k-th feature based on the selected s-th feature respectively according to Eq. 14.

    End

  End

  b) Choose the feature where the minimum of CMI based on the selected features is the maximum. That is:

$$v(t) = \arg\max_k\{\min_s\{I(Y; X_k | X_s)\}\}$$

End

Output: The series of selected feature set $\{X_{v(1)}, X_{v(2)}, \ldots, X_{v(T)}\}$.

Fig. 7. CMI based feature selection procedure.

$$
\begin{aligned}
&= H(Y|X_{v(s)}) - H(Y|X_{v(k+1)}, X_{v(s)}) \\
&= H(Y, X_{v(s)}) - H(X_{v(s)}) \\
&\quad - H(Y, X_{v(k+1)}, X_{v(s)}) \\
&\quad + H(X_{v(k+1)}, X_{v(s)}).
\end{aligned}
\tag{14}
$$

B. LDA

LDA [4] is a representative subspace learning method which has achieved great success in face recognition. In this part, we conduct LDA on the selected features to learn the most discriminant subspace for classification. The essential idea of LDA is to disperse the samples from different classes and meanwhile gather the samples from the same class. Given the training samples $Z = \{Z_1, Z_2, \ldots, Z_n\}$, the between class scatter matrix S_b and within class scatter matrix S_w are defined as

$$S_b = \frac{1}{n} \sum_{i=1}^{L} n_i (m_i - m)(m_i - m)^T$$

$$S_w = \frac{1}{n} \sum_{i=1}^{L} \sum_{Z_j \in C_i} (Z_j - m_i)(Z_j - m_i)^T \tag{15}$$

where $m_i = 1/n_i \sum_{Z_j \in C_i} Z_j$ is the mean vector of data in class C_i, L is the class number and $m = 1/n \sum_{i=1}^{L} \sum_{Z_j \in C_i} Z_j$ is the global mean vector. LDA aims to find the projective directions W which maximize the ratio of between class scatter matrix to within class scatter one as

$$J = \frac{tr(\tilde{S}_b)}{tr(\tilde{S}_w)} = \frac{tr(W^T S_b W)}{tr(W^T S_w W)}. \tag{16}$$

The optimal projection matrix W_{opt} can be obtained by solving the following eigen-value problem

$$S_w^{-1} S_b W = W \Lambda \tag{17}$$

Fig. 8. Face examples of FERET databases.

where Λ is the diagonal matrix whose diagonal elements are eigenvalues of $S_w^{-1} S_b$.

In classification phase, after projecting the original data onto discriminant subspace, cosine distance is utilized to measure the dissimilarity of two samples in subspace.

V. EXPERIMENTS

Extensive experiments have been carried out to illustrate the efficacy of the proposed method. Specifically, three large publicly available face databases, FERET [26], AR [27] and FRGC ver 2.0 [28], are used to evaluate the performance of different methods. These face databases contain various changes of face images, including expression, lighting, aging, occlusion etc. The proposed methods, GV-LBP-TOP and E-GV-LBP, have shown their robustness and accuracy in these variations.

A. Experiment I: FERET Database

The FERET database is one of the largest publicly available databases. In this experiment, the training set contains 731 images. In test phase, the gallery set contains 1196 images from 1196 subjects. Four probe sets (fb, fc, dup1, and dup2) including expression, illumination and aging variations are used to compare the performance of different methods. All the images are rotated, scaled and cropped into 88×80 size according to the provided eye coordinates (Fig. 8).

