Monogenic Binary Coding: An Efficient Local Feature Extraction Approach to Face Recognition

Meng Yang, *Student Member, IEEE*, Lei Zhang, *Member, IEEE*, Simon Chi-Keung Shiu, *Member, IEEE*, and David Zhang, *Fellow, IEEE*

Abstract—Local-feature-based face recognition (FR) methods, such as Gabor features encoded by local binary pattern, could achieve state-of-the-art FR results in large-scale face databases such as FERET and FRGC. However, the time and space complexity of Gabor transformation are too high for many practical FR applications. In this paper, we propose a new and efficient local feature extraction scheme, namely monogenic binary coding (MBC), for face representation and recognition. Monogenic signal representation decomposes an original signal into three complementary components: amplitude, orientation, and phase. We encode the monogenic variation in each local region and monogenic feature in each pixel, and then calculate the statistical features (e.g., histogram) of the extracted local features. The local statistical features extracted from the complementary monogenic components (i.e., amplitude, orientation, and phase) are then fused for effective FR. It is shown that the proposed MBC scheme has significantly lower time and space complexity than the Gabor-transformation-based local feature methods. The extensive FR experiments on four large-scale databases demonstrated the effectiveness of MBC, whose performance is competitive with and even better than state-of-the-art local-feature-based FR methods.

Index Terms—Face recognition, Gabor filtering, LBP, monogenic binary coding, monogenic signal analysis.

Fig. 1. Pipeline of local statistical feature-based face recognition (LSF-FR).

I. INTRODUCTION

AUTOMATIC face recognition (FR) has been extensively studied in the past two decades, yet it is still an active research topic, partially because of its many challenges in practice such as uncontrolled environments, occlusions and variations in pose, illumination, expression and time. FR has a wide range of applications, including entertainment, smart cards, information security, law enforcement, access control and video surveillance [1], [2]. Various methods have been proposed for facial feature extraction and classification, among which the representatives include subspace learning (e.g., Eigenface [5], Fisherface [6], Laplacianfaces [37], 2DPCA-based discriminant analysis [17]), discriminative models for age invariant FR [3], [4], Gabor feature based classification (GFC) [7], local binary pattern (LBP) and its variants [8]–[11], [22], [23], local appearance and shape based methods for 3-D FR [24], [38], the recently developed

sparse representation based methods [29], [30], [35], [39], [40], and the like.

Unlike many appearance-based FR methods, which are either holistic feature based (e.g., Eigenface and Fisherface) or local feature based (e.g., GFC), the adoption of LBP in FR [8] triggers the use of local statistical feature (LSF) in the FR field. Fig. 1 shows the pipeline of LSF based FR. One can see that LSF-FR methods have two main phases: statistical histogram feature extraction and statistical feature combination. Histogram feature extraction could be further divided into three steps: feature map generation (e.g., original image, Gabor feature), pattern map coding (e.g., LBP) and subregion histogram computing. Almost all the LSF-FR methods [8]–[10], [11], [22], [23] have similar procedures of subregion histogram computing (i.e., extracting the statistical information of pattern feature in each subregion), which shows certain robustness to local deformations (e.g., pose, expression, and occlusion) of face images. However, different schemes of feature map generation and pattern map coding lead to different LSF-FR methods.

The well-known LBP operator [8] directly encodes the image intensity to extract image local pattern features. In order to overcome the sensitiveness of intensity to image variations (e.g., illumination), Zhang *et al.* [9] proposed to extract directional Gabor magnitude features at multiple scales, and then apply

LBP to the Gabor magnitude maps for robust LSF. The studies of Gabor phase based LSF-FR were conducted in [10], [11], [22]. Zhang *et al.* [10] adopted multiscale Gabor phase to take the place of Gabor magnitude in [9], and the global and local variations of real part and imagery part of complex Gabor filtering coefficients were encoded in [11]. Recently, Xie *et al.* [22] utilized XOR (*exclusive or*) operator to encode the local variation of Gabor phase, and then fused Gabor-magnitude local pattern and Gabor-phase local pattern. This scheme achieves very promising FR results.

For the second phase of LSF-FR, many feature combination methods have been proposed [8]–[11], [22], [26], [36]. The most commonly used strategy is weighting the histogram feature extracted in different blocks [8]–[11]. In [26] and [36] the kernel linear discriminant analysis was used to reduce the dimension of histogram feature, and in [22] a block-based Fisher linear discriminant (BFLD) method was proposed to extract the low-dimensional discriminative features. Compared with weighting based histogram feature combination, the discriminant subspace learning methods [26], [36], [22] are much more preferred because they can not only improve the feature discrimination, but also greatly reduce the storage space.

Despite the great success of Gabor feature-based LSF-FR methods [9]–[11], [22], the expenses of Gabor transformations in these methods are also rather high, including both the computational cost and the storage space, because the Gabor transformations of an image need to be implemented at multiple scales and orientations. For example, 5-scale and 8-orientation Gabor transformations are used in [7], [9]–[11], [22]. Therefore, for a single image 80 convolutions (40 real convolutions and 40 imagery convolutions) are required to generate the Gabor features (e.g., 40 Gabor magnitude feature maps). The many convolutions and the Gabor feature maps make Gabor feature generation a process of high time and space complexity, preventing its wide acceptance in practical applications.

In our previous work [23], we investigated the use of monogenic signal analysis [12] for LSF-FR, and our proposed monogenic signal decomposition based local feature extraction has much lower time and space complexity than the Gabor filtering based methods [9], [11] but leads to competitive performance in FR. As a two-dimensional generalization of one-dimensional analytic signal representation, monogenic signal representation decomposes an image into amplitude, phase, and orientation components, which represent the signal energetic, structural, and geometric information, respectively [12]. Different from Gabor transformation, monogenic signal representation does not use steerable filters to extract multiple-orientation features, and thus it has much lower time (e.g., 3 convolutions on each scale) and space (e.g., 3 feature maps on each scale) complexity than Gabor transformations.

To further exploit the discrimination information embedded in the amplitude, phase and orientation components of monogenic signal representation, in this paper we propose an efficient and effective LSF-FR scheme, namely monogenic binary coding (MBC), which encodes the local pattern in different monogenic feature maps. The BFLD [22] is then adopted to extract the low-dimensional discriminative features from the generated MBC feature maps of amplitude, phase and

orientation, respectively. Finally, the three types of LSFs are fused for face classification. One of the most important advantages of the proposed MBC scheme is its low time and space complexity while achieving very competitive performance with the Gabor-feature based LSF-FR methods [9]–[11], [22]. The proposed MBC method is validated on benchmark large scale face databases, including Multi-PIE [49], FERET [14], [41], FRGC 2.0 [13] and PolyU-NIR [28]. The experimental results verified the efficiency and effectiveness of the proposed MBC based LSF-FR method.

The rest of the paper is organized as follows. Section II introduces briefly the monogenic signal representation. Section III presents in detail the MBC algorithm. Section IV describes the whole scheme of MBC based LSF-FR. Section V presents the experimental results, and Section 6 concludes the paper.

II. MONOGENIC SIGNAL REPRESENTATION

Monogenic signal is an important generalization of the analytic signal from one dimension (1-D) to two dimensions (2-D), and it preserves the desired properties of 1-D analytic signal. One prominent property of monogenic signal representation lies in that its feature extraction process (phase, amplitude and orientation estimation) is truly rotation-invariant [32]. Monogenic signal representation has been used in many applications such as face recognition [23], scale-space analysis [31], and texture classification [33]. In this section, we first briefly introduce the analytic signal, and then describe the monogenic signal representation.

A. Analytic Signal Representation

The analytic signal is a complex-valued representation in 1-D signal processing. Given a real valued 1-D signal $f(x)$, its complex analytic signal is defined as

$$f_{\mathrm{anal}}(x) = f(x) + j \cdot f_{\mathrm{H}}(x) \tag{1}$$

where $f_{\mathrm{H}}(x) = f(x) * h(x)$, and $h(x) = 1/\pi x$ refers to the Hilbert transform kernel in the spatial domain. In general the Hilbert transform is performed in frequency domain and the response of $h(x)$ in frequency domain is $H(\omega) = -j \cdot \mathrm{sign}(\omega) = -j\omega/|\omega|$.

The analytic signal representation allows one to retrieve the local amplitude $A(x)$ and local phase $\varphi(x)$:

$$A(x) = \|f_{\mathrm{anal}}(x)\| = \sqrt{f^2(x) + f_{\mathrm{H}}^2(x)} \tag{2}$$

$$\varphi(x) = \mathrm{atan2}(f_{\mathrm{H}}(x), f(x)), \varphi(x) \in [0, 2\pi) \tag{3}$$

where the local phase $\varphi(x)$ is invariant with respect to the local energy of the signal but changes if the local structure varies, and the local amplitude $A(x)$ is invariant with respect to the local structure but represents the local energy. Table I shows the relation between feature type and phase angle [18], respectively.

B. Monogenic Signal Representation

Monogenic signal was introduced by Felsberg and Sommer in 2001 [12] to generalize the analytic signal from 1-D to 2-D. The monogenic signal is built around the Riesz transform which is a natural multidimensional extension of the Hilbert transform
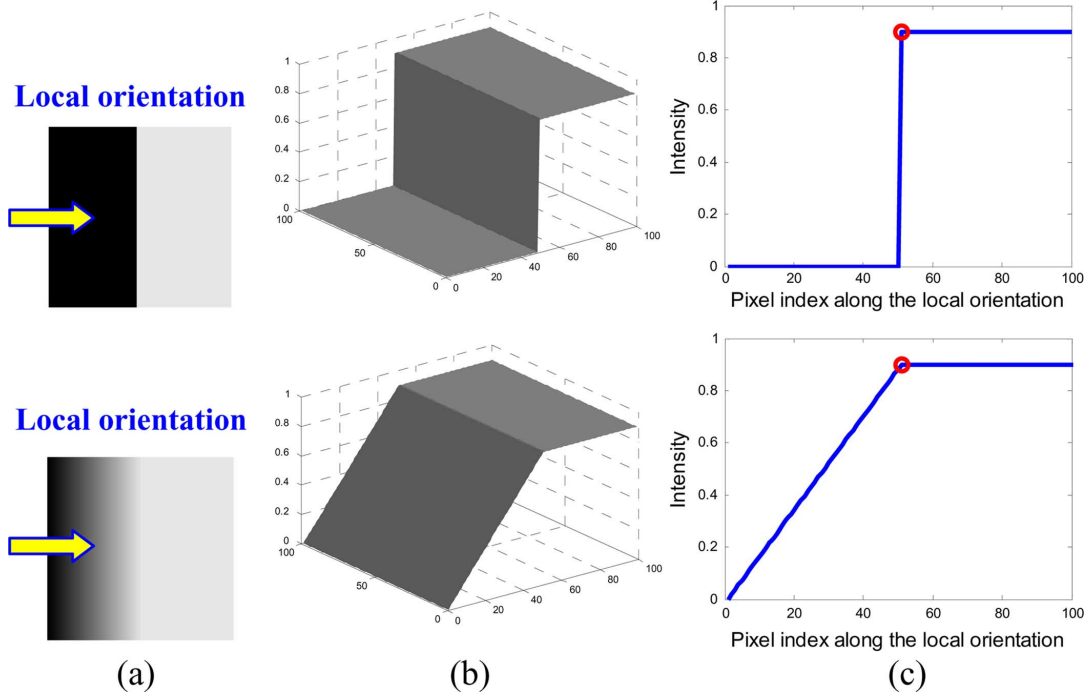
Fig. 2. Toy examples of monogenic phase. (a) Two image structures; (b) their shapes shown in 3-D; and (c) the monogenic phases of them.

TABLE I
SOME 1-D FEATURES AND THEIR CORRESPONDING PHASE ANGLES (DEGREE)

Feature						
Feature type	step edge (Positive going)	Step edge (Negative going)	Spot (Negative going)	Spot (Positive going)	Roof edge (Positive)	Roof edge (Negative)
Phase angle	0	180	270	90	44.2	224.2

[19], [32]. The Riesz transform is the scalar-to-vector signal transformation whose frequency response is $-j\boldsymbol{\omega}/\|\boldsymbol{\omega}\|$. In 2-D, the Riesz transform can be expressed as

$$\boldsymbol{f}_{\mathrm{R}}(\boldsymbol{z}) = \begin{pmatrix} f_x(\boldsymbol{z}) \\ f_y(\boldsymbol{z}) \end{pmatrix} = \begin{pmatrix} h_x * f(\boldsymbol{z}) \\ h_y * f(\boldsymbol{z}) \end{pmatrix} \qquad (4)$$

where $f(\boldsymbol{z})$ with $\boldsymbol{z} = (x, y)$ is the input signal, and filters h_x and h_y are characterized by the 2-D frequency responses $H_x = -j\omega_x/\|\boldsymbol{\omega}\|$ and $H_y = -j\omega_y/\|\boldsymbol{\omega}\|$ with $\boldsymbol{\omega} = (\omega_x, \omega_y)$. It is easy to obtain the spatial representation of Riesz kernel, which is

$$(h_x, h_y) = \left(\frac{x}{2\pi\|\boldsymbol{z}\|^3}, \frac{y}{2\pi\|\boldsymbol{z}\|^3} \right) \qquad (5)$$

For an image $f(\boldsymbol{z})$, its monogenic signal is defined as the combination of f and its Riesz transform

$$f_{\mathrm{M}}(\boldsymbol{z}) = (f(\boldsymbol{z}), f_x(\boldsymbol{z}), f_y(\boldsymbol{z})) \qquad (6)$$

In (6), $f(\boldsymbol{z})$ is called the real part of the monogenic signal, and $f_x(\boldsymbol{z})$ and $f_y(\boldsymbol{z})$ the two imagery parts. Based on the real and imagery parts, the original image signal $f(\boldsymbol{z})$ could be orthogonally decomposed into three components, local amplitude, local phase and local orientation, which are defined as [12]

$$\begin{cases} \text{local amplitude}: \ A = \sqrt{f^2 + f_x^2 + f_y^2} \\ \text{local phase}: \ \phi = -\mathrm{sign}(f_x)\mathrm{atan2}\left(\sqrt{f_y^2 + f_x^2} \Big/ f \right) \\ \text{local orientation}: \ \theta = \mathrm{atan}(f_y/f_x) \end{cases} \qquad (7)$$

where A describes the local energetic information, ϕ describes the local structural information, and θ describes the geometric information.

The local phase and local orientation compose the monogenic phase of a 2-D signal [12]. In contrast to 1-D case, the monogenic phase includes additional geometric information, which indicates the main orientation in ideal case (i.e., the signal is constant in one direction). Two toy examples of image features (e.g., step and roof) are shown in Fig. 2. Fig. 2(a) shows the step edge and roof edge features in 2-D image, with Fig. 2(b) the corresponding 3-D shapes of them. The geometric information (i.e., the main orientation) is illustrated by local orientation shown in Fig. 2(a), with the structure information shown by local phase in Fig. 2(c).

C. Multiresolution Monogenic Signal Representation

In practice, the signals are of finite length, and we need to perform band-pass filtering to an image before applying the Riesz transforms. On the other side, as indicated by [12], band-pass filtering has the benefits of maintaining the invariance-equivariance property of signal decomposition. Here the invariance-equivariance property [21], [12] means that energy (local amplitude) and structure (local phase on local orientation) are independent information.

Classical Gabor filters are commonly used as band-pass filters for that they could offer the best simultaneous localization of spatial and frequency. However, these filters overlap more in low frequencies than in high frequencies, leading to a nonuniform coverage of the Fourier domain. Moreover, Gabor filters have nonzero means, and thus they are affected by direct current (DC) components [34]. An alternative to the Gabor filters is the log-Gabor filters proposed in [20]. The log-Gabor filters discard the DC components and can overcome the bandwidth limitation of traditional Gabor filters. Meanwhile, it has a Gaussian shaped response along the logarithmic frequency scale
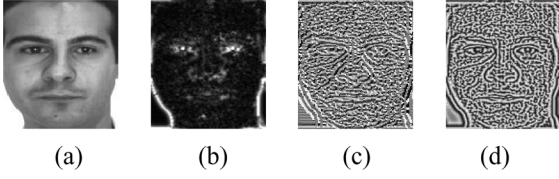
Fig. 3. (a) Face image and its monogenic representation on one scale: (b) amplitude component, (c) orientation component, and (d) phase component.

instead of a linear one. This allows more information to be captured in the high frequency band and endows desirable high pass characteristics.

The frequency response of log-Gabor filters can be described as

$$G(\omega) = \exp\{-[\log(\omega/\omega_0)]^2/2[\log(\sigma/\omega_0)]^2\} \qquad (8)$$

where ω_0 is the center frequency and σ is the scaling factor of the bandwidth. In order for filters with the constant shape ratio, we set σ/ω_0 as a constant as suggested in [43]. It is worth noting that a log-Gabor filter with a 3-octave bandwidth has the same spatial width as that of a 1-octave Gabor filter, demonstrating its ability to capture broad spectral information with compact support [42].

The band-pass monogenic signal representation is defined as

$$\begin{aligned} f_{\text{lg}-\text{M}}(\boldsymbol{z}) &= (f_{\text{lg}}(\boldsymbol{z}), f_{\text{lg}-x}(\boldsymbol{z}), f_{\text{lg}-y}(\boldsymbol{z})) \\ &= (f_{\text{lg}}(\boldsymbol{z}), h_x * f_{\text{lg}}(\boldsymbol{z}), h_y * f_{\text{lg}}(\boldsymbol{z})) \end{aligned} \qquad (9)$$

where $f_{\text{lg}}(\boldsymbol{z}) = f(\boldsymbol{z}) * F^{-1}(G(\omega))$, and F^{-1} represents the 2-D inverse Fourier transform. Similarly, the local amplitude A_{lg}, orientation θ_{lg} and phase ϕ_{lg} of a 2-D signal f can be computed by [12]

$$A_{\text{lg}} = \sqrt{f_{\text{lg}}^2 + f_{\text{lg}-x}^2 + f_{\text{lg}-y}^2}; \ \theta_{\text{lg}} = \text{atan}(f_{\text{lg}-y}/f_{\text{lg}-x});$$

$$\phi_{\text{lg}} = -\text{sign}(f_{\text{lg}-x})\text{atan2}\left(\sqrt{f_{\text{lg}-y}^2 + f_{\text{lg}-x}^2} \Big/ f_{\text{lg}}\right) \qquad (10)$$

Since log-Gabor filters are band-pass filters, usually multiscale monogenic representation is required to fully describe a signal. In multiscale log-Gabor filters, the parameters σ and ω_0 can be rewritten as

$$\sigma = \sigma_{\text{ratio}}\omega_0, \quad \omega_0 = (\lambda_{\text{min}}\mu^{s-1})^{-1} \qquad (11)$$

where λ_{min} is the minimal wavelength, μ is the multiply factor of wavelength, s is the scale index, and σ_{ratio} is the ratio σ/ω_0.

As an example, Fig. 3 shows the monogenic representation of a face image on one scale. We can see that the amplitude components reflect the local variation of gray value; while the facial local structures are well captured in the local phase and orientation components.

D. Multiscale Monogenic Signal Representation versus Multiscale Gabor Wavelet Representation

Traditional 2-D Gabor filters are highly jointly localized in spatial location, orientation and frequency. Usually 5-scale and 8-orientation Gabor filters are used in face recognition

[7], [9]–[11], [22], [36], where the Gabor features at each orientation are extracted by using steerable filters.

It is easy to see that both multiscale 2-D Gabor wavelet (i.e., Gabor filters) representation and multiscale monogenic signal representation are redundant. For multiscale Gabor wavelets, the redundancy comes from two aspects: one is the redundancy across multiple scales, and the other is the redundancy across multiple orientations on each scale. However, for multiscale monogenic signal representation, the redundancy only comes from the multiscale representation, because on each scale the local amplitude, local phase, and local orientation are orthogonal. Although multiscale Gabor filtering is effective for FR, the high cost of Gabor representation (80 convolutions for complex Gabor filtering across 5 scales and along 8 orientations) limits its application in practice. On the other side, although the multiscale monogenic signal representation has no redundancy along orientation, this does not mean that it is not effective for face representation. Our previous work in [23] has shown that the combination of orientation and amplitude components of monogenic signal representation could achieve even better recognition rates than the multiscale Gabor wavelet representation but with significantly lower time and space complexity.

III. MONOGENIC BINARY CODING

In this section, we describe the proposed algorithm of coding the monogenic signal features, which contains two parts: monogenic local variation coding and monogenic local intensity coding. The first part encodes the variation between the central pixel and its surrounding pixels in a local patch, while the second part encodes the value of central pixel itself. Obviously, these two parts are complementary.

A. Binary Coding of Monogenic Local Variation

Refer to Fig. 4, for each local patch (e.g., 3×3 neighborhood) a typical local variation coding usually consists of the following three stages: binary quantification, binary sequence generation, and binary sequence to decimal value conversion. In the procedures of local variation coding, the binary sequence generation and binary sequence to decimal value conversion are the common stages in most binary pattern coding methods, such as LBP [8], [16], LGBP [9], HGPP [11] and LGXP [22]. An example of these two steps is shown in the second row of Fig. 4. The decimal value of the binary code is defined as follows:

$$\begin{aligned} C(\boldsymbol{z}_c) &= [C_N, C_{N-1}, \ldots, C_1]_{\text{binary}} \\ &= \left[\sum_{i=1}^N 2^{i-1} \cdot C_i\right]_{\text{decimal}} \end{aligned} \qquad (12)$$

where $\boldsymbol{z}_c$ denotes the central pixel position in the local patch of feature map, C_i denotes the binary code (0 or 1) of the ith neighbor of the central pixel and N is the number of involved neighbors (we fix $N = 8$).

However, the first step of local variation coding, i.e., binary quantization, usually depends on the property of feature map, and should be designed based on the physical meaning of the feature. For example, in the binary quantization of LBP [8] which regards original image as feature map, the surrounding
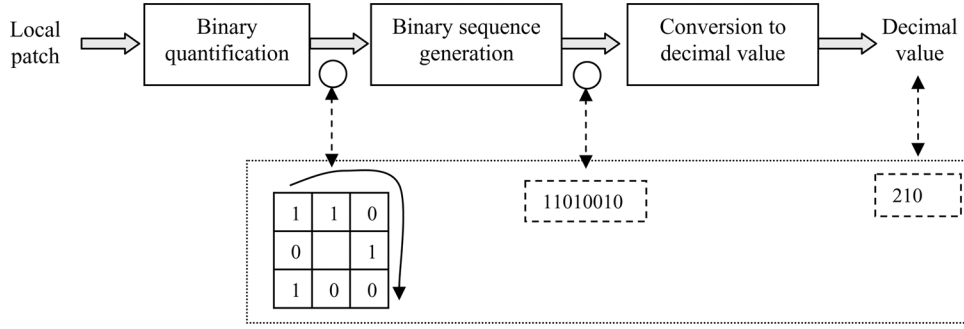
Fig. 4.   Typical process of local variation coding.

pixel is coded as 1 if its value is not less than the value of central pixel; otherwise, the surrounding pixel is coded to 0. In HGPP [11], the surrounding pixel is coded to 0 if it has the same sign as that of the central pixel; otherwise, it is quantified to 1. The monogenic signal representation has three components: local amplitude, local orientation and local phase. The local orientation and local phase compose monogenic phase. According to the properties of local amplitude and monogenic phase, different binary quantization strategies are utilized to generate the binary code.

1) Local Variation of Monogenic Amplitude: The local amplitude of monogenic signal representation is a measurement of local energetic information. For example, high amplitude usually indicates higher energetic local features (e.g., edges, lines, textures). Therefore, the local variation of monogenic amplitude could be coded by comparing the amplitude value of central location with those of its neighbors, which is similar to LBP [8]. For a local patch, let's denote by $A(z_c)$ the amplitude value of the central pixel z_c, and by $A(i)$ the amplitude value of the ith neighbor of the central pixel. Usually, we normalize the range of the amplitude to [0, 255]. Then the amplitude binary code of the ith neighbor is defined as

$$C_i^A(z_c) = \begin{cases} 1 & \text{if } A(i) \geq A(z_c) \\ 0 & \text{else} \end{cases} \tag{13}$$

Consequently, the amplitude binary code, denoted by C_A, of the central pixel z_c is formed as $C_A(z_c) = [C_N^A, C_{N-1}^A, \ldots, C_1^A]_{\text{binary}}$.

2) Local Variation of Monogenic Phase: For 2-D signals, monogenic phase is decomposed into local orientation and local phase. The local orientation describes the geometric information, more specifically the main orientation of the local structure, and the local phase describes the information of the local structure along the main orientation (i.e., structure type as shown in Table I). Both local orientation and local phase are represented by angle degree. In this paper, we unwrap both of them into the range of [0, 360). Because the angle of local phase or local orientation indicates the structural or geometric information, it is more meaningful to compare the feature type than to directly compare the angle's degree. Therefore, the binary coding of local phase and local orientation variations should adopt a different strategy from that of local amplitude. Inspired by the coding strategy of Gabor phase in HGPP [11] and LGXP [22], we consider two cases when comparing the monogenic features: one is that the local phases/orientations

are the same or close enough to each other and the other is that the local phases/orientations are different.

In order to increase the robustness of binary coding, we firstly divide the range [0, 360) into K intervals (we fix $K = 4$ in this paper). If two phases/orientations belong to the same interval, they are believed to be similar local features; otherwise, they are regarded as different. For a local patch, denote by $\phi(z_c)$ the local phase of the central pixel z_c, and by $\phi(i)$ the phase of its ith neighbor. The phase binary code of the ith neighbor is

$$C_i^P(z_c) = \begin{cases} 0 & \text{if } Q(\phi(z_c)) = Q(\phi(i)) \\ 1 & \text{else} \end{cases} \tag{14}$$

where $Q(x)$ is the quantification function, defined as

$$Q(x) = q \quad \text{if } \frac{360 \cdot (q-1)}{K} \leq x < \frac{360 \cdot q}{K} \tag{15}$$

Then the local phase binary code, denoted by C_P, of the central pixel z_c is formed as $C_P(z_c) = [C_N^P, C_{N-1}^P, \ldots, C_1^P]_{\text{binary}}$.

Similarly, let's denote by $O(z_c)$ the local orientation of the central pixel z_c, and by $O(i)$ the local orientation of its ith neighbor. The orientation binary code of the ith neighbor is

$$C_i^O = \begin{cases} 0 & \text{if } Q(O(z_c)) = Q(O(i)) \\ 1 & \text{else} \end{cases} \tag{16}$$

And then the local orientation binary code, denoted by C_O, of the central pixel z_c is formed as $C_O(z_c) = [C_N^O, C_{N-1}^O, \ldots, C_1^O]_{\text{binary}}$.

B. Binary Coding of Monogenic Local Imagery Intensity

Almost all the LSF-FR methods, including LBP [8], LGBP [9] and LGXP [22], code the variation information of a patch for pattern recognition but omit the feature information of the central pixel in the local patch. In fact, the information of central pixel itself can have discriminative information which may not be carried out by the local variation. For instance, the two pixels having the same local variation pattern may have very different intensities. In [25], Guo *et al.* proposed to quantify the local intensity of central pixel based on the average gray level of the whole image for texture classification; however, this binary quantization scheme may not be robust, especially for FR with illumination changes.

In the monogenic signal representation, the Riesz transform is asymmetric and it can suppress the DC component [12]. Such properties make the imaginary part of Riesz transform able to enhance the feature intensity while being robust to disturbances
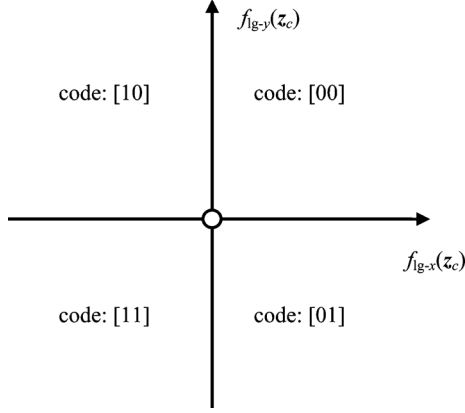
Fig. 5. Quadrant bit coding of monogenic imaginary feature at each location.

Method	MBC-X	LBP	Gabor-related coding methods
Number of patterns	1024	256	256

this sparser histogram can still have enough (even higher) discrimination capability and result in very good FR performance (please refer to the section of experimental results).

The above three MBC maps can be used individually as the feature for face classification, and they can also be used together to enhance the FR performance, as we will describe in the next section.

IV. FACE RECOGNITION BY MONOGENIC BINARY CODE

A. Face Recognition by Individual MBC

The statistical information of image local areas can be described by local histograms, which are robust to the image occlusion and variations of pose, expression, and noise, etc. After computing the multiscale MBC-A, MBC-P and MBC-O feature maps of the face image, we can construct three histograms, $H_{\mathrm{MBC-A}}$, $H_{\mathrm{MBC-P}}$ and $H_{\mathrm{MBC-O}}$, through the following procedures. For each kind of pattern map on each scale, it is partitioned into multiple nonoverlapping regions, and then the local histogram is built for each subregion. Finally, all the local histograms across different scales and different regions are concatenated into a single histogram vector to represent the face image. Formally, the $H_{\mathrm{MBC-A}}$, $H_{\mathrm{MBC-P}}$ and $H_{\mathrm{MBC-O}}$ histograms are formulated as

$$H_t = (h_t(r,s), r = 1, \ldots, L;\ s = 1, 2, 3),$$
$$t \in \{\mathrm{MBC-A}, \mathrm{MBC-P}, \mathrm{MBC-O}\} \tag{21}$$

where L is the number of subregions on each scale; s denotes the scale index; and $h_t(r,s)$ is the histogram of feature map t in the rth subregion on scale s.

To be identical to other LSF based FR methods [9]–[11], [22], the similarity between two histograms h_a and h_b is defined as their intersection:

$$S_H(h_a, h_b) = \sum_{i=1}^{M} \min\left(h_a^i, h_b^i\right) \tag{22}$$

where M is the number of bins in the histogram, and h_a^i and h_b^i denote the frequency of the ith bin in h_a and h_b, respectively. By measuring the histogram intersection, the similarity of $H_{\mathrm{MBC-A}}$, $H_{\mathrm{MBC-P}}$ and $H_{\mathrm{MBC-O}}$ histogram features from gallery and probe face images could be computed for classification.

B. Face Recognition by Fused MBC

It is obvious that different subregions in the face image will have different discriminative power for FR. Therefore, many methods [8]–[11], [23] assign different weights to the histograms of different subregions before matching. This is actually a feature combination process and could improve

such as illumination changes. Therefore, in this paper we propose to use the imagery part of monogenic signal representation to encode the local feature intensity information of the central pixel. Since there are two components (e.g., $f_{\mathrm{lg}-x}$ and $f_{\mathrm{lg}-y}$) in the imagery part of monogenic signal representation, the binary coding leads to a quadrant bit for the monogenic imaginary feature at a location, as illustrated in Fig. 5.

Refer to Fig. 5, the imagery part of monogenic signal representation at pixel z_c is encoded into two bits, $[C_x^I(z_c), C_y^I(z_c)]_{\mathrm{binary}}$, by the following rule:

$$C_d^I(z_c) = \begin{cases} 0 & \text{if } f_{\mathrm{lg}-d}(z_c) > 0 \\ 1 & \text{if } f_{\mathrm{lg}-d}(z_c) \le 0 \end{cases}, \quad d \in \{x, y\} \tag{17}$$

where $f_{\mathrm{lg}-x}$ and $f_{\mathrm{lg}-y}$ (refer to (9) please) are respectively the horizontal and vertical Riesz transform outputs of monogenic signal representation.

C. Monogenic Binary Code (MBC)

With the coding procedures described in Sections III-A and III-B, for each of the amplitude, phase and orientation components of the monogenic signal representation, we can have a monogenic binary code (MBC) map. We denote by MBC-A, MBC-P and MBC-O the code maps for amplitude, phase and orientation, respectively. At each location, MBC-A, MBC-P and MBC-O are formed by combining the local imagery intensity code and the local variation code as follows

$$\mathrm{MBC-A}(z_c) = \left[C_x^I(z_c), C_y^I(z_c), C_A(z_c)\right]_{\mathrm{binary}} \tag{18}$$

$$\mathrm{MBC-P}(z_c) = \left[C_x^I(z_c), C_y^I(z_c), C_P(z_c)\right]_{\mathrm{binary}} \tag{19}$$

$$\mathrm{MBC-O}(z_c) = \left[C_x^I(z_c), C_y^I(z_c), C_O(z_c)\right]_{\mathrm{binary}} \tag{20}$$

It can be seen that when the 8 closest neighbors of a pixel are involved in local variation coding, each MBC pattern will have 10 bits. Then the number of possible patterns for each MBC is 1024, which is larger than that of previous binary coding methods such as LBP, LGBP, HGPP and LGXP (refer to Table II please). The more patterns can characterize more accurately the local signal structure; at the same time, it will also make the generated histogram in each subregion a little sparser. However,

the FR rate. However, this weighting scheme cannot reduce the feature dimensionality. Considering that the kernel linear discrimination analysis adopted in [26], [36] has high time complexity, we adopt the BFLD [22] scheme to reduce the histogram feature dimension while enhancing its discrimination. The MBC feature map is first partitioned into $M_b \times M_b$ blocks, and then each block is further partitioned into $M_r \times M_r$ subregions. In each subregion, the histogram of feature map on each scale is built as described in Section IV-A. The histogram feature of each block is the concatenation of all the histograms of its subregions. Then for the histogram feature of each block, LDA [6] is used to learn a projection matrix from the training set, and then the dimensionality reduced histogram feature can be obtained by using this projection matrix.

Denote by $F_t^p(i)$ and $F_t^g(i)$ the dimensionality reduced features of H_t in the ith block of a probe face image and a gallery image, respectively, where $t \in \{\mathrm{MBC-A}, \mathrm{MBC-P}, \mathrm{MBC-O}\}$. Since F_t^p and F_t^g are not histogram features anymore, we compute their similarity using the cosine distance:

$$S_t\left(F_t^p(i), F_t^g(i)\right) = \frac{\langle F_t^p(i), F_t^g(i)\rangle}{\|F_t^p(i)\| \|F_t^g(i)\|} \qquad (23)$$

where $\langle \bullet \rangle$ is the inner product operator. The cosine distance is widely used in face recognition [22], [26], [36], and in [45] it has been reported that the cosine distance performs better than L_1, L_2 and Mahalanobis distances in most face recognition experiments. In addition, using cosine distance could also make the comparison with [22] fair. The similarity between the whole probe and gallery images is computed as

$$S_t\left(F_t^p, F_t^g\right) = \sum_{i=1}^{M_b \times M_b} S_t\left(F_t^p(i), F_t^g(i)\right) \qquad (24)$$

The monogenic amplitude, phase and orientation components describe the energetic, structural and geometric information of a 2-D signal, respectively, and they are complementary to each other. Therefore, it is useful to fuse these three types of features for a more robust FR result. In this paper, we use the simple weighted average to fuse the three similarities, i.e., $S_{\mathrm{MBC-A}}$, $S_{\mathrm{MBC-P}}$ and $S_{\mathrm{MBC-O}}$, as follows

$$S_{\mathrm{MBC-F}} = (1 - 2 \cdot w)S_{\mathrm{MBC-A}} + wS_{\mathrm{MBC-P}} + wS_{\mathrm{MBC-O}} \qquad (25)$$

where w is the weight. Note that in (25) we let the weights assigned to $S_{\mathrm{MBC-P}}$ and $S_{\mathrm{MBC-O}}$ equal because the local orientation and local phase have almost the same performance in various databases. Therefore, there is only one weight parameter, i.e., w, to be determined in our fusion scheme, and accordingly we call the FR scheme with similarity score $S_{\mathrm{MBC-F}}$ as fused MBC, denoted by MBC-F.

V. EXPERIMENTAL RESULTS

In this section, we first discuss the parameter settings, and then give the time and space complexity comparison between MBC and some representative and state-of-the-art LSF-FR methods, including LGBP [9], HGPP [11] and LGXP [22]. Finally, we evaluate the proposed MBC algorithm on four large

scale benchmark face databases: Multi-PIE [49], FERET [14], [41], FRGC 2.0 [13] and PolyU-NIR [28]. The Matlab source code of the proposed MBC algorithm can be downloaded at http://www4.comp.polyu.edu.hk/~cslzhang/code.htm.

A. Parameter Settings

There are some parameters in different stages of MBC, e.g., multiscale log-Gabor filtering, subregion histogram computing, and feature combination by BFLD [22]. In order to be consistent with other methods and provide a fair comparison, no preprocessing is used in all the following experiments.

Refer to Section II-C, the parameters in multiscale log-Gabor filtering are the minimal wavelength $\lambda_{\min}$, the ratio factor σ_{ratio}, the multiply factor of wavelength μ and the number of scale (i.e., the scale index s). In our paper, if no specific instruction, we fix $\lambda_{\min} = 4, \mu = 0.64, \sigma_{\mathrm{ratio}} = 1.7$, and the number of scale as 3. Because one-scale log-Gabor filter could capture more frequency information than one-scale Gabor filter [42], [43], the number of scales in multiscale log-Gabor filtering is often set to 3 [23], [47], [48]. Since the number of scales will affect much the time and space complexity of the proposed method, we perform two experiments on Fb and DupI datasets of the FERET database to discuss the selection of scale parameter s.

The detailed experimental settings on Fb and DupI are presented in Section V-D. We change the number of scales from 1 to 4, and plot the FR rates of MBC versus the scale parameter s in Fig. 6. It could be seen that in most cases (especially for the test in DupI), the recognition rate will benefit from the increase of s from 1 to 3 or 4. It is interesting to see that in the test on Fb, there are some fluctuations as the scale number increases. This is possibly because the test data in Fb has little variation compared to the gallery set Fa, so it is not necessary to use multiscale representation. Nevertheless, considering the balance of recognition rate and complexity, particularly in more challenging FR scenarios (e.g., DupI), we prefer to set the scale number to 3 in all the experiments.

In the stage of subregion histogram computing, the whole image needs to be divided into $M_b \times M_b$ blocks, which is further divided into $M_r \times M_r$ subregions. In each subregion, the histogram is extracted; and in each block, the discriminative feature is extracted by BFLD. In the stage of fusing, three similarities, i.e., $S_{\mathrm{MBC-A}}$, $S_{\mathrm{MBC-P}}$ and $S_{\mathrm{MBC-O}}$, will be averaged by using the weight w. If no specific instruction, we fix $M_r = 2$, $M_b = 4$, $w = 0.275$, and the dimensionality of discriminative feature in each block as 200.

B. Time and Space Complexities

In the representative LSF based FR methods [8]–[11], [22], [23], the feature map generation, pattern map coding, histogram computing, and histogram similarity computing are the main time-consuming parts. The number of image convolution directly decides the computational burden of feature map generation, while the number of pattern maps directly affects the running time of other parts and space complexity of algorithms. Fig. 7 compares the time and space complexity of MBC (including MBC-A, MBC-O and MBC-P) and MBC-F with some state-of-the-art LSF-FR methods, including LGBP [9], HGPP
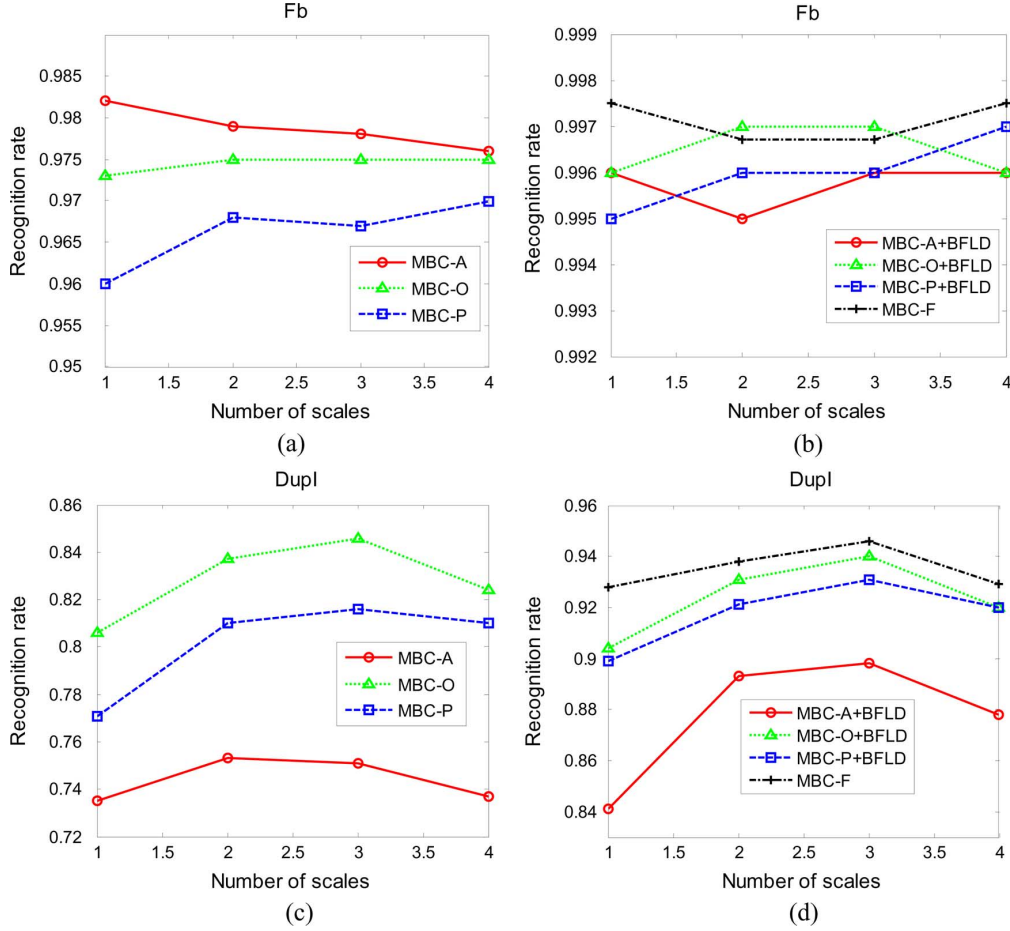
Fig. 6. Curves of FR rate versus the number of scales by the proposed MBC methods. (a) MBC on Fb; (b) MBC+BFLD and MBC-F on Fb; (c) MBC on DupI; (d) MBC+BFLD and MBC-F on DupI.
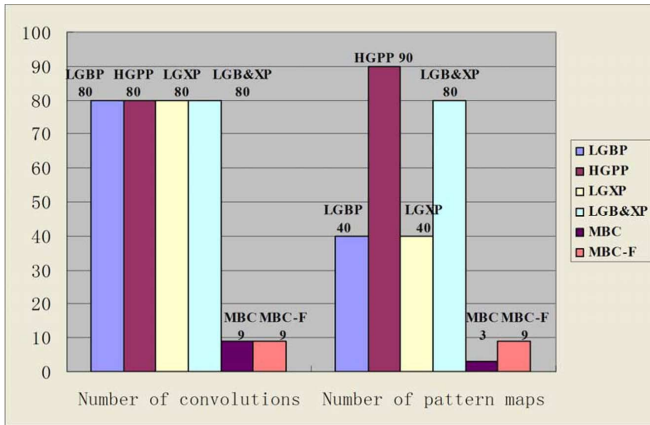

Fig. 7. Comparison of the time and space complexities among LGBP [9], HGPP [11], LGXP [22], LGB&XP [22] (fusing LGXP and LGBP), and the proposed MBC (MBC denotes any one of MBC-A, MBC-P, and MBC-O) and MBC-F.

[11], LGXP [22], and LGB&XP (fusing Gabor magnitude and phase parts in [22]).

All the methods of LGBP, HGPP, LGXP, and LGB&XP use Gabor transformations (with 5 scales and 8 orientations) to generate feature maps, which means that 80 convolutions (40 real convolutions and 40 imagery convolutions) are needed in processing a single image. However, for the process of feature gen-

eration in the proposed MBC and MBC-F, only 9 convolutions are executed (1 log-Gabor filtering and 2 Riesz transformations on each scale) [12], [23], [32], [33]. Fig. 7 clearly shows that the time complexity of MBC in feature map generation is only 1/9 times of that of [9]–[11], [22].

In the pattern map coding, both LGBP and LGXP encode 40 feature maps into 40 pattern maps, and HGPP produces 90 pattern maps, including 10 global phase pattern maps and 80 local pattern maps, with LGB&XP outputting 40 Gabor magnitude and 40 Gabor phase pattern maps. In comparison, our proposed MBC-A, MBC-P and MBC-O each only needs to encode 3 monogenic pattern maps. Even the fusing version of MBC, i.e., MBC-F, only has 9 pattern maps together. Therefore the proposed MBC has significantly lower time and space complexity than the previous Gabor-based ones [9]–[11], [22] in pattern map coding, as illustrated in Fig. 7.

Considering that the time and space complexity for histogram generating depends on the number of pattern maps, the proposed MBC methods will also need much less computational cost and space in histogram feature generation than [9]–[11], [22].

In similarity computing, when the original histogram features are used, according to the number of pattern maps, the time and space complexity of proposed MBC scheme is much less than that of LGBP, HGPP and LGXP. However, when BFLD [22] is used to reduce the dimensionality of histogram features, the dimensionality of projected discriminative features for all
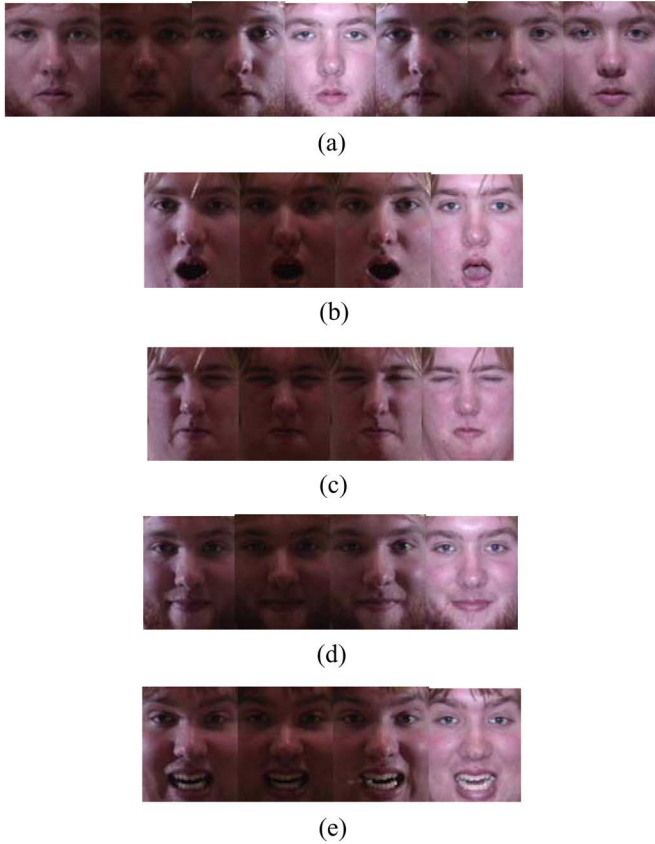
(a)

(b)

(c)

(d)

(e)

Fig. 8. Subject in Multi-PIE database. (a) Training samples with only illumination variations. (b) Testing samples with surprise expression and illuminations in Session 2. (c) Testing samples with squint expression and illuminations in Session 2. (d) and (e) Testing samples with smile expression and illumination variations in Sessions 1 and 3, respectively.

TABLE III
FACE RECOGNITION RATES (%) AND AVERAGE RUNNING TIME (SECOND) ON THE MULTI-PIE EXPRESSION DATABASE

Method	Smile-S1	Smile-S3	Surprise-S2	Squint-S2	Running Time
LBP	96.9	68.4	68.8	76.8	**0.313**
LGBP	**99.2**	77.7	72.0	**83.4**	12.702
HGPP	**99.2**	80.3	**75.2**	**83.4**	27.532
LGXP	**99.2**	77.7	72.4	**83.4**	12.236
MBC-A	**99.2**	**85.8**	72.9	**83.4**	1.016
MBC-O	99.0	80.6	72.9	**83.4**	1.072
MBC-P	99.1	78.2	73.2	81.7	1.025

(a)

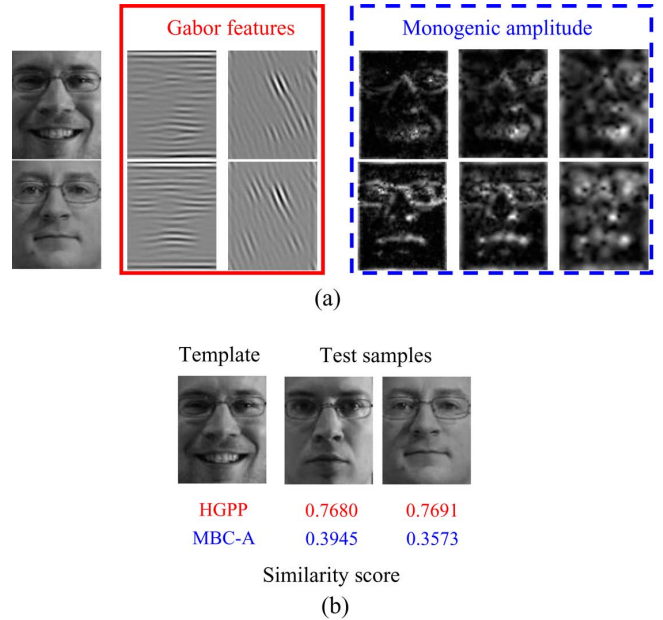|	Template	Test samples
HGPP		0.7680	0.7691
MBC-A		0.3945	0.3573

Similarity score

(b)

Fig. 9. (a) Examples of Gabor feature and monogenic amplitude feature maps. (b) Similarity scores by HGPP and MBC-A, where the left image is a template image, the middle image is a test image which is from the same class as the template image, and the right test image is from a different class.

methods, including MBC and LGBP [9], HGPP [11], LGXP [22], is nearly the same as each other. Therefore, when coupled with BFLD [22], all the methods would have similar time and space complexity in the phase of similarity computing. Note that BFLD will increase the computational complexity in offline learning stage, but it will not add computational burden in the online recognition stage. We will compare the empirical running time of the competing methods in Section V-C.

C. Experiments on the Multi-PIE Expression Database

In this section, we use the large-scale Multi-PIE [49] to verify the effectiveness and efficiency of MBC without considering dimensionality reduction. All the 249 subjects in Session 1 are used as the training set in this experiment. To make the FR more challenging, four subsets with both illumination and expression variations in Sessions 1, 2 and 3 are used for testing. For the training set, we use the 7 frontal images with extreme illuminations $\{0, 1, 7, 13, 14, 16, 18\}$ and neutral expression (refer to Fig. 8(a) for examples). For the testing set, 4 typical frontal images with illuminations $\{0, 2, 7, 13\}$ and different expressions (smile in Sessions 1 and 3, squint and surprise in Session 2) are used (refer to Fig. 8(b) for examples with surprise in Session 2, Fig. 8(c) for examples with squint in Session 2, Fig. 8(d) for examples with smile in Session 1, and Fig. 8(e) for examples with smile in Session 3). We crop and normalize the images to 100×82. Here the programming environment is Matlab version

R2011a. The desktop used is of 1.86 GHz CPU and with 2.99G RAM.

Here LBP [8] is used as the baseline method, and we compare the proposed MBC with the state-of-the-art methods, such as LGBP [9], HGPP [11] and LGXP [22]. Table III lists the face recognition rates and average running time of all the competing methods. It is clear to see that LBP is the fastest one, but with the worst recognition rate. The proposed MBC has very competitive recognition rates as LGBP, HGPP and LGXP. Especially for MBC-A, it has the highest recognition rates in the testing sets of "smile-S1", "smile-S3" and "squint-S2". The running time of MBC is about 1 second for each image, less than 1/10 times of that of LGBP, HGPP and LGXP. It is noted that HGPP has the longest running time. The running time results are very consistent with our analysis in Section V-B.

In this experiment, it can be seen that MBC-A outperforms LGBP and LGXP in all cases. Particularly, MBC-A outperforms much LGBP, HGPP and LGXP on Smile-S3. This is because that on the Smile-S3 dataset, the Gabor features from different subjects may not have enough discrimination. Fig. 9(a) shows some examples of Gabor feature and monogenic amplitude feature maps. One can see that the Gabor feature maps of the two images, which are from two different classes, are very similar, while the monogenic amplitude feature maps of them can have clear difference. Fig. 9(b) shows the similarity scores between

Fig. 10. Some samples of normalized face images of FERET.

TABLE IV
RECOGNITION RATES (%) BY USING LOCAL VARIATION CODE ONLY AND
BY USING THE COMBINATION OF LOCAL VARIATION CODE AND IMAGERY
INTENSITY CODE ON THE FB AND DUP1 PROBE SETS

	Method	Local variation code	Local variation & imagery intensity code
Fb	MBC-A	96.3	**98.1**
	MBC-A + BFLD	99.2	**99.4**
	MBC-P	95.1	**96.6**
	MBC-P + BFLD	99.0	**99.6**
	MBC-O	96.4	**97.6**
	MBC-O + BFLD	99.2	**99.6**
DupI	MBC-A	66.9	**74.4**
	MBC-A + BFLD	81.4	**88.8**
	MBC-P	79.5	**81.6**
	MBC-P + BFLD	89.2	**92.9**
	MBC-O	80.3	**85.2**
	MBC-O + BFLD	90.9	**93.6**

TABLE V
RECOGNITION RATES (%) OF THE PROPOSED MBC METHODS
ON THE FERET PROBE SETS

Method	FERET probe sets			
	Fb	Fc	DupI	DupII
LGBP (Baseline)	*96.0*	*96.0*	*74.0*	*70.0*
MBC-A	**98.1**	97.4	74.4	72.2
MBC-P	96.6	**99.0**	81.6	78.2
MBC-O	97.6	**99.0**	**85.2**	**83.3**
LGBP +BFLD (Baseline)	*99.0*	*99.0*	*91.0*	*91.0*
MBC-A+BFLD	99.4	99.0	88.8	86.3
MBC-P+BFLD	**99.6**	**99.5**	92.9	**91.0**
MBC-O+BFLD	**99.6**	**99.5**	93.6	90.2
MBC-F	**99.7**	**99.5**	93.6	**91.5**

TABLE VI
RECOGNITION RATES (%) OF DIFFERENT METHODS ON THE
FERET FACE DATABASE

Methods	FERET probe sets			
	Fb	Fc	Dup I	Dup II
Xie's method in [22]	99.0	99.0	**94.0**	**93.0**
Tan's method in [26]	98.0	98.0	90.0	85.0
Zou's method in [27]	99.5	99.5	85.0	79.5
HGPP [11]	97.5	99.5	79.5	77.8
LGBP [9]	98.0	97.0	74.0	71.0
LBP [8]	97.0	79.0	66.0	64.0
MBC-F	**99.7**	**99.5**	**93.6**	91.5

two test images and a template image by HGPP and MBC-A. We can see that MBC-A could lead to the correct classification, but Gabor feature based HGPP will lead to a misclassification. Note that the MBC-A similarity score values are usually much lower than those of HGPP because the histogram feature of MBC-A is sparser than that of HGPP. However, this does not reduce the discrimination capability of MBC-A.

D. Experiments on the FERET Database

The FERET database [14], [41] is often used to validate an algorithm's effectiveness because it contains many kinds of image variations. By taking Fa subset as a gallery, the probe subsets Fb and Fc are captured with expression and illumination variations. Especially, DupI and DupII consist of images taken at different times. For some people, the time intervals are more than two years between the gallery set and DupI or DupII set. The facial portion of each image is cropped based on the locations of eyes. The cropped face images are then normalized to 150×130 pixels. Fig. 10 shows some examples of normalized face images. Here the projection matrix of BFLD is learnt from the training set which has 1 002 frontal images from 429 subjects and we set the number of blocks as $M_b = 5$.

In order to illustrate that the combination with local imaginary intensity code could improve much the local variation code, in Table IV we report the recognition rates by using local variation code only and by using the combination of local variation code and local intensity code on the probe sets of Fb and DupI. It can be clearly seen that in all the tests, the latter one outperforms much the former one. Without BFLD, the combination with local intensity code could improve in average 4.6% the recognition rate for MBC-A, 1.8% for MBC-P, and 3% for MBC-O, respectively. With BFLD, the combination with

local intensity code could still improve in average 3.8% for MBC-A, 2.2% for MBC-P, and 1.5% for MBC-O, respectively. This shows that the local imaginary code could provide complementary information to the local variation code.

Table V lists the FR results by each individual MBC (e.g., MBC-A, MBC-P or MBC-O), their combinations with BFLD, and the fused MBC (MBC-F). Meanwhile, we give the results by LGBP and LGBP+BFLD reported in [22] as the baseline methods. From Table V, we can see that all the proposed monogenic codes, MBC-A, MBC-P or MBC-O, could achieve higher recognition rate than LGBP, especially with the biggest improvement of about 11% in DupI and DupII for MBC-O. BFLD could improve the performance of all methods. When combined with BFLD, MBC-A, MBC-P or MBC-O could achieve higher recognition rates than the method LGBP+BLFD in most cases. In this experiment, the weight w in MBC-F is set to 0.45 since the amplitude component performs a little worse than the monogenic phase part.

The FR results of MBC-F and other state-of-the-art LSF-FR methods, including the feature fusion version of LGXP [22], Tan's method in [26], Zou's method in [27], HGPP [11], LGBP [9] and LBP [8], are compared in Table VI. It can be seen that the proposed MBC-F is the best one for Fb and Fc. In DupI and DupII, the recognition rate of MBC-F is slightly lower than the method in [22]. However, it should be noted that our proposed MBC-F method has much lower time and space complexity than the method in [22]. Therefore, the overall performance of MBC-F is better.

E. Experiments on the FRGC Database

FRGC 2.0 [13] is a very large scale face database, which consists of 50 000 recordings divided into training and validation

Fig. 11. Some samples of normalized face images of FRGC 2.0.

TABLE VII
VERIFICATION RATES BY THE PROPOSED MBC METHODS ON FRGC 2.0
WHEN FAR = 0.1%

Method	Exp.1			Exp.4		
	ROC1	ROC2	ROC3	ROC1	ROC2	ROC3
LGBP (baseline)	*80.3*	*76.3*	*71.9*	*11.3*	*11.4*	*11.7*
MBC-A	83.8	80.4	76.7	11.0	11.5	12.1
MBC-P	**88.0**	**85.0**	**82.0**	16.2	16.3	16.8
MBC-O	87.7	85.0	81.8	**17.4**	**17.7**	**18.4**
LGBP+BFLD (baseline)	*98.1*	*97.5*	*96.8*	*79.6*	*80.3*	*81.0*
MBC-A+BFLD	97.6	96.9	96.2	79.3	79.3	80.5
MBC-P+BFLD	97.8	97.0	96.3	71.6	72.5	73.6
MBC-O+BFLD	98.0	97.4	**96.8**	71.8	72.8	74.0
MBC-F	**98.5**	**98.0**	**97.5**	**82.1**	**83.0**	**83.9**

partitions. The FRGC challenge problem consists of six experiments, which evaluate FR performance on various conditions, e.g., controlled or uncontrolled lighting and background, 3-D imagery, and multistill imagery [13]. There are three masks, which could be used to generate three receiver operator characteristic (ROC) curves. ROC1 could verify FR on the images collected within semesters, while ROC2 and ROC3 evaluate FR on the images collected within a year and between semesters, respectively.

In our experiments, we learn the projection matrix of BFLD based on the given training set (12 776 images of 222 persons) and implement our tests on the 1st and 4th experiments, denoted by Exp.1 and Exp.4. Exp.4 is much more difficult than Exp.1 since the query images in Exp.4 are taken under uncontrolled environment. All the face images are cropped based on the manually located eye centers provided by the original database and normalized to 168×128 pixels. Some samples are shown in Fig. 11. Here M_r is set as 4. In order for a stable and effective discriminative feature extraction by dimensionality reduction, BFLD usually requires that the input should have more features. For the feature of monogenic signal representation, the local amplitude (i.e., A_{lg} in (10)) is a compact composition of three partial features (i.e., f_{lg}, f_{lg-x}, and f_{lg-y} in (10)). In order to capture more information in amplitude, for MBC-A+BFLD in Exp4, instead of directly encoding on the amplitude, we first encode the local variations of f_{lg}, f_{lg-x}, and f_{lg-y}, and then in each block we concatenate the three kinds of histogram features. Finally we apply BFLD to the concatenated higher dimensional histogram feature. Although a higher dimensional histogram feature of monogenic amplitude is generated, the time and space complexity is only slightly increased, and the dimension of the final discriminative feature generated by BFLD does not increase.

Similar to Table V, we list the verification rates of different MBC based methods when the FAR is 0.1% in Table VII. It can be seen that MBC could achieve much better performance than LGBP in most cases. However, for Exp.4, all MBC and
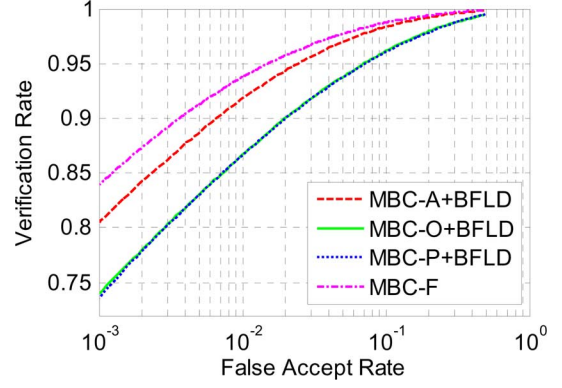

Fig. 12. ROC3 curves of different methods on Exp4 of FRGC 2.0 databases.

TABLE VIII
VERIFICATION RATES (%) BY DIFFERENT METHODS ON FRGC 2.0
FACE DATABASE

Methods	FRGC 2.0	
	Exp.1	Exp.4
Xie's method in [22]	97.5	84.9
Tan's method in [26]	N/A	83.6
Huang's method in [44]	91.5	74.3
Liu's method in [45]	92.0	76.0
Liu's method in [46]	N/A	81.3
Su's method in [15]	**98.0**	**86.0**
MBC-F	97.5	83.9

LGBP methods have low validation rates, not larger than 20%. When BLFD is used to extract the discriminative features, all the methods have great improvements, e.g., more than 60% for MBC-A and LGBP. We could also see that MBC+BLFD, especially MBC-A+BFLD, has very competing performance with LGBP but with much lower complexity. The ROC3 curves of MBC+BFLD and MBC-F are plotted in Fig. 12, which clearly show that MBC-F could improve the final performance by fusing the complementary monogenic information.

The recognition results of MBC-F and other state-of-the-art methods, including the fusion version of LGXP [22], Tan's method in [26], Huang's method in [44], Liu's method in [45], Liu's method in [46], Su's method in [15], are listed in Table VIII. In the table, all the results for the comparison are directly cited from the related papers. We can see that MBC-F achieves very close results to the best results reported in [15] and [22]. However, considering that both Xie's method in [22] and Su's method in [15] adopt multiscale and multiorientation Gabor transformation to generate features, MBC-F has much lower complexity than them, which further validates the effectiveness of the proposed techniques.

F. Experiments on the Polyu NIR Database

The PolyU-NIR face database [28] is a large scale near-infrared face database, consisting of 350 subjects, each subject providing about 100 samples. Various variations of face images, such as expression, pose, scale, focus, time, are involved in the capturing. In this paper, we do three tests following the experiment setting in [28]. These three subsets are Exp. 1 (419 training samples, 574 gallery samples and 2 762 probe samples), Exp. 2 (1 876 training samples, 1 159 gallery samples and 4 747 probe

Fig. 13. Some samples of normalized face images of Exp.3 on the PolyU NIR database.

TABLE IX
RECOGNITION RATES BY THE PROPOSED MBC METHODS ON THE POLYU-NIR DATABASE

Methods	PolyU-NIR probe sets		
	Exp.1	Exp.2	Exp.3
Gabor-DBC (baseline)	*97.6*	*92.3*	*92.5*
MBC-A	98.7	93.5	94.1
MBC-P	98.3	92.3	92.7
MBC-O	98.4	92.6	93.2
MBC-A + BFLD	98.6	94.0	94.8
MBC-P + BFLD	98.5	93.5	94.4
MBC-O + BFLD	98.7	93.6	94,4
MBC-F	**98.8**	**94.5**	**95.3**

TABLE X
RECOGNITION RATES (%) BY DIFFERENT METHODS ON THE POLYU-NIR PROBE SETS

Methods	PolyU-NIR		
	Exp.1	Exp.2	Exp.3
Gabor-PCA	82.4	90.2	60.8
Gabor-LDA	91.8	92.1	82.4
LGBP	95.1	89.3	89.8
Gabor-DBC	97.6	92.3	92.5
MBC-F	**98.8**	**94.5**	**95.3**

samples), and Exp. 3 (576 training samples, 951 gallery samples and 3 648 probe samples). Some samples of Exp. 3 are shown in Fig. 13. Because the face images are normalized to 64×64 pixels, which are much smaller than that of Multi-PIE, FERET and FRGC, we set $\lambda_{\min} = 0.55$, $M_b = 2$ and the dimensionality of discriminative feature as 100.

Table IX lists the recognition rates by MBC methods and the baseline method, Gabor-DBC, proposed in [28]. From Table IX, we can see that all the MBC methods are better than the baseline method; and BFLD could further improve MBC's performance. The recognition results of MBC-F and other state-of-the-art results reported in [28] are shown in Table X. It can be seen that the proposed MBC-F outperforms the second best method by 1.2% in Exp.1, 2.2% in Exp.2 and 2.8% in Exp.3. Meanwhile, compared to these Gabor feature based methods, the proposed MBC methods have much lower time and space complexity.

G. Discussions

From the above complexity analysis and experimental results, it can be seen that the proposed MBC has competitive accuracy with or higher accuracy than the Gabor feature based methods (e.g., LGBP [9], HGPP [11], and LGXP [22]), while it has much lower cost (e.g., less than 1/10 times running time on the MPIE database). This merit of MBC comes from two parts: the power of monogenic signal representation and the effectiveness of MBC's pattern coding strategy.

Like Gabor wavelet representation, monogenic signal representation can effectively extract the discrimination information embedded in the original signal by decoupling the local energy (local amplitude) and structure (local phase and local orientation). Different from Gabor features, which are extracted along several orientations by using steerable filters, in monogenic signal representation the local orientation is the main orientation of the linear structure with large support, and the local phase describes the structure along the main orientation [12]. This strategy of extracting only the information along the dominant orientation is somewhat similar to the maximum-like pooling mechanism in the visual processing in cortex [50], and it leads to a compact but effective representation of the image. Our experimental results validate that monogenic signal analysis is very suitable for face image representation, probably because human facial local structures can be well characterized along one dominant orientation. In addition, the adaptation of log-Gabor filters in multiscale monogenic signal representation makes it efficient in capturing the multiscale information.

The promising performance of MBC also comes from the proposed binary coding scheme, which is well suited to describe the monogenic signal representation. In MBC, we not only encode the local variation of monogenic amplitude and monogenic phase, but also encode the monogenic local imagery intensity to provide complementary information to the local variation. As shown in Table IV, this coding strategy could exploit more useful information of monogenic signal representation. In addition, the three parts of MBC: MBC-A, MBC-P and MBC-O, encode the energetic, structural and geometric information of a 2-D signal, respectively, which are complementary to each other. The experimental results show that MBC on monogenic amplitude, orientation and phase could individually achieve high recognition rate on different datasets. More importantly, a strategy by using a weighted average to fuse these three similarity scores (e.g., MBC-A, MBC-P and MBC-O) could further improve the recognition accuracy and achieve state-of-the-art FR performance.

VI. CONCLUSION

We proposed a novel face representation model, namely monogenic binary coding (MBC), based on the monogenic signal representation. One of the best merits of MBC is that it has much less time and space complexity than the widely used Gabor transformation based local feature extraction method. Through multiscale monogenic signal representation, three kinds of features (e.g., local amplitude, local orientation and local phase) can be generated, and then encoded by the proposed monogenic local variation coding and monogenic imagery intensity coding procedures. The produced MBC pattern maps are used to compute the statistical features (e.g., histogram), which are then used to measure the similarity for face recognition. The extensive experiments on the benchmark face databases, including FERET, FRGC 2.0, Multi-PIE and PolyU-NIR, clearly showed that the proposed MBC methods not only have significantly lower time and space complexity than the state-of-the-art Gabor feature based face recognition methods, but also have very competitive or even better recognition rates.

REFERENCES

[1] W. Y. Zhao, R. Chellppa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Survey*, vol. 35, no. 4, pp. 399–458, 2003.

[2] N. Poh, C. H. Chan, J. Kittler, S. Marcel, C. M. Cool, E. A. Rúa, J. L. A. Castro, M. Villegas, R. Paredes, V. Struc, N. Pavesić, A. A. Salah, H. Fang, and N. Costen, "An evaluation of video-to-video face verification," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 4, pp. 781–801, Dec. 2010.

[3] Z. F. Li, U. Park, and A. K. Jain, "A discriminative model for age invariant face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 1028–1037, Sep. 2011.

[4] H. B. Ling, S. Soatto, N. Ramanathan, and D. W. Jacobs, "Face verification across age progression using discriminative methods," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 1, pp. 82–91, Mar. 2010.

[5] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neurosci.*, vol. 13, no. 1, pp. 71–86, 1991.

[6] P. Belhumeur, J. Hespanha, and D. Kriegman, "Egienfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[7] 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.

[8] A. Timo, H. Abdenour, and P. Matti, "Face recognition with Local binary patterns," in *Proc. Eur. Conf. Computer Vision*, Prague, Czech Republic, 2004.

[9] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," in *Proc. IEEE Int. Conf. Computer Vision*, Beijing, China, 2005.

[10] W. Zhang, S. Shan, L. Qing, X. Chen, and W. Gao, "Are Gabor phases really useless for face recognition?," *Pattern Anal. Appl.*, vol. 12, no. 3, pp. 301–307, 2009.

[11] 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. 2006.

[12] M. Felsberg and G. Sommer, "The monogenic signal," *IEEE Trans. Signal Process.*, vol. 49, no. 12, pp. 3136–3144, Dec. 2001.

[13] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recogniton grand challenge," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, San Diego, CA, 2005.

[14] 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.

[15] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical ensemble of global and local classifiers for face recognition," in *Proc. Int. Conf. Computer Vision*, Rio de Janeiro, Brazil, 2007.

[16] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, 1996.

[17] J. Yang and C. J. Liu, "Horizontal and vertical 2DPCA-based discriminant analysis for face verification on a large-scale database," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 4, pp. 781–792, Dec. 2007.

[18] S. Venkatesh and R. Owens, "On the classification of image features," *Pattern Recognit. Lett.*, vol. 11, pp. 339–349, 1990.

[19] E. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton, NJ: Princeton Univ. Press, 1971.

[20] D. Fields, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer.*, vol. 4, no. 12, pp. 2379–2394, 1987.

[21] G. H. Granlund and H. Knutsson, *Signal Processing for Computer Vision*. Norwell, MA: Kluwer, 1995.

[22] S. F. Xie, S. G. Shan, X. L. Chen, and J. Chen, "Fusing local patterns of gabor magnitude and phase for face recognition," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1349–1361, May 2010.

[23] M. Yang, L. Zhang, L. Zhang, and D. Zhang, "Monogenic binary pattern (MBP): A novel feature extraction and representation model for face recognition," in *Proc. Int. Conf. Pattern Recognition*, Istanbul, Turkey, 2010.

[24] H. K. Ekenel, H. Gao, and R. Stiefelhagen, "3-D face recognition using local appearance-based models," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 630–636, Sep. 2007.

[25] Z. H. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1657–1663, Jun. 2010.

[26] X. Tan and B. Triggs, "Fusing gabor and LBP feature sets for kernel-based face recognition," in *Proc. IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures*, Rio de Janeiro, Brazil, 2007.

[27] J. Zou, Q. Ji, and G. Nagy, "A comparative study of local matching approach for face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 10, pp. 2617–2628, Oct. 2007.

[28] B. C. Zhang, L. Zhang, D. Zhang, and L. L. Shen, "Directional binary code with application to PolyU near-infrared face database," *Pattern Recognit. Lett.*, vol. 31, no. 14, pp. 2337–2344, 2010.

[29] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[30] M. Yang and L. Zhang, "Gabor feature based sparse representation for face recognition with gabor occlusion dictionary," in *Proc. Eur. Conf. Computer Vision*, Crete, Greece, 2010.

[31] M. Felsberg and G. Sommer, "The monogenic scale-space: A unifying approach to phase-based image processing in scale-space," *J. Math. Imag. Vis.*, vol. 21, no. 1, pp. 5–26, 2004.

[32] M. Unser, D. Sage, and D. V. D. Ville, "Multiresolution monogenic signal analysis using the riesz-laplace wavelet transform," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2402–2418, Nov. 2009.

[33] L. Zhang, L. Zhang, Z. H. Guo, and D. Zhang, "Monogenic-LBP: A new approach for rotation invariant texture classification," in *Proc. Int. Conf. Pattern Recognition*, Istanbul, Turkey, 2010.

[34] S. V. Fischer, F. Sroubek, L. Perrinet, R. Redondo, and G. Cristobal, "Self-invertible 2D log-gabor wavelet," *Int. J. Comput. Vis.*, vol. 75, no. 2, pp. 231–246, 2007.

[35] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Robust sparse coding for face recognition," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Colorado Springs, CO, 2011.

[36] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010.

[37] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, Mar. 2005.

[38] T. Inan and U. Halici, "3D face recognition with local shape descriptors," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 577–587, Jun. 2012.

[39] Z. Zhou, A. Wagner, H. Mobahi, J. Wright, and Y. Ma, "Face recognition with contiguous occlusion using markov random fields," in *Proc. IEEE Int. Conf. Computer Vision*, Kyoto, Japan, 2009.

[40] E. Elhamifar and R. Vidal, "Robust classification suing structured sparse representation," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Colorado Springs, CO, 2011.

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

[42] J. Cook, V. Chandran, S. Sridharan, and C. Fookes, "Gabor filter bank representation for 3D face recognition," in *Proc. Digital Imaging Computing: Techniques and Applications*, Cairns, Australia, 2005.

[43] P. Kovesi, "Invariant Measures of Image Features From Phase Information," Ph.D. Thesis, Univ. of Western Australia, Perth, Australia, 1996.

[44] W. Huang, G. Park, and J. Lee, "Multiple face model of hybrid fourier feature for large face image set," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, New York, 2006.

[45] 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.

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

[47] J. Cook, C. McCool, V. Chandran, and S. Sridharan, "Combined 2D/3D face recognition using log-gabor templates," in *Proc. IEEE Int. Conf. Video and Signal Based Surveillance*, Sydney, Australia, 2006.

[48] J. Cook, V. Chandran, and C. Fookes, "3D face recognition using log-gabor templates," in *Proc. British Machine Vision Conf.*, Edinburgh, Scotland, U.K., 2006.

[49] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image Vis. Comput.*, vol. 28, pp. 807–813, 2010.

[50] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," *Nature Neurosci.*, vol. 2, no. 11, pp. 1019–1025, 1999.

Meng Yang (S'11) received the B.Sc. degree in information engineering and the M.Sc. degree in control theory and applications from the Northwestern Polytechnical University, Xi'an, China, in 2006 and 2009, respectively. He received the Ph.D. degree in computer science at The Hong Kong Polytechnic University, Hong Kong, in 2012.

His research interests include computer vision, pattern recognition, sparse representation, and face recognition.
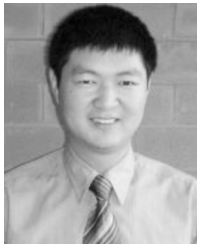
Lei Zhang (M'04) received the B.S. degree in 1995 from Shenyang Institute of Aeronautical Engineering, Shenyang, China, and the M.S. and Ph.D. degrees in automatic control theory and engineering from Northwestern Polytechnical University, Xi'an, China, in 1998 and 2001, respectively.

From 2001 to 2002, he was a research associate in the Department of Computing, The Hong Kong Polytechnic University. From January 2003 to January 2006, he worked as a Postdoctoral Fellow in the Department of Electrical and Computer Engineering, McMaster University, Canada. In 2006, he joined the Department of Computing, The Hong Kong Polytechnic University, as an Assistant Professor. Since September 2010, he has been an Associate Professor in the same department. His research interests include image and video processing, biometrics, computer vision, pattern recognition, multisensor data fusion and optimal estimation theory, etc.

Dr. Zhang is an Associate Editor of IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and *Image and Vision Computing Journal*. Dr. Zhang was awarded the Faculty Merit Award in Research and Scholarly Activities in 2010 and 2012, and the Best Paper Award of SPIE VCIP2010.

Simon Chi-Keung Shiu (M'90) received the M.Sc. degree in computing science from University of Newcastle Upon Tyne, U.K., in 1985, the M.Sc. degree in business systems analysis and design from City University, London, in 1986, and the Ph.D. degree in computing in 1997 from The Hong Kong Polytechnic University.

He is an assistant professor in the Department of Computing, Hong Kong Polytechnic University, Hong Kong. He worked as a system analyst and project manager between 1985 and 1990 in several business organizations in Hong Kong. His current research interests include case-base reasoning, machine learning, and soft computing. He has coauthored the book *Foundations of Soft Case-Based Reasoning*, published by John Wiley and Sons, and co-guest edited a special issue on soft case based reasoning of the *Applied Intelligence Journal*. He is a member of the British Computer Society.

David Zhang (F'09) graduated in computer science from Peking University in 1974 and received the M.Sc. and Ph.D. degrees in computer science and engineering from the Harbin Institute of Technology (HIT), Harbin, China, in 1983 and 1985, respectively. He received the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, Canada, in 1994.

From 1986 to 1988, he was a Postdoctoral Fellow at Tsinghua University, Beijing, China, and became an Associate Professor at Academia Sinica, Beijing, China. Currently, he is a Professor with the Hong Kong Polytechnic University, Hong Kong. He is Founder and Director of Biometrics Research Centers supported by the Government of the Hong Kong SAR (UGC/CRC). He is also Founder and Editor-in-Chief of the *International Journal of Image and Graphics* (IJIG), Book Editor, *The Kluwer International Series on Biometrics*, and an Associate Editor of several international journals. His research interests include automated biometrics-based authentication, pattern recognition, biometric technology and systems. As a principal investigator, he has finished many biometrics projects since 1980. So far, he has published over 200 papers and 10 books.