Neurocomputing 355 (2019) 1–12

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Local polynomial contrast binary patterns for face recognition

Zhen Xu[a,b], Yinyan Jiang[a,b], Yichuan Wang[a,b], Yicong Zhou[c], Weifeng Li[a,b,*], Qingmin Liao[a,b]

[a] Department of Electronic Engineering/Graduate School at Shenzhen, Tsinghua University, China
[b] Shenzhen Key Laboratory of Information Science and Technology, Shenzhen, China
[c] Department of Computer and Information Science, University of Macau, Macau, China

ARTICLE INFO

Article history:
Received 21 August 2017
Revised 1 May 2018
Accepted 24 September 2018
Available online 27 September 2018

Communicated by Dr Xiaoming Liu

Keywords:
Face recognition
Polynomial filters
Local binary patterns
Surface fitting

ABSTRACT

We propose a novel face representation model, called the polynomial contrast binary patterns (PCBP), based on the polynomial filters, for robust face recognition. It is assumed that the discrete array of pixel values comes about by sampling an underlying smooth surface in an image. The proposed method efficiently estimates the underlying local surface information, which is approximately represented as linear projection coefficients of the pixels in a local patch. The decomposition using polynomial filters can capture rich image information at multiple orientations and frequency bands. This guarantees its robustness to illumination and expression variations. The weighting scheme embeds different discriminative powers of each filter response image. We also propose to carry out a subsequent Fisher linear Discriminant (FLD) on each decomposed image for dimension reduction of features. Our extensive experiments on the public FERET and LFW databases demonstrate that the non-weighted Polynomial contrast binary patterns performs better than most of methods and the weighting scheme further improves the recognition rates. WPCBP+FLD(CD) and WPCBP+FLD(HI) can achieve much competitive or even better recognition performance compared with the state-of-the-art face recognition methods.

1. Introduction

Automatic face recognition has been extensively studied over the past two decades due to the broad prospect in real-world applications such as surveillance, biometrics, human computer interaction. Face representation is the most key stage in a face recognition system, because face images captured in real world environments are largely affected by intra-personal variations, such as expression, illumination, pose, noise, occlusion, aging and so on. An effective face representation should be able to extract discriminative features to make face images more separable. However, the large variations occurring in face images reduce the intra-personal similarity and increase the inter-personal similarity, which make face representation phase one of the key challenges for excellent face recognition performance.

A variety of face representation methods focus on seeking effective and efficient feature to enhance face recognition performance [1,2]. Basically, feature extraction methods can be mainly divided into two categories: subspace based holistic features and local appearance features. Among the holistic features, principle component analysis (PCA) [3] and Fisher linear discriminant (FLD) [4] are the two most well-known linear subspace learning methods in the field of pattern recognition tasks. PCA provides an optimal linear transformation in sense of the least mean square reconstruction error. FLD attempts to seek a linear transformation by maximizing the ratio of the variance between the classes to the variance within the class. Various manifold learning methods such as ISOMAP [5], LLE [6], LPP [7], NPE [8] etc, can also be cast into this category. Furthermore, the kernel methods [9] are introduced to handle the nonlinear structure of data. Holistic features performs well under controlled environments but they share some common shortcomings. For instance, they usually need a representative training set, and the recognition performance may degrade due to the partial variations occurring under uncontrolled environments.

In recent years, using deep neural networks to learn effective feature representations has become popular in face recognition. Through training on large-scale face images, DeepID [10] and DeepFace [11] could learn discriminative deep face representation. Then DeepID2 [12] and DeepID2+ [13] as improved methods of DeepID, were proposed and effectively promoted the accuracy on face recognition. Although the methods based on deep leaning have nearly achieved the best performance, the huge amount of training data restricts their application in many practical cases.

As opposed to holistic features, local appearance features provide powerful robustness to partial variations. Local binary patterns (LBP) [14] and Gabor wavelets [15] are two representative

* Corresponding author at: Department of Electronic Engineering/Graduate School at Shenzhen, University, Tsinghua, China.
E-mail address: li.weifeng@sz.tsinghua.edu.cn (W. Li).

https://doi.org/10.1016/j.neucom.2018.09.056
0925-2312/

methods in face recognition applications. It has been proven that LBP is a discriminant and effective feature for face recognition. LBP can capture the local structure information via the difference between the neighbors and central pixel. The binary encoding scheme of LBP ensures its invariant to any monotonic changes of intensity values in a local region. The spatial concatenating of local statistical histograms provides more robustness to partial variations of local features. In the last few years, the face representation based on binary encoding and spatial histogram arouses increasing interest and inspires rich researches into LBP variants. Three-Patch LBP (TPLBP) and Four-Patch LBP (FPLBP) [16] are proposed to capture the local patch similarities instead of pixel similarities. Local ternary pattern (LTP) [17] adopts a ternary instead of binary encoding rule to alleviate the sensitivity to the near-uniform image regions.

Gabor wavelets resemble the two-dimensional receptive fields of the mammalian cortical smile cells, which enable Gabor wavelets to capture the local structure corresponding to specific spatial frequency (scale), spatial locality and selective orientation. Gabor feature and its variants have been demonstrated to be robust to noise, illumination and expression changes. Recently, some researchers attempted to apply local features on stable pixel attributes (e.g, Gabor magnitude responses and phase responses) rather than the pixel intensity to obtain more sufficient and stable representations. There are lots of work focusing on the combinations between Gabor attributes and LBP. Zhang et al. [18] proposed the local Gabor binary patterns (LGBP) by applying LBP on Gabor magnitude responses. Zhang et al. [19] developed the local Gabor XOR patterns (LGXP) by performing local XOR patterns (LXP), which is a variant of LBP, on Gabor phase responses. Xie et al. [20] presented a face representation by fusing LGBP and LGXP, which has shown impressive performance. Chai et al. [21] takes advantage of different kinds of ordinal measures on magnitude, phase, real, and imaginary components of Gabor images. Besides Gabor wavelets, researchers attempted to explore more efficient and effective attributes. Vu et al. [22] proposed to extend the pixel-based self-similarity in LBP to patch-based self-similarity, in which the intensity values of a central pixel is replaced by the accumulated gradient magnitudes across different directions. Recently, there have been some works on learning discriminant attributes rather than pre-defined ones. Lei et al. [23] proposed to learn a set of image filters using the LDA criterion. Lu et al. [24] also attempted to learn image filters by maximizing the variance of all binary codes.

Taken together, it is vital to seek stable and powerful attributes which can capture well local structure information of a face image for face representation. In [25], Robert M. Haralick proposed the facet model which exploits surface fitting concept [26] and applied for step edge detection. The facet model assumes that there exists an underlying gray tone intensity surface and the digital image should be regarded as an observed noisy sampling of the surface. In some sense, numeric digital image operations should be explained in terms of their actions on the underlying surface. Therefore, there must then involve fitting the underlying surface corresponding to the sampled data before operating.

In this paper, we attempt to represent the underlying surface as linear combinations of the two-dimensional discrete Chebyshev polynomials and thus to estimate the surface fitting coefficients by minimizing the reconstruction error between the underlying surface and the observed image. The surface fitting coefficients can reflect the local surface structure information, and meanwhile provide stable and powerful attributes for further patterns extraction. Therefore, we propose a novel feature extraction method called polynomial contrast binary patterns (PCBP) based on a surface fitting concept via polynomial approximation. PCBP first captures the local structure information corresponding to different orientations and spatial frequencies by the linear surface fitting coefficients, which are used to represent pixel values with a set of 2-D Chebyshev polynomials. Here, the linear representation coefficients at a pixel are derived by convolving the local patch centered at the pixel with a series of filters, which proves to be just an approximate normalization of an evaluation of the 2-D polynomials. We name the derived Chebyshev polynomial filters (CPF), which can decompose an image into a series of responses with different and compact structure information. Then we extract local pattern which captures background intensity contrast information on all polynomial filter responses, and the spatial histogram model is also used to form a joint face representation. We further propose Weighted PCBP (WPCBP) to enhance the discriminative power of PCBP by assigning different polynomial filter responses with unequal weights, which are learned based on Fisher separation criterion. Finally, we also exploit the Fisher linear discriminant analysis for discriminative feature selection.