The uniform pattern number and the block size are two parameters that impact the performance of the proposed method. Fig. 9 shows the recognition rate on fb probe set by varying the number of uniform patterns and block size for E-GV-LBP representation on Gabor magnitude and phase responses, respectively. Here, the uniform patterns are computed on the FERET training set and the unweighted histogram intersection measure (7) is used. As expected, too large or too small block size would result in a decreased face recognition rate because of the loss of spatial information or sensitivity to local variations. A smaller size of uniform pattern set would lose the discriminative information and a larger one would increase the computational cost. Considering the tradeoff between the recognition rate and computational cost, in the following experiments, the face image is divided into 11×10 nonoverlapped blocks with the size of 8×8, and the number of uniform patterns is set to 8. For the LBP, LGBP, GV-LBP-TOP and E-GV-LBP methods, weighted histogram intersection measure is adopted. The uniform codes of GV-LBP-TOP and E-GV-LBP and the weights of different blocks are statistically calculated on the FERET
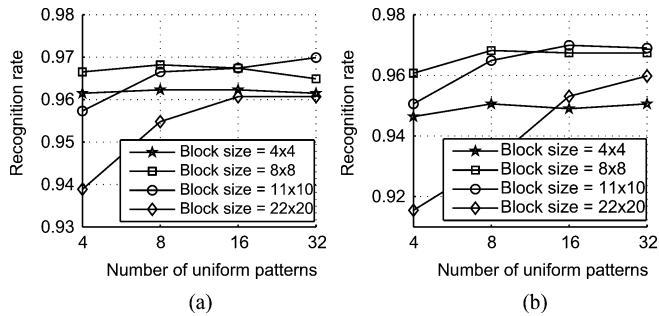
Fig. 9. Face recognition rate of E-GV-LBP on fb probe set with different uniform pattern numbers and block sizes: (a) E-GV-LBP-M and (b) E-GV-LBP-P.
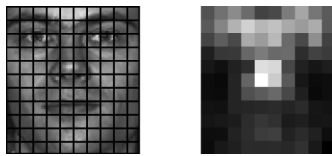

Fig. 10. Face partition and their corresponding weights for E-GV-LBP.

Fig. 11. Face recognition performance of GV-LBP-TOP and E-GV-LBP on (a) fb, (b) fc, (c) dup1, and (d) dup2 probe sets.

training set. Fig. 10 shows the face partition mode and the comparative weights of 11×10 blocks for E-GV-LBP. It is shown that the regions around eyes and nose have more contributions for face recognition which is consistent to people's intuition and the previous research work [16], [17]. In following results, the character 'M' denotes the Gabor magnitude feature and 'P' denotes the Gabor phase one.

Table I lists the recognition rates of different methods on FERET database, and Fig. 11 illustrates the corresponding cumulative match curves for the proposed method. From the results, we can observe the following.

1) The proposed methods, GV-LBP-TOP and E-GV-LBP, achieve better results than LGBP both on Gabor magnitude and phase faces which strongly demonstrates that there is complementary discriminative information among spatial, frequency and orientation domains and that the proposed descriptor is effective to explore this information for better face representation and recognition.

2) Though many previous works claimed Gabor phase information may not be robust enough for face recognition due to its sensitivity to displacement, in our experiments, the performance of methods utilizing Gabor phase information is comparable with (or even slightly better than) that based upon Gabor magnitude feature. As pointed in [18], Gabor phase feature is also able to provide discriminant information which can be fused with Gabor magnitude feature to further improve the face recognition performance.

3) Comparing the results of LBP on original face images with those of LBP on Gabor faces, the combination of Gabor and LBP effectively reduces the affect of expression, illumination and aging variations and significantly improves the performance.

4) Regarding to the complexity and accuracy, E-GV-LBP codings based upon Gabor magnitude and phase representations are the best choice among these methods for face representation and recognition.
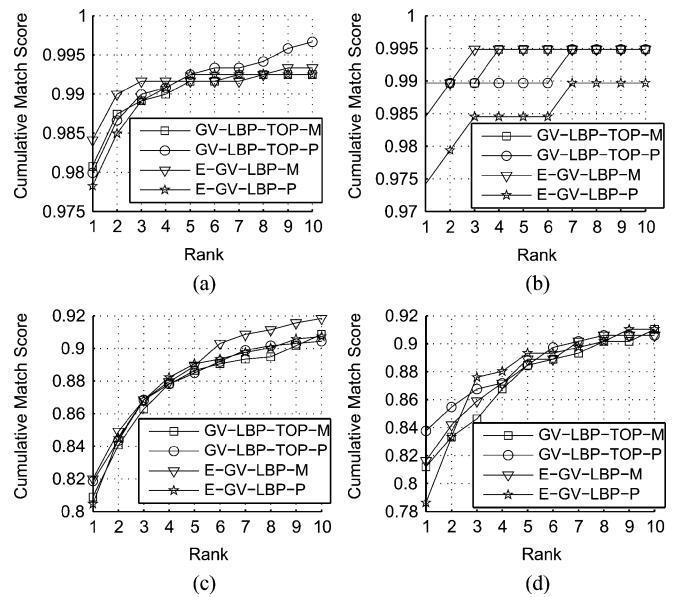
TABLE I
RECOGNITION RATES OF DIFFERENT METHODS ON THE FERET DATABASE

Methods	Fb	Fc	Dup I	Dup II
PCA	0.7891	0.0979	0.3366	0.1197
FLDA	0.8778	0.4742	0.4432	0.2009
LBP [16]*	0.97	0.79	0.66	0.64
Gabor-M+FLDA	0.9615	0.7629	0.5817	0.3419
Gabor-P+FLDA	0.9172	0.6907	0.6247	0.4188
LGBP-M [17]*	0.98	0.97	0.74	0.71
LGBP-P [18]*	0.96	0.94	0.72	0.69
GV-LBP-TOP-M	0.9808	0.9845	0.8089	0.8120
GV-LBP-TOP-P	0.9799	0.9897	0.8186	0.8376
E-GV-LBP-M	0.9841	0.9897	0.8199	0.8162
E-GV-LBP-P	0.9782	0.9742	0.8047	0.7863

* Note that the results are from the original paper.

The results reported previously are all based upon the well aligned images which were cropped using the manually labeled eye positions. However, in practice, due to various types of noise, there are usually errors for the detected eye coordinates which is so-called misalignment problem. To evaluate the robustness of different methods in this case, we disturb the eye positions of probe images by adding Gaussian white noise with different variances. In this part, we combine the four probe sets together and report the rank-1 recognition rate and the verification rate when the false accept rate is 0.001. Fig. 12 shows the face recognition performance of different methods on the different disturbed images. Since the performance of GV-LBP-TOP is close to that of E-GV-LBP, we just plot the results of E-GV-LBP in the figure for a more clear view. It can be seen that a) compared to the subspace methods such as LDA based upon pixel level, the histogram features computed on a local region are much more robust. b) With LDA, the performance of Gabor phase decreases much faster than that of Gabor magnitude. That's because phase information is more sensitive to displacement and is not so robust as Gabor magnitude responses. The histogram based features greatly alleviate this disadvantage and both the magnitude and phase features
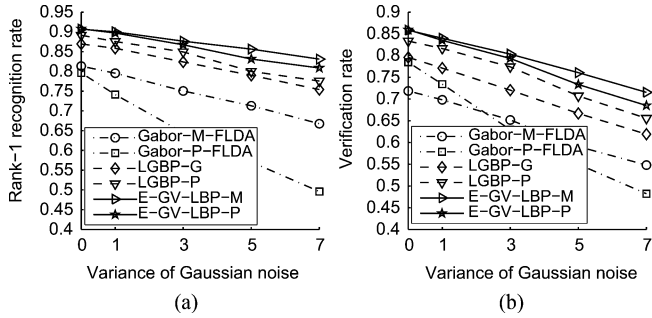
Fig. 12. Rank-1 recognition rate and $VR@FAR = 0.001$ curves with respect to different variances of Gaussian noise on FERET database.
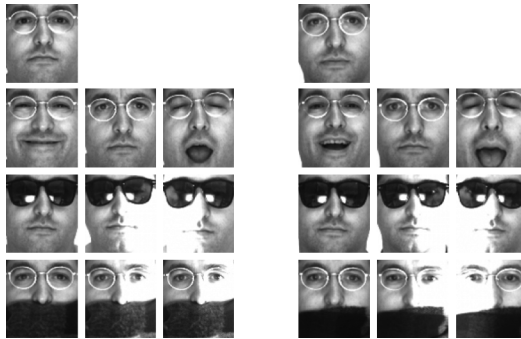

Fig. 13. Face example images for one person on AR database. The left and right ones are the images collected in two sessions. The first row is the image with neutral expression, the second to the fourth rows are the images with various expressions, sunglass and scarf occlusions, respectively.

are robust to the misalignment problem and, hence, are more practical in real application. The proposed E-GV-LBP coding combined with the statistical uniform mechanism achieves the best performance in the case of misalignment.

B. Experiment II: AR Database

The AR database consists of more than 4,000 images of 126 subjects, including 70 males and 56 females. The images were taken in two sessions separated by two weeks, considering expression (neutral, smile, anger and scream) and occlusion (sunglass and scarf) variations. In this experiment, we randomly select 45 men and 45 women, respectively, use the neutral images in the two sessions as the gallery set, and compare the performance of different methods in expression and occlusion cases. Therefore, there are 180 images in the gallery set (two images per person) and 540 images (six images per person) for the three probe sets (expression, sunglass and scarf occlusion) respectively. All the images are cropped into 88×80 according to the eye positions (Fig. 13).

Table II compares the recognition rates of different methods on AR expression, sunglass and scarf occlusion sets, respectively. For PCA and LDA related methods, the subspaces are learned from the FERET training set. For the histogram feature based methods, the way of block partition of face image and the weights for different blocks are the same as in the former experiment. The results in Table II show that the PCA and LDA related holistic methods could achieve comparatively good results on expression subset, but it is really difficult for them to deal with

TABLE II
RECOGNITION RATE OF DIFFERENT METHODS ON THE AR DATABASE

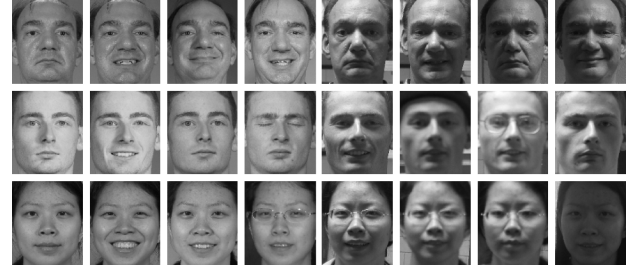| Methods	Expression	Sunglass	Scarf
PCA	0.7407	0.1296	0.0241
FLDA	0.7241	0.1185	0.0981
Gabor-M+FLDA	0.8630	0.2111	0.3148
Gabor-P+FLDA	0.7704	0.2889	0.5537
LBP [16]	0.8704	0.3463	0.4704
LGBP-M [17]	0.8611	0.3759	0.8259
LGBP-P [18]	0.8593	0.3704	0.8352
GV-LBP-TOP-M	0.9056	0.5389	0.8741
GV-LBP-TOP-P	0.9111	0.4611	0.9037
E-GV-LBP-M	0.9093	0.4722	0.8278
E-GV-LBP-P	0.8981	0.4407	0.8667

Fig. 14. Face examples of FRGC databases.

occlusion variations. The LBP descriptor is robust to expression variation, but the performance degrades dramatically in the case of sunglass and scarf occlusion. Comparatively, the performances of LGBP, GV-LBP-TOP and E-GV-LBP are much better than that of LBP, especially on scarf occlusion set which shows these methods are more robust to the occlusion compared to the LBP. However, their performance on sunglass occlusion set is still far from satisfactory. It is mainly due to the fact that the area around the eyes contains the most important clues for face recognition and the sunglass occlusion drops too much distinct information. The proposed methods, GV-LBP-TOP and E-GV-LBP, whatever with magnitude or phase representation, all yield better performance than LGBP, which strongly suggests the effectiveness of the proposed method in expression and occlusion variations.

C. Experiment III: FRGC Database

In this experiment, we evaluate the proposed E-GV-LBP representation with CMI and LDA based feature dimensionality reduction technique on FRGC database following the experiment 4 protocol which is considered as the most challenging case. The training set consists of 12776 face images from 222 individuals, including 6360 controlled images and 6416 uncontrolled ones. In the test set, there are 16028 controlled images as the target images and 8014 query images which are uncontrolled ones, from 466 persons. All the images are rotated, scaled and cropped to 142×120 according to the provided eye positions succeeded by histogram equalization preprocessing. No further preprocessing is applied. Fig. 14 illustrates some cropped face examples.

In order to explore as much information as possible, histogram features calculated on multiscale windows which are slided with a step of four pixels across the whole face are extracted. There are totally 3,474,560 features in candidate feature set. CMI is then utilized to select 6000 dimension features
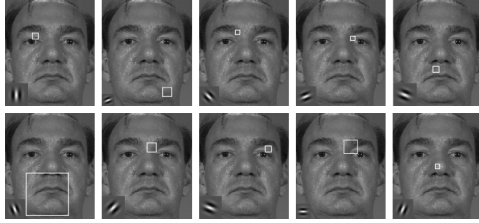
Fig. 15. First five selected E-GV-LBP-M (up) and E-GV-LBP-P (down) features. The white rectangle is the local histogram region and the left-down is the real part of Gabor kernel at the central point.

TABLE III
PERFORMANCE COMPARISON OF THE PROPOSED METHOD
AND THE EXISTING ONES

Method	ROC I	ROC II	ROC III
BEE Baseline [28]	16.08%	15.18%	14.01%
Hwang's method[30]	75.7%	75.1%	74.3%
Liu's method [31]	—	—	76%
Su's method [29]	—	—	85.8%
E-GV-LBP-M	83.47%	84.45%	85.21%
E-GV-LBP-P	82.55%	83.63%	84.79%
E-GV-LBP-M+P	88.71%	89.38%	89.99%

from the original feature set. After that, LDA learning is applied to the selected features to find 221 dimension discriminant subspace where the classification is finally performed. All of the previously shown operations are conducted on FRGC training set. In test phase, we follow the experiment 4 protocol to report the results as ROC I, ROC II, and ROC III corresponding to different time intervals.

Fig. 15 shows the first five selected features with Gabor magnitude and phase information, respectively. It can be seen the selected E-GV-LBP-M and E-GV-LBP-P features are different in Gabor kernel, feature position and local region size. Therefore, it is possible that there exists complementary information in Gabor magnitude and phase responses useful for face recognition. Table III shows the comparative results of the proposed E-GV-LBP+CMI+LDA method with some state-of-the-art methods. It should be noted our result is based upon one or two models while methods in [29] is the ensemble results of a good many of (more than 20) models. From the results, we can see that both E-GV-LBP on Gabor magnitude and phase faces are effective for face recognition and the combination of these two representations can further improve the performance, which is comparable with the state-of-the-art method, such as [29], but requires much less computational and storage cost because of the much less models used in our method.

VI. CONCLUSION

This paper proposes two novel face representations. Different from LGBP [17], we first formulate Gabor faces as a third-order volume and then apply LBP operators on three orthogonal planes (GV-LBP-TOP), encoding discriminative information not only in spatial domain, but also in frequency and orientation domains. In order to reduce the computational complexity, an effective GV-LBP (E-GV-LBP) descriptor is further

proposed to describe the changes in spatial, frequency and orientation domains simultaneously. The statistical uniform pattern mechanism is proposed to improve the effectiveness and robustness of the proposed representations. In face recognition phase, CMI and LDA are utilized to reduce the redundancy and make the representation more compact and, thus, to improve the efficiency of the algorithm. Experimental results validate the efficacy of the proposed method.

VII. ACKNOWLEDGEMENT

REFERENCES

[1] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, pp. 399–458, 2003.
[2] S. Z. Li and A. K. Jain, Eds., Handbook of Face Recognition New York, Springer-Verlag, 2005.
[3] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1991, pp. 586–591.
[4] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
[5] P. Comon, "Independent component analysis—a new concept?," *Signal Process.*, vol. 36, pp. 287–314, 1994.
[6] X. Wang and X. Tang, "A unified framework for subspace face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1222–1228, Sep. 2004.
[7] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.
[8] J. Tenenbaum, V. Silva, and J. Langford, "A global geometric framework for nonlinear dimensinality reduction," *Science*, vol. 290, no. 22, pp. 2319–2323, 2000.
[9] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 22, pp. 2323–2326, 2000.
[10] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, "Face recognition using laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, Mar. 2005.
[11] X. He, D. Cai, S. Yan, and H. Zhang, "Neighborhood preserving embedding," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2005, pp. 1208–1213.
[12] B. Schölkopf, A. Smola, and K. R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, pp. 1299–1319, 1999.
[13] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Müller, "Fisher discriminant analysis with kernels," in *Proc. Neural Netw. Signal Process.*, 1999, pp. 41–48.
[14] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 467–476, Apr. 2002.
[15] Z. Lei, S. Z. Li, R. Chu, and X. Zhu, "Face recognition with local gabor textons," in *Proc. IAPR/IEEE Int. Conf. Biometr.*, 2007, pp. 49–57.
[16] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.
[17] W. C. Zhang, S. G. Shan, W. Gao, and H. M. Zhang, "Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2005, pp. 786–791.
[18] W. C. Zhang, S. G. Shan, X. L. Chen, and W. Gao, "Are gabor phases really useless for face recognition?," in *Proc. Int. Conf. Pattern Recognit.*, 2006, pp. 606–609.
[19] B. Zhang, S. Shan, X. Chen, and W. Gao, "Histogram of gabor phase patterns (hgpp): A novel object representation approach for face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 57–68, Jan. 2007.
[20] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Rev. Neurosci.*, vol. 2, no. 3, pp. 194–203, 2001.
[21] G. Zhao and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 915–928, Jun. 2007.

[22] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Second ed. Hoboken, NJ: Wiley, 2001.

[23] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, 1997.

[24] F. Fleuret, "Fast binary feature selection with conditional mutual information," *J. Mach. Learn. Res.*, vol. 5, pp. 1531–1555, 2004.

[25] B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian face recognition," *Pattern Recognit.*, vol. 33, no. 11, pp. 1771–1782, 2000.

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

[27] A. Martinez and R. Benavente, "The AR face database, CVC," Tech. Rep 24, 1998.

[28] P. J. Phillips, P. J. Flynn, W. T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. J. Worek, "Overview of the face recognition grand challenge," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 947–954.

[29] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical ensemble of global and local classifiers for face recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2007, pp. 1–8.

[30] W. Hwang, G. t. Park, J. H. Lee, and S.-C. Kee, "Multiple face model of hybrid fourier feature for large face image set," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2006, pp. 1574–1581.

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

Zhen Lei received the B.S. degree in automation from the University of Science and Technology of China (USTC), Hefei, China, in 2005 and the Ph.D. degree from Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2010.

He is currently with the Center for Biometrics and Security Research and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His research interests are in computer vision, pattern recognition, image processing, and face recognition in particular.

Shengcai Liao received the B.S. degree in mathematics and applied mathematics from the Sun Yat-sen University, Guangzhou, China, in 2005 and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2010.

He is currently with the Center for Biometrics and Security Research and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His research interests include machine learning, pattern recognition, biometrics, and visual surveillance.

Dr. Liao was awarded the Excellence Paper of Motorola Best Student Paper and the 1st Place Best Biometrics Paper in the International Conference on Biometrics on 2006 and 2007, respectively, for his works on face recognition.

Matti Pietikäinen (S'75–M'77–SM'95) received the D.Sc. degree in technology from the University of Oulu, Finland, in 1982.

He is currently with the Machine Vision Group, University of Oulu. From 1980 to 1981 and from 1984 to 1985, he visited the Computer Vision Laboratory, University of Maryland, College Park. His research interests are in texture-based computer vision, face analysis, activity analysis, and their applications in human computer interaction, person identification and visual surveillance. He has authored about 250 papers in international journals, books, and conference proceedings, and over 100 other publications or reports. His research is frequently cited and its results are used in various applications around the world.

Dr. Pietikäinen has been Associate Editor of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and *Pattern Recognition* journals, and is currently an Associate Editor of *Image and Vision Computing* journal. He was President of the Pattern Recognition Society of Finland from 1989 to 1992. From 1989 to 2007 he served as Member of the Governing Board of the International Association for Pattern Recognition (IAPR), and became one of the founding fellows of the IAPR in 1994. He was a Vice-Chair of IEEE Finland Section.

Stan Z. Li (M'92–SM'99–F'09) received the B.Eng. degree from Hunan University, Changsha, China, the M.Eng. degree from the National University of Defense Technology, China, and the Ph.D. degree from Surrey University, Surrey, U.K.

He is currently a Professor and the Director of Center for Biometrics and Security Research (CBSR), Institute of Automation, Chinese Academy of Sciences (CASIA). He worked at Microsoft Research Asia as a researcher from 2000 to 2004. Prior to that, he was an Associate Professor at Nanyang Technological University, Singapore. His research interest includes pattern recognition and machine learning, image and vision processing, face recognition, biometrics, and intelligent video surveillance. He has published over 200 papers in international journals and conferences, and authored and edited eight books.

Dr. Li is currently an Associate Editor of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and is acting as the Editor-in-Chief for the *Encyclopedia of Biometrics*. He served as a co-chair for the International Conference on Biometrics 2007 and 2009, and has been involved in organizing other international conferences and workshops in the fields of his research interest.