The rest of our paper is organized as follows: Section 2 introduces in detail the polynomial filters. Section 3 introduces the feature extraction algorithm of the polynomial contrast binary patterns (PCBP). Section 4 proposes the scheme of the PCBP based face recognition. Section 5 presents and discusses our experimental results and finally Section 6 concludes the paper.

2. Polynomial filters

Based on facet model, we attempt to exploit the underlying surface structure information as the attributes for local patterns extraction. As for discrete image analysis, the underlying surface is a real-valued function s defined on the domain of the image, which is a bounded and connected subset of the real plane $\mathbb{R}^2$. As we know, it is not possible to recover the true function s directly from the observed noisy sample. However, it can be assumed that the underlying function s takes some kind of parametric form, and these parameters are kind of representations of the underlying function s, i.e, smooth surface. To fit the function s, one must first assume some parametric form of s, and then exploit the intensity values of the sampled image to estimate the parameters, which reflect the local surface structure information of the image. Therefore, we can consider these parameters as more stable and powerful attributes, and then perform some operations based on the estimated values of the parameters.

Assuming that in each neighborhood of the image the function s takes the parametric form of a polynomial in the row and column coordinates, then we can have

$$s(x, y) = k_1 + k_2x + k_3y + k_4x^2 + k_5xy + k_6y^2 + \cdots, \quad (1)$$

which shows that the underlying surface function is easy to represent as linear combinations of the simple polynomials $\{1, x, y, x^2, xy, y^2, \ldots\}$. These polynomials are not orthogonal to each other, and may lead to redundancy. In this paper, we exploit the discrete two-dimensional chebyshev polynomials as the basis set as in [25]. The Chebyshev polynomials are unique and orthogonal to each other, thus providing compactness for surface fitting. Specifically, the underlying surface function s can be the linear combination of the discrete two-dimensional chebyshev polynomials, and the linear combination coefficients would represent the local surface structure information of the image. Therefore, we should first construct the discrete two-dimensional chebyshev polynomials and then compute the linear combination coefficients via the image and polynomials.

2.1. Discrete orthogonal polynomial construction

To construct the 2-dimensional discrete orthogonal polynomials for image processing, we first construct the 1-dimensional discrete

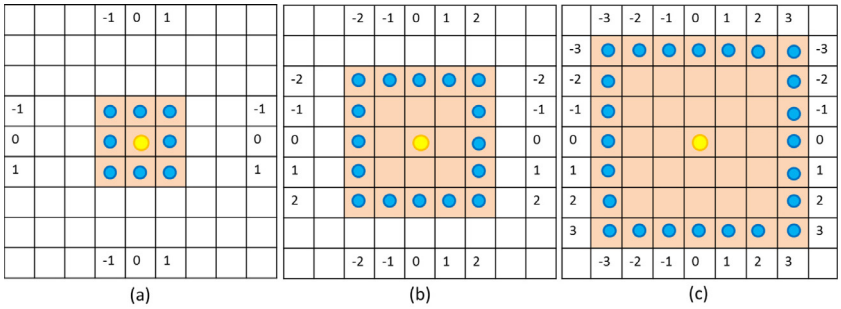

Fig. 1. Two-dimensional squarely symmetric neighborhood index sets for different size. (a) 3×3 patch with $\{-1,0,1\}\times\{-1,0,1\}$; (a) 5×5 patch with $\{-2,-1,0,1,2\}\times\{-2,-1,0,1,2\}$; (a) 7×7 patch with $\{-3,-2,-1,0,1,2,3\}\times\{-3,-2,-1,0,1,2,3\}$.

orthogonal polynomials $\{P_i(x)\}$, where $P_i(x)$ denotes the ith order 1-dimensional polynomial. Let the discrete index set Ω be symmetric around the zero (e.g, $\{-2,-1,0,1,2\}$). First define $P_0(x)=1, x\in\Omega$. Then $P_i(x)$ is assumed to take the form

$$P_i(x)=a_0+a_1x+\cdots+a_{i-1}x^{i-1}+x^i. \tag{2}$$

Suppose that the first n polynomials $P_0(x),\ldots,P_{n-1}(x)$ have been defined, then $P_n(x)$ must be orthogonal to $P_0(x),\ldots,P_{n-1}(x)$. Therefore, we will have n linear equations

$$\sum_{x\in\Omega}P_i(x)P_n(x)=0,\qquad i=0,\ldots,n-1. \tag{3}$$

Substituting Eq. (2) into Eq. (3), we will have

$$\begin{bmatrix}\mu_0 & \mu_1 & \mu_2 & \cdots & \mu_n\\ \mu_1 & \mu_2 & \mu_3 & \cdots & \mu_{n+1}\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ \mu_{n-2} & \mu_{n-1} & \mu_n & \cdots & \mu_{2*n-2}\\ \mu_{n-1} & \mu_n & \mu_{n+1} & \cdots & \mu_{2*n-1}\end{bmatrix}\begin{bmatrix}a_0\\ a_1\\ \vdots\\ a_{n-2}\\ a_{n-1}\\ 1\end{bmatrix}=0, \tag{4}$$

where

$$\mu_n=\sum_{x\in\Omega}x^n, \tag{5}$$

is the summation of the nth order index set. Therefore, by computing all the μ_n and solving Eq. (4), we will obtain the desired representation coefficients $a_0,a_1,\ldots,a_{n-1}$ for the nth order Chebyshev polynomial $P_n(x)$. Finally, we sequentially construct the 1-D polynomials $\{P_i(x)\}$.

Consider a 2-D patch corresponding to a 2-D index set $\Omega_x\times\Omega_y$, where Ω_x and Ω_y respectively denote 1-D index set along x and y axis, and the operator $\times$ is tensor product. Fig. 1 shows several examples about two-dimensional index set corresponding to the local neighborhood. By solving

$$\begin{aligned}&\sum_{x\in\Omega_x}P_i(x)P_n(x)=0,\qquad i=0,\ldots,n-1,\\ &\sum_{y\in\Omega_y}Q_j(y)Q_m(y)=0,\qquad j=0,\ldots,m-1,\end{aligned} \tag{6}$$

we will sequentially obtain the 1-D Chebyshev polynomials $\{P_0(x),\ldots,P_n(x),\ldots,P_{N-1}(x)\}$ and $\{Q_0(y),\ldots,Q_m(y),\ldots,Q_{M-1}(y)\}$ on Ω_x and Ω_y respectively and the number of elements of Ω_x and Ω_y are N and M respectively. Then the 2-D orthogonal polynomials can be derived from two sets of 1-D polynomials by taking tensor products. Therefore, the polynomial set

$$\begin{aligned}T_{n,m}(x,y)=P_n(x)Q_m(y),\qquad &n=0,\ldots,N-1,\\ &m=0,\ldots,M-1,\end{aligned} \tag{7}$$

is the desired 2-D discrete orthogonal polynomials on $\Omega_x\times\Omega_y$. The number of the 2-D polynomials is $N\times M$, which is the product of the number of the two corresponding 1-D polynomials. It is easy to prove the orthogonality of the 2-D polynomials when the 1-D polynomials are orthogonal to each other. According to the orthogonality of $\{P_n(x)\}$ and $\{Q_m(y)\}$, we will have

$$\begin{aligned}&\sum_{x\in\Omega_x}P_i(x)P_n(x)=\delta_{i,n},\\ &\sum_{y\in\Omega_y}Q_j(y)Q_m(y)=\delta_{j,m},\end{aligned} \tag{8}$$

where

$$\delta_{i,n}=\begin{cases}const.,\ i=n,\\ 0,\qquad\quad i\neq n,\end{cases} \tag{9}$$

is the Kronecker function. Considering the orthogonality of $T_{i,j}(x,y)$ and $T_{n,m}(x,y)$, we can have

$$\begin{aligned}&\sum_{x\in\Omega_x}\sum_{y\in\Omega_y}T_{i,j}(x,y)T_{n,m}(x,y)\\ &=\sum_{x\in\Omega_x}\sum_{y\in\Omega_y}P_i(x)Q_j(y)P_n(x)Q_m(y)\\ &=\sum_{x\in\Omega_x}P_i(x)P_n(x)\sum_{y\in\Omega_y}Q_j(y)Q_m(y)\\ &=\delta_{i,n}\delta_{j,m},\end{aligned} \tag{10}$$

where at least one of the right two terms will equal to 0 when $n\neq i$ or $m\neq j$, and thus the constructed 2-D polynomials are orthogonal to each other.

With regard to a 3×3 patch, the index set Ω is $\{-1,0,1\}\times\{-1,0,1\}$, and the 2-D discrete polynomial set $T_{n,m}(x,y)$ is as follows

$$\begin{aligned}&T_{0,0}=1, T_{1,0}=x, T_{0,1}=y,\\ &T_{2,0}=x^2-2/3, T_{1,1}=xy, T_{0,2}=y^2-2/3,\\ &T_{2,1}=\left(x^2-2/3\right)y, T_{2,1}=\left(y^2-2/3\right)x,\\ &T_{2,2}=\left(x^2-2/3\right)\left(y^2-2/3\right).\end{aligned} \tag{11}$$

2.2. Polynomial filters construction

After obtaining the 2-D polynomials, we will exploit them as the basis set for representing the underlying surface of an image. Denote the 2-D index set as Ω, and then for each $(x,y)\in\Omega$, we can rewrite Eq. (1) into the form

$$s(x,y)=\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}c_{n,m}T_{n,m}(x,y), \tag{12}$$

where the coefficients $c_{n,m}, (n=0,\ldots,N-1, m=0,\ldots,M-1)$ represent the surface structure information to be recovered in a local neighborhood. Let the observed data be $f(x,y)$, which is the

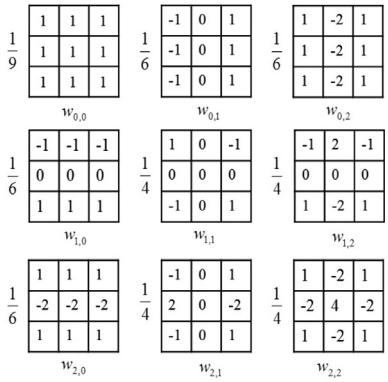

Fig. 2. Examples of 9 polynomial filters (weights) for local neighborhood of size 3×3.

pixel intensity value in the position (x, y) of the local neighborhood. To estimate the coefficients, we may minimize the total reconstruction error in the local neighborhood

$$\epsilon^2 = \sum_{(x,y)\in\Omega} [f(x,y) - s(x,y)]^2$$
$$= \sum_{(x,y)\in\Omega} \left[f(x,y) - \sum_{n=0}^{N-1}\sum_{m=0}^{M-1} c_{n,m} T_{n,m}(x,y) \right]^2. \quad (13)$$

Setting the derivatives of the above objective with respect to $c_{n,m}$ to zero, we obtain for each index $(x, y) \in \Omega$

$$c_{n,m} = \sum_{(x,y)\in\Omega} w_{n,m}(x,y) f(x,y) = \mathbf{w}_{n,m}^T \mathbf{f}, \quad (14)$$

where

$$w_{n,m}(x,y) = \frac{T_{n,m}(x,y)}{\sum_{(x,y)\in\Omega} T_{n,m}^2(x,y)} \quad (15)$$

is just an approximate normalization of an evaluation of the polynomial $T_{n,m}$ at the index (x, y), $\mathbf{w}_{n,m} \in \mathbb{R}^{NM\times 1}$ and $\mathbf{f} \in \mathbb{R}^{NM\times 1}$ are the vector notation of the weights $\{w_{n,m}(x,y), (x,y) \in \Omega\}$ and patch $\{f(x, y), (x, y) \in \Omega\}$ respectively. Fig. 2 shows the weights $\{w_{n,m}(x,y), (x,y) \in \Omega\}$ corresponding to Eq. (11) for a 3×3 patch. Eq. (14) implies the fitting coefficient $c_{n,m}$ can be computed as the linear combination of the pixel intensity values, and in vector form, $c_{n,m}$ can also be considered as the inner product of patch vector $\mathbf{f}$ and weights vector $\mathbf{w_{n,m}}$. If we consider Eq. (14) from an image filtering perspective, the surface fitting coefficient $c_{n,m}$ can be computed as the filter response of the mask vector $\mathbf{w}_{n,m}$ filtering over the image patch $\mathbf{f}$. By rearranging the vector $\mathbf{w}_{n,m}$ into matrix from, we will obtain the desired filters or masks. We name the weights $\mathbf{w}_{n,m}$ in vector or matrix form as Chebyshev polynomial filter (CPF), from which we could capture the underlying surface structure information. Regarding to a patch of $N \times M$ pixels, we can obtain $N \times M$ filters $\mathbf{w}_{n,m}$ for each polynomial basis $T_{n,m}$. In this paper, we only consider the symmetric square neighborhood, i.e, we set $N = M$. The construction of polynomial filters is summarized in Algorithm 1.

Fig. 3 shows the polynomial filters with different local neighborhood of size 3×3, 5×5, 7×7. The filters shown in Fig. 3 are arranged in square form, i.e, each row and column have N filters of $N \times N$ size. As can be seen, the most left-upper filter in each polynomial set is a mean filter, and the rest are gradient operator filters which are zero-mean due to the symmetric index set. It can be observed that our polynomial filters can compute not only lower-order derivatives but also higher-order derivative. Moreover, our polynomial filters decompose an image into multiple frequencies from coarse to fine. The lower frequency components are mainly

Algorithm 1 The construction of the Chebyshev Polynomial filters (CPF)

Input: The size of the local symmetrical neighborhood Ω, $N \times N$; The index set $(x, y) \in \Omega$;
Output: N^2 polynomial filters of size $N \times N$, $\mathbf{w}_{n,m}, n, m = 0, 1, \cdots, N-1$;
1: Define $P_0(x) = 1, Q_0(y) = 1$;
2: Construct the 1-D Chebyshev polynomials along x and y axis:
3: **for** each $P_n(x), Q_m(y)$, $n, m \in [1, N-1]$ **do**
4: 1) Compute a series of μ_i^p and μ_j^q according to Eq. (5), for $i \in [0, 2n-1]$ and $j \in [0, 2m-1]$;
5: 2) Compute the coefficients of $P_n(x), Q_m(y)$ according to Eq. (4);
6: 3) Obtain the $P_n(x), Q_m(y)$ by substituting the coefficients in 2) into Eq. (2);
7: **end for**
8: Compute the 2-D orthogonal polynomials $T_{n,m}(x, y)$ using 1-D polynomials $P_n(x)$ and $Q_m(y)$ according to Eq. (7);
9: Calculate all the N^2 polynomial filters $\mathbf{w_{n,m}}$ using $T_{n,m}(x, y)$ according to Eq. (15).

captured by the left and upper filters, and most of the higher frequency parts are reflected by the right and bottom filters. From the view of the arrangement of filters, it is obvious that polynomial filters are able to capture horizontal and vertical directions, as well as the combinations of horizontal and vertical directions. The filters lie in diagonal line can capture the symmetric structure of the image, and the two filters which are relevant to the symmetry of main diagonal can capture similar structure with just 90° rotation. As we know, the polynomial filter is a normalization approximation of the 2-D orthogonal polynomials, therefore polynomial filters are also orthogonal to each other in vector form. A $N \times N$ patch vector lies in the space $\mathbb{R}^{N^2\times 1}$ and any given $N \times N$ patch vector can be linearly represented using our proposed N^2 orthogonal polynomial filters as the basis set. Therefore, the polynomial filters can be considered as a set of complete orthogonal basis, which implies that the response of the filters is unique and compact. Polynomial filters can represent the space in corresponding local neighborhood without losing significative and inherent information. In summary, polynomial filters can achieve a powerful decomposition which could not only capture the details in multiple orientations and scales, but also keep the discriminant information. Our polynomial filters have the following characteristics:

- The filters can reflect the underlying surface structure information.
- The first filter is a mean filter, and all the other filters are zero-mean.
- The filters are multi-frequency, multi-orientation, and symmetric.
- The filters are orthogonal to each other and complete, compact.

In [25], the 2-D discrete Chebyshev polynomials were first proposed and exploited to fit the underlying surface for digital image. We share the construction of the masks with the polynomials and surface fitting concept. But the subsequent processing is different. They tried to recover the input image by further convolving the polynomials with the surface fitting coefficients which are obtained using the masks. Then they applied other operators for edge detection. However we have discovered and systematically summarized the potential and characteristics of the masks which we called chebyshev polynomial filters. We exploit the masks to provide a powerful decomposition of an input face image. The decomposition can transform the input image into a relatively stable and meaningful attributes domain, where we can make advantage

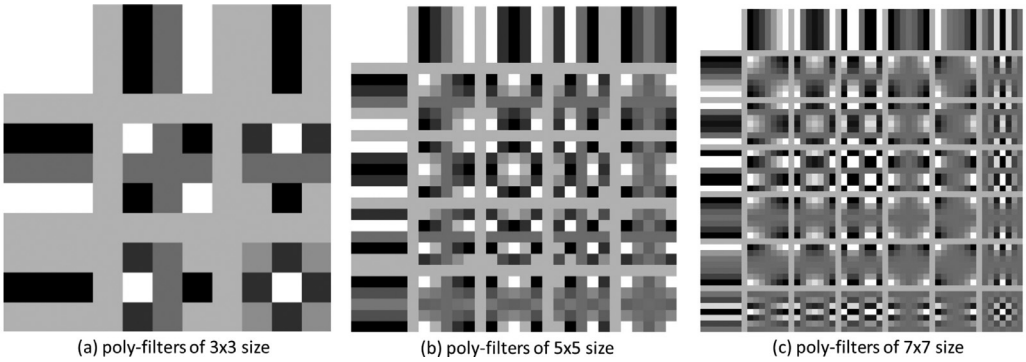

Fig. 3. Examples of polynomial filters for local neighborhood of size (a) 3×3, (b) 5×5 and (c) 7×7.

of the different details in multiple frequencies, orientations and scales. [25] mainly applies the polynomials and surface fitting for image enhancement or denoising while we make use of the powerful decomposition and representation of the masks for face descriptor extraction.

3. Polynomial contrast binary patterns

3.1. Polynomial contrast binary patterns

In this section, we will describe in detail our proposed polynomial contrast binary patterns (PCBP). As shown in above section, polynomial filters decompose a face image into multiple directions from coarser to finer scales. In the coarser scale the facial components are well represented with strong edges, while in the finer scale the details are mainly exhibited with weaker edges such as curves and lines. As polynomial filters are a set of complete orthogonal basis, it is easy to see that if two patches are the same (or different), the polynomial responses of them will also be the same (or different). Two similar patches cannot produce the same filter responses. Therefore, response of polynomial filters can keep inter-personal variations but is not robust enough to variations in intra-personal face images.

As we know, the quantified binary encoding scheme of LBP can reduce the intra-personal distance caused by local variations. Therefore we attempt to incorporate quantified binary encoding and polynomial filter responses to keep the inter-personal distance and reduce the intra-personal distance. LBP encodes each pixel in a self-similarity manner with eight binary bits by quantifying the difference of pixel intensity between the central pixel and its neighbors. LBP measures the local neighborhood difference towards only central pixel, which may fail to take advantage of gray-scale dependent information and reduce the intra-personal distance. Moreover, using only central pixel intensity as threshold for quantification may be sensitive to noise. Therefore, we propose to exploit background intensity contrast information for extracting local neighborhood relationship on polynomial filter responses to enhance the accuracy. We name the resulting feature sets as the polynomial contrast binary patterns (PCBP). Denote $C_{n,m}$ as the response map of polynomial filter $\mathbf{w}_{n,m}$ over the input face image. Formally, our PCBP could be described as follows:

$$PCBP_{n,m}(x_c, y_c) = \sum_{p=0}^{7} 2^n s\left(C_{n,m}^p(x_c, y_c) - \mu_{n,m}(x_c, y_c)\right), \tag{16}$$

in which (x_c, y_c) is the location of the central pixel, and m, n is the index of polynomial filter response. $C_{n,m}^p(x_c, y_c)$ is the intensity of the central pixel and its pth neighbor, $\mu_{n,m}(x_c, y_c)$ represents the averaged background intensity for each polynomial attribute

$$\mu_{n,m}(x_c, y_c) = \frac{1}{|Nr(x_c, y_c)|} \sum_{(x,y) \in Nr(x_c, y_c)} C_{n,m}(x, y), \tag{17}$$

where $Nr(x_c, y_c)$ denotes the local neighborhood centered at pixel (x_c, y_c) and $|Nr(x_c, y_c)|$ is the number of the pixels of the local neighborhood. Moreover, $s(u)$ is a non-linear mapping function

$$s(u) = \begin{cases} 1, & u \geq 0, \\ 0, & u < 0. \end{cases} \tag{18}$$

We then apply uniform pattern scheme to further encode these patterns. Fig. 4 demonstrates the filter responses and their PCBP encoded maps for an input image and given polynomial filters. We can see that the responses echo the filters very well. For example, the filters in the first row are designed to extract basic vertical textures and the corresponding responses reflect the filters' excellent ability to capture the expected texture information. Because the filters are multiscale and multidirectional, the responses shown in Fig. 4(c) are able to capture very rich texture information, which ensures our features' performance. As we know, face representation by the spatial histogram can encode both texture and structure information, and handle well with partial variations. The spatial histogram scheme is then exploited to provide a more reliable description. The scheme firstly divides PCBP maps into several non-overlapping blocks, then computes a histogram within each block, and concatenating the histograms of each block into a joint histogram vector for the face representation. The framework of face representation based on PCBP is illustrated in Fig. 5. The difference between PCBP, LGBP and LBP is kinds of filters. As the filters become complicated, the method is more time-consuming.

3.2. Weighted polynomial contrast binary patterns

The decomposition using polynomial filters can capture rich information at multiple scales, orientations, and frequencies. It is reasonable to assume that different polynomial response maps own different discriminative powers and should make unequal contribution to the performance of face recognition. Thus different weights should be assigned to the polynomial response maps. We adopt the Fisher separation criterion (FSC) [19,27] to evaluate the discriminative significance of different polynomial response maps as the weights.

First denote $S(\cdot, \cdot)$ as the similarity measure, such as cosine distance. And $F_{n,m}$ is used to denote the PCBP feature vector for the response maps of the polynomial filter $\mathbf{w}_{n,m}$. Let the number of the subjects in the training set be C. Then for the (n, m)th polynomial maps, we can compute the mean and variance of the intra-class similarities, respectively denoted as $m_I^{n,m}$ and $\sigma_I^{n,m}$, as

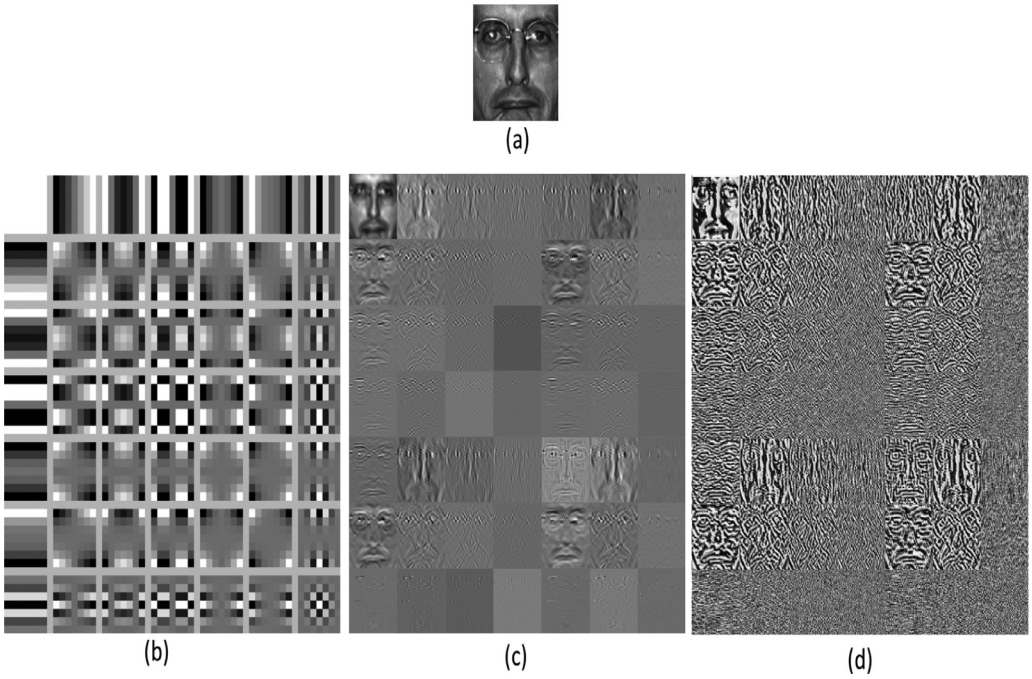

Fig. 4. (a) An input face image; (b) 49 polynomial filters of size 7×7; (c) The responses of (a) using the polynomial filters in (b); (d) The polynomial contrast binary patterns (PCBP) encoded maps corresponding to (c).

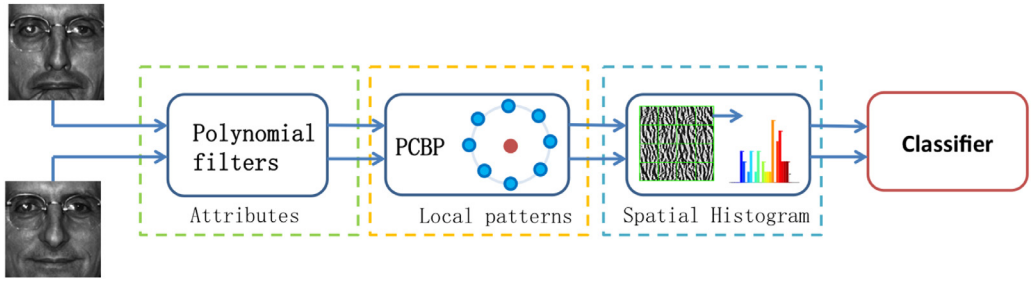

Fig. 5. The four stage flow-chart of the PBP-based face representation and recognition method. For each gallery and probe face image, we apply the polynomial filters, extract polynomial contrast binary patterns, divide them into blocks and concatenate histograms within blocks into a feature vector for face representation. Finally, the nearest neighborhood classifier is used to measure the sample similarity.

follows:

$$m_I^{n,m} = \frac{1}{C}\sum_{i=1}^{C}\frac{2}{N_i(N_i-1)}\sum_{k=2}^{N_i}\sum_{j=1}^{k-1} S(F_{i,j}^{n,m}, F_{i,k}^{n,m}), \tag{19}$$

$$\sigma_I^{n,m} = \left(\frac{2}{\left(\sum_{i=1}^{C} N_i(N_i-1)\right)-2} \times \sum_{i=1}^{C}\sum_{k=2}^{N_i}\sum_{j=1}^{k-1}\left(S(F_{i,j}^{n,m}, F_{i,k}^{n,m}) - m_I^{n,m}\right)^2\right)^{\frac{1}{2}}, \tag{20}$$

where $F_{i,j}^{n,m}$ denotes the histogram vector of the (n, m)th polynomial maps for the jth sample of the ith class, and N_i is the sample number of the ith class in the training set. Likewise, we can also compute the mean and variance of the extra-class similarities, denoted as m_E^l and σ_E^l, as follows:

$$m_E^{n,m} = \frac{2}{C(C-1)}\sum_{i=1}^{C-1}\sum_{j=i+1}^{C}\frac{1}{N_iN_j}\sum_{k=1}^{N_i}\sum_{m=1}^{N_j} S(F_{i,k}^{n,m}, F_{i,m}^{n,m}), \tag{21}$$

$$\sigma_E^{n,m} = \left(\frac{1}{\left(\sum_{i=1}^{C-1}\sum_{j=i+1}^{C} N_iN_j\right)-1} \times \sum_{i=1}^{C-1}\sum_{j=i+1}^{C}\sum_{k=1}^{N_i}\sum_{l=1}^{N_j}(S(F_{i,k}^{n,m}, F_{j,l}^{n,m}) - m_E^{n,m})^2\right)^{\frac{1}{2}}. \tag{22}$$

Finally, we can compute the weight for the (n, m)th polynomial maps, denoted as $\omega^{n,m}$, based on the Fisher criterion [27] as follows:

$$\omega^{n,m} = \frac{(m_I^{n,m} - m_E^{n,m})^2}{(\sigma_I^{n,m})^2 + (\sigma_E^{n,m})^2}. \tag{23}$$

The larger $\omega^{n,m}$ implies the corresponding polynomial map holds more discriminative power. Therefore, our proposed WPCBP is composed of the weighted feature vector of PCBP for each polynomial response map. In other words, the weighting scheme can be applied when we compute the similarity of two face images in the testing phase. The similarity between the gallery and probe face images could be calculated as a weighted summation of the similarities between the corresponding feature vector:

$$S(F_g, F_p) = \sum_{n=0}^{N-1}\sum_{m=0}^{M-1} \omega^{n,m} S(F_g^{n,m}, F_p^{n,m}). \quad (24)$$

We apply the weight learning scheme in two different ways. On one hand, the weights can be learned using the histogram vector of PCBP (using PCBP histogram as the $F^{n,m}$ in Eq. (24)) and histogram intersection (HI) similarity [28]. As the dimensionality of the obtained feature vectors is relatively high, we can employ the Fisher linear discriminant (FLD) [4] to reduce the feature dimensionality and select features. The corresponding method with the weighting and feature selection scheme is called WPCBP+FLD(HI). Similarly, we can also learn weights using the feature vector after feature selection using FLD as the $F^{n,m}$ and cosine distance (CD) as similarity measure. The corresponding method is named WPCBP+FLD(CD).

The decomposition using polynomial filters can capture rich information at multiple scales, orientations, and frequencies. Applying contrast patterns and the spatial histogram model to decomposed images can well capture the discriminative texture and structure information and make the face representation less sensitive to the expression and pose variations and to small registration errors. The weighting scheme can further enhance the discriminative power of PCBP. In summary, these properties of PCBP allow our features to convey rich information of face images, to minimize the intra-person variations caused by illumination, expression, and aging, and to inherit their discriminative powers as well.

4. Experiments for face recognition

Extensive experiments have been carried out to illustrate the effectiveness of the proposed method. In our experiments, two large publicly available face databases, FERET [29], LFW [30] are used to evaluate the performance of different methods. These face databases contain face images with various appearance changes, including expression, illumination, aging, occlusion etc. The proposed method has shown its robustness and accuracy in these variations.

In our experiments, we divide the PCBP encoded maps into 8×8 blocks, and adopt the uniform patterns with 59 bins. The nearest neighbor classifier is used throughout all experiments. For feature sets after feature selection using FLD, only the most critical 200 dimensions are retained for each polynomial maps and the cosine distance is adopted as the similarity measure.

4.1. Experiments on the FERET database

The FERET database is widely used to evaluate face recognition algorithms. The basic gallery *fa* contains 1196 images of 1196 subjects. There are four probe sets with different environment variations. The *fb* set contains 1195 images of 1195 subjects taken at the same time as the gallery images but with different expressions; the *fc* set includes 194 images of 194 subjects taken at the same time under significantly different illumination conditions; the *dup I* set contains 722 images of 243 subjects taken between one minute and 1031 days after the gallery image was taken; the *dup II* set contains 234 images of 75 subjects taken at least 18 months after the gallery image was taken. For our experiment, all images are aligned, cropped and resized to 128×128 with the centers of the eyes located at (29,34) and (99,34). Some sample face images from FERET are demonstrated in Fig. 6.

Fig. 6. Sample face images in FERET face database, and the images in each row correspond to the same subject.

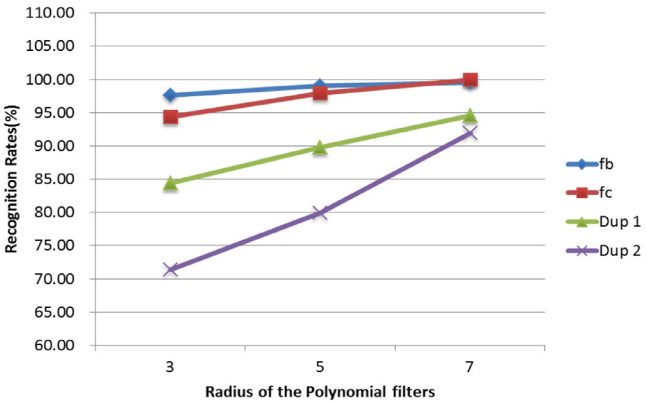

Fig. 7. The recognition rates versus the filter radius upon the FERET database.

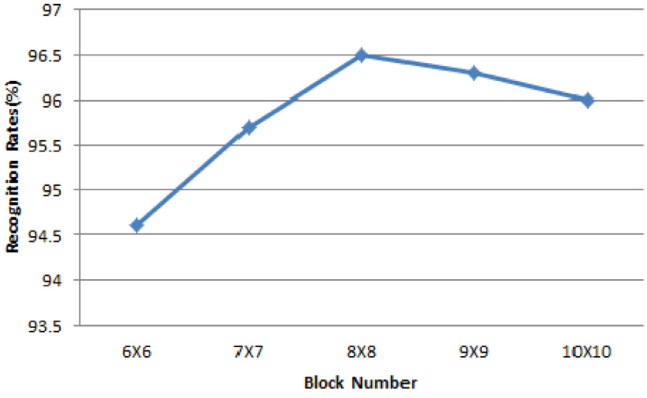

Fig. 8. The recognition rates versus block number upon the FERET database.

We first evaluate 3 different filter radius (3,5,7) on FERET database and set the parameters empirically. Fig. 7 plot the recognition rates versus variant filter radius. It can seen that when radius is 7, the recognition rates are relatively higher. Increasing radius may help improve the accuracy but would also brings extra cost both in time and storage. Setting radius as 7 may be a useful trade-off for accuracy and computational cost. Therefore, in the following experiments we all use the 49 filters of 7×7.

We also test different block sizes and FLD dimensions. Figs. 8 and 9 show the recognition rates on FERET with different block sizes and FLD dimensions, respectively. It is clear that our setting of parameters is the best choice.

To better evaluate the effectiveness of our proposed method, we compare our methods with other state-of-the-art methods

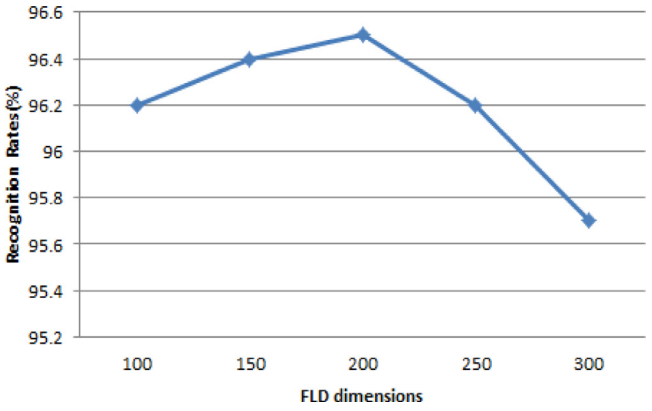

Fig. 9. The recognition rates versus the FLD dimensions upon the FERET database.

Table 1
Recognition rates in percentage compared with the state-of-the-art methods tested with the FERET evaluation protocol. Avg denotes the average rate across the four probe sets. Note the results with * are cited from their original papers.

Methods	fb	fc	dup I	dup II	Avg
LBP [14]*	97.0	79.0	66.0	64.0	82.7
LGBP [18]*	98.0	97.0	74.0	71.0	87.8
HGPP [19]*	97.5	99.5	79.5	77.8	90.2
LLGP [31]*	99.0	99.0	80.0	78.0	91.1
high-dim LBP [32]	99.3	73.7	88.4	78.6	91.8
DT-LBP [33] *	99.0	100.0	84.0	80.0	92.6
DLBP [34]*	99.0	99.0	86.0	85.0	93.6
Tan's method [35]*	98.0	98.0	90.0	85.0	94.2
Zou's method [36]*	99.5	99.5	85.0	79.5	93.0
LGBP+LGXP [20]*	99.0	99.0	94.0	93.0	96.9
POEM+WPCA [22]*	99.6	99.5	88.8	85.0	94.8
DFD+WPCA [23]*	99.4	100.0	91.8	92.3	96.4
CBFD+WPCA [24]*	99.8	100.0	93.5	93.2	97.2
GOM [37]*	**99.9**	100.0	**95.7**	93.1	**97.9**
PCBP + FLD	99.5	**100.0**	93.1	88.9	96.5
WPCBP + FLD (CD)	99.5	**100.0**	94.0	91.9	97.1
WPCBP + FLD (HI)	99.5	**100.0**	94.7	**94.0**	97.5

reported in literatures. To obtain the results of high-dimensional LBP [32] on FERET, we find 20 facial feature points based on the method proposed in [38], then extract LBP features from patches centered at the feature points in multiscale images and finally adopt PCA and FLD to select features. All the results are summarized in Table 1. As can be seen, PCBP+FLD with no weighting processing performs better than most of the methods, which demonstrates the effectiveness of our proposed PCBP algorithm. Moreover, the weighting scheme can further improve the performance. The weighting versions WPCBP+FLD (CD) and WPCBP+FLD (HI) achieve higher recognition rates than PCBP+FLD, by mainly improving the performance on the aging probe subsets Dup I and Dup II. The proposed WPCBP+FLD (HI) can achieve the best face recognition results on the fc and Dup II subsets, and comparable results on the fb and Dup I subsets. In average WPCBP+FLD (HI) performs very excellently with accuracy up to 97.5%, just 0.4% lower than Gabor Ordinal Measures (GOM) [37]. GOM is an approach using ordinal measures to encode Gabor filtering responses. It costs much computational time and space because it applies 8 ordinal filters to encode a total of 80 Gabor filtering responses to binary codes. Different from GOM, our method does not focus on complicated coding schemes and aims at searching for the novel polynomial filters to capture facial textures better. In the end of this subsection, we will present some experiments to compare our polynomial filters with Gabor filters to further explain the difference between GOM and our PCBP. In Table 1, we can also see that the high-dim LBP performs not so well on FERET. This is because all the images in FERET database are aligned and cropped properly and high-dim LBP's advantage over pose variations by finding out facial points is much reduced. In summary, our best result is much competitive compared with the state-of-the-art methods.

It can also be observed from Table 1 that, both non-weighted PBCP and weighted PCBP all achieve the best accuracy up to 100% on fc subset, which demonstrates the effectiveness of our methods towards illumination variations. We further display some face images under different illumination conditions along with their corresponding PCBP encoded images in Fig. 10. The PCBP encoded images are obtained using 49 Chebyshev polynomial filters of 7. As can be seen, though the input face images are under quite different illumination conditions, their corresponding PCBP encoded images can extract the similar structure patterns and resist the illumination changes.

Figs. 11 and 12 respectively show the weight values for the polynomial filter maps in WPCBP+FLD(CD) and WPCBP+FLD(HI). From left to right every 7 bars correspond to the filters in the same row of Fig. 4(b). It can be seen that the lower frequency components hold the larger weights, which implies that they own stronger discriminative powers. The 7th filters of each row

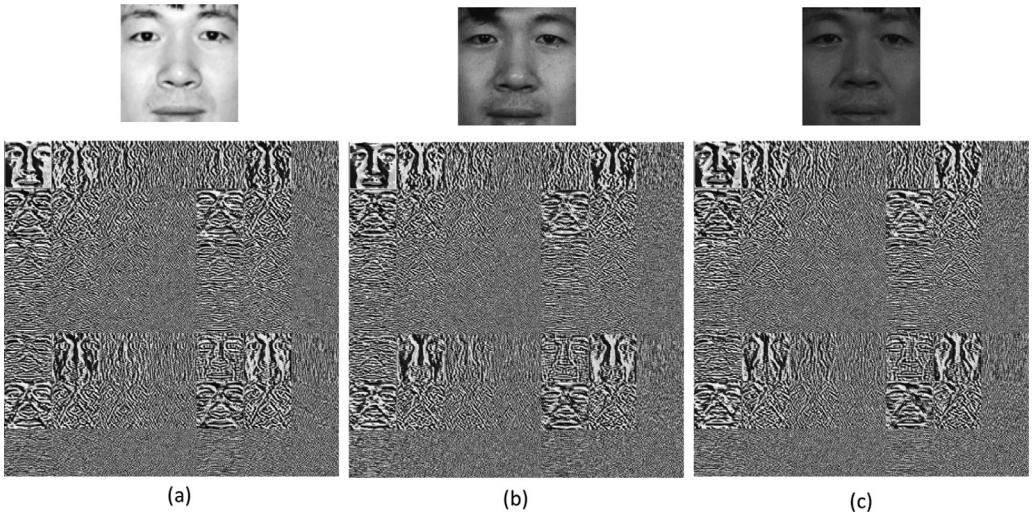

Fig. 10. Face images and corresponding PCBP encoded images under different illumination conditions of FERET database.

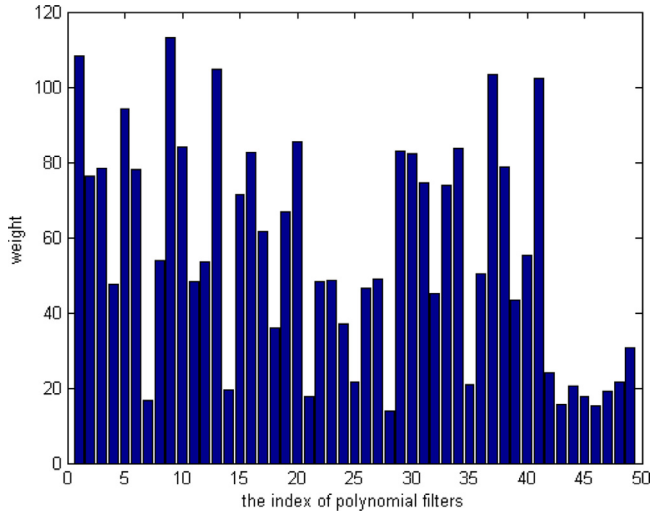

Fig. 11. Weight values (estimated using cosine distance) for 49 polynomial filters in WPCBP+FLD(CD).

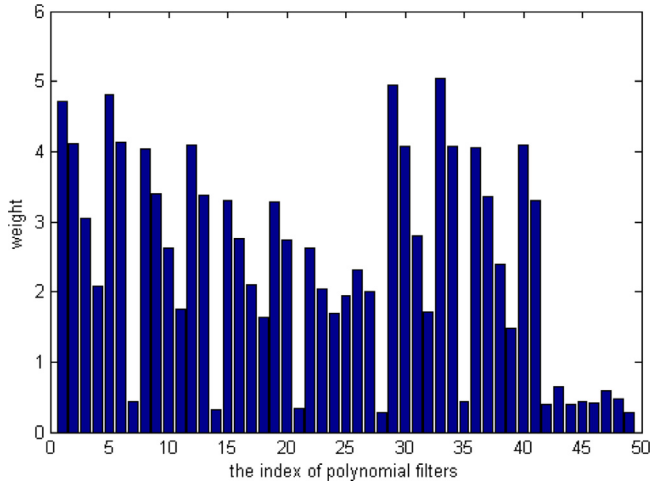

Fig. 12. Weight values (estimated using histogram intersection) for 49 polynomial filters in WPCBP+FLD(HI).

(7,14,21,28,35,42,49th filter) hold the smallest weights, and they mainly capture the higher frequency parts. As we know, the noise and wrinkle textures occurred in aged faces which would reduce the accuracy mainly in high-frequency bands. It is advantageous to reduce the significance of the higher-frequency bands by assigning them smaller weights. Therefore, the improvement on Dup I and Dup II via weighting scheme is due to these two subsets reflect the aging variation which brings many high-frequency components like wrinkle textures.

To further demonstrate our polynomial filters' effectiveness, we compare our method with Gabor filters-based method and LBP. We use Gabor filters (8 directions in 5 scales) to process the images and encode the responses the same as our method did (the same CBP operations and the same histogram intersection as distance metric). Table 2 shows the performance of the two filters on FERET. All the results here were obtained without FLD or any other feature selection method. From Table 2 we can obtain two conclusions. First, the methods combining filters and binary coding generally perform better than simple LBP. Second, our polynomial filters are able to capture more useful information and perform better in face recognition than Gabor filters do. To analyze the reason of the results, we take out the responses of these filters and show an instance in Fig. 13. From the figure, both the two kinds of filters can capture multiscale and multidirectional texture information, however our polynomial filters seem to be able to extract these textures finer than Gabor filters do. The textures represented by the responses of polynomial filters look more clear and regular than those of Gabor filters, which makes us believe that the polynomial filters can enhance texture information better than Gabor filters. That's why PCBP outperforms the Gabor-based method in this experiment. Considering the methods shown in Table 1, we believe that it's possible to improve our accuracy to exceed the performance of Gabor-based GOM [37] through modifying other parts in recognition, such as the coding method and the classifier, which will be our future work.

Table 2
Recognition rates in percentage compared with Gabor filters method and LBP tested with the FERET evaluation protocol.

Methods	fb	fc	dup I	dup II	Avg
PCBP	94.8	98.5	78.5	73.5	88.0
Gabor + CBP	94.6	93.3	69.7	61.1	83.5
LBP [14]*	97.0	79.0	66.0	64.0	82.7

4.2. Experiments on the LFW database

To further demonstrate the effectiveness of our proposed method, we carry out experiments for unconstrained face verification task using the real-world Labeled Faces in the Wild (LFW) database [30]. LFW database contains up to 13,233 face images of 5749 subjects collected from the web including large variation in pose, lighting, expression, age, etc. In our experiment, we exploit the aligned version (LFW-a) of the faces as provided in [39]. Then the aligned images are cropped and resized to 88×88 pixels. Some sample face images from LFW are shown in Fig. 14.

According to the database protocol, the LFW database is divided into two "Views": View 1 is designed for model selection and View 2 is mainly for performance reporting and comparison. In our experiment, View 2 is used for performance evaluation and the View 1 dataset is not used. View 2 provides 10 random splits to obtain 10 subsets. There are 300 intra-class pairs and 300 inter-class pairs in each subset. The performance is reported using the 10-fold cross validation. We use 5400 pairs of images in the 9 subsets for the configuration of matching threshold, and report the accuracy on the remaining 10th subset.

We follow unrestricted protocol (know identity information) during training. As we employ FLD as feature selection method, we need a training set for learning transform matrix. For each of the 10-fold cross-validation tests, we use identities with at least three images for training. We first pick out the identities with at least three images and for the identities owning larger than 10 images, only randomly chosen 10 images are used. The number of training images is about 4862.

Then we make a comparison with the state-of-the-art methods using single and multi-descriptor representations under the unrestricted protocol. Fig. 15 and Table 3 are the comparison of our method WPCBP+FLD (HI) with the state-of-the-art methods, evaluated in terms of the ROC curves and mean verification rates of the 10-fold cross validation test, respectively.

As shown in Fig. 15 and Table 3, our method achieves the face verification accuracy 92.85%, significantly outperforming most of the state-of-the-art methods. Our method adopts a single feature PCBP as the face representation, and outperforms a number of methods with combined multiple-features [40–42]. The accuracy of our method is only slightly less than two methods, The Fisher vector faces (FVF) and the high-dim LBP. The Fisher vector faces

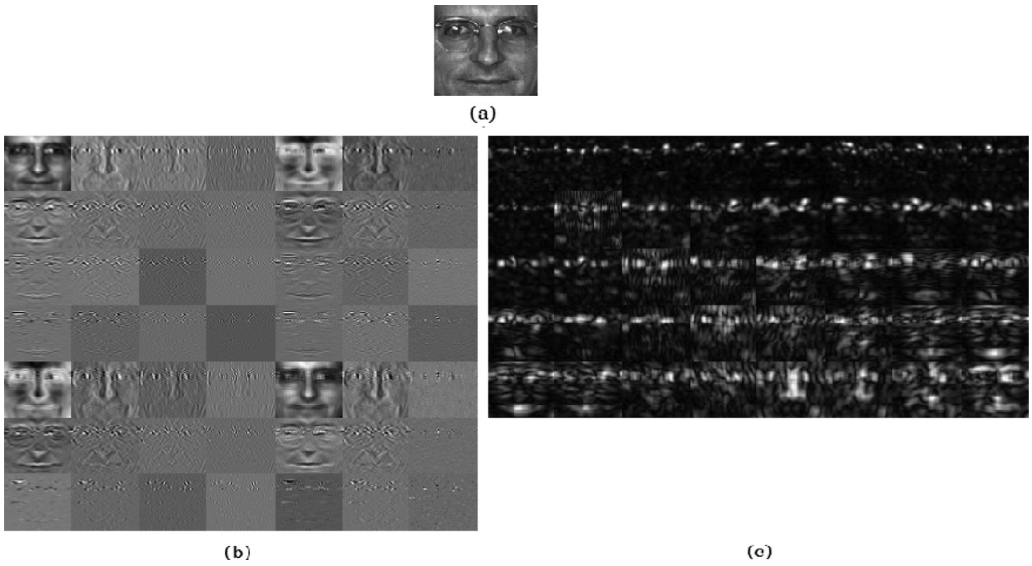

Fig. 13. (a) An input face image; (b) The responses of (a) using the polynomial filters of size 7×7; (c) The responses of (a) using the Gabor filters of 8 directions in 5 scales.

Fig. 14. Cropped sample face images in the LFW face database. The LFW database includes great variations in expression, illumination, occlusion, pose, etc.

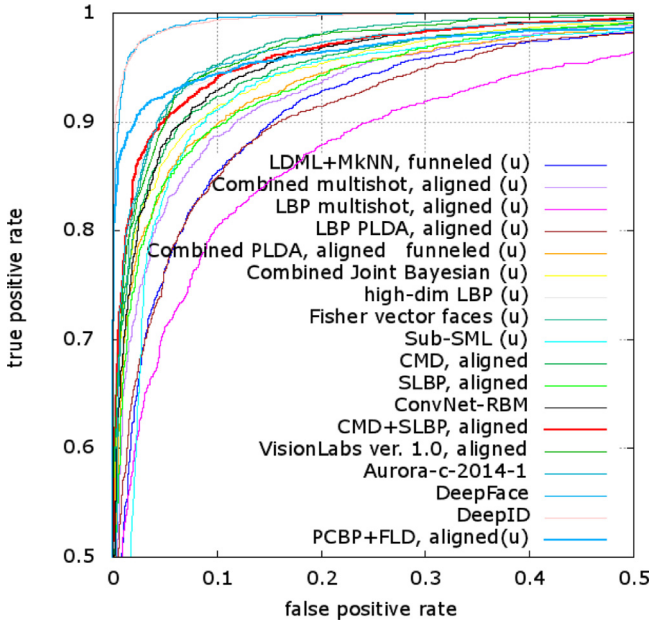

Fig. 15. Comparison of the ROC curves over View 2 on the LFW database between our method and other state-of-the-art techniques for face verification in the LFW-unrestricted setting.

(FVF) exploits the dense SIFT feature and Fisher Vectors (FVs) encoding, followed by a discriminative dimension-reduction with the Joint Bayesian method. It uses a different alignment method from the one we use. Moreover, it takes more computational cost with the dense SIFT operator and sophisticated encoding algorithms.

Table 3
Comparison of classification accuracy ($\pm$ standard error) for our method and other state-of-the-art methods operating in the unrestricted setting.

Methods	Accuracy
LDML-MkNN [40]	0.8750 $\pm$ 0.0040
LBP multishot [41]	0.8517 $\pm$ 0.0061
Combined multishot [41]	0.8950 $\pm$ 0.0051
LBP PLDA [42]	0.8733 $\pm$ 0.0055
Combined PLDA [42]	0.9007 $\pm$ 0.0051
Combined Joint Bayesian [44]	0.9090 $\pm$ 0.0148
Sub-SML [45]	0.9075 $\pm$ 0.0064
VMRS [46]	0.9205 $\pm$ 0.0045
Fisher vector faces [47]	0.9303 $\pm$ 0.0105
high-dim LBP [32]	0.9318 $\pm$ 0.0107
DeepFace [11]	0.9725 $\pm$ 0.0081
DeepID [10]	0.9745 $\pm$ 0.0026
LightCNN [43]	0.9883 $\pm$ 0.0083
WPCBP + FLD (HI)	0.9285 $\pm$ 0.0094

The high-dim LBP proposes a rotated sparse regression technique as the feature selection method. For LFW database, the high-dim LBP features are sampled around 27 landmarks in 5 scales which rely on the sophisticated face landmark detectors. When facial images have pose variations, high-dim LBP perhaps performs better than general features because it first preprocesses images with a face alignment algorithm. As [32] presents, if the face alignment algorithm is not very accurate or there are only few feature points, the performance of high-dim LBP will degrade severely. To obtain very accurate feature points and to store very high dimensional features, the high-dim LBP costs much more time and space for the computations than PCBP does. Deep learning methods (DeepID [10], DeepFace [11], LightCNN [43]) can achieve much better accuracy than the other methods because they use huge amounts of data to train a deep network. For example, DeepFace made use of millions of images from other databases for training. Compared with deep learning methods, our method extracts local texture information very directly using the proposed polynomial filters and does not need any training. In summary, the results in the unrestricted setting clearly confirm that our proposed methods can obtain an effective and efficient face descriptor for real-world face verification task.

5. Conclusion

We proposed a novel and effective face representation model, namely polynomial contrast binary pattern (PCBP), which is based on the polynomial filters. The effectiveness of PCBP comes from several aspects including the decomposition of polynomial filters from coarser to finer, local contrast binary coding, the spatial histogram modeling. We also adopt the Fisher separation criterion to learning weights for different polynomial filter maps. The weighting scheme helps improve the performance when handling noise and aging variations. To further reduce the feature dimensionality, we exploited FLD to select the most discriminative information. Experimental results on several publicly available databases have evidently illustrated the effectiveness of the proposed method. Due to its promising performance in the face recognition applications, we can expect that the proposed method is a good choice for the recognition of other objects.

Acknowledgments

This work was supported in part by the National Science Foundation of China under Grant No. 61271393, and the Macau Science and Technology Development Fund under Grant 106/2013/A3 and by the Research Committee at University of Macau under Grants MYRG113(Y1-L3)-FST12- ZYC, MRG001/ZYC/2013/FST and MYRG2014- 00003-FST.

References

[1] W. Zhao, R. Chellappa, A. Rosenfeld, P. Phillips, Face recognition: a literature survey, ACM Comput. Surv. (2003) 399–458.
[2] S. Li, A. Jain, Handbook of Face Recognition (2nd Edition), Springer-Verlag, 2011.
[3] M.A.T.A.P. Pentland, Eigenfaces for recognition, J. Cognit. Neurosci. 3 (1) (1991) 71–86.
[4] P. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[5] J. Tenenbaum, V. De Silva, J. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.
[6] S. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[7] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, Face recognition using laplacianfaces, IEEE Trans. Pattern Anal. Mach. Intell. 27 (3) (2005) 328–340.
[8] X. He, D. Cai, S. Yan, H. Zhang, Neighborhood preserving embedding, in: International Conference on Computer Vision (ICCV), 2, 2005, pp. 1208–1213.
[9] M. Yang, Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods, in: International Conference Automatic Face and Gesture Recognition, 2002, pp. 215–220.
[10] Y. Sun, X. Wang, X. Tang, Deep learning face representation from predicting 10,000 classes, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
[11] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, Deepface: Closing the gap to human-level performance in face verification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701–1708.
[12] Y. Sun, Y. Chen, X. Wang, X. Tang, Deep learning face representation by joint identification-verification, in: Advances in Neural Information Processing Systems, 2014, pp. 1988–1996.
[13] Y. Sun, X. Wang, X. Tang, Deeply learned face representations are sparse, selective, and robust, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
[14] T. Ahonen, A. Hadid, M. Pietikäinen, Face description with local binary patterns: Application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (12) (2006) 2037–2041.
[15] C. Liu, H. Wechsler, Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition, IEEE Trans. Image Process. 11 (4) (2002) 467–476.
[16] L. Wolf, T. Hassner, T. Taigman, Effective unconstrained face recognition by combining multiple descriptors and learned background statistics, IEEE Trans. Pattern Anal. Mach. Intell. 33 (10) (2011) 1978–1990.
[17] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, in: Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures, 2007, pp. 235–249.
[18] W. Zhang, S. Shan, W. Gao, X. Chen, H. Zhang, Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition, in: International Conference on Computer Vision (ICCV), 2005, pp. 786–791.
[19] B. Zhang, S. Shan, X. Chen, W. Gao, Histogram of gabor phase patterns (hgpp): a novel object representation approach for face recognition, IEEE Trans. Image Process. 16 (1) (2007) 57–68.
[20] S. Xie, S. Shan, X. Chen, J. Chen, Fusing local patterns of gabor magnitude and phase for face recognition, IEEE Trans. Image Process. 19 (5) (2010) 1349–1361.
[21] Z. Chai, Z. Sun, H. Mendez-Vazquez, R. He, T. Tan, Gabor ordinal measures for face recognition, IEEE Trans. Inf. Forensics Secur. 9 (1) (2014) 14–26.
[22] N. Vu, A. Caplier, Enhanced patterns of oriented edge magnitudes for face recognition and image matching, IEEE Trans. Image Process. 21 (3) (2011) 1352–1365.
[23] Z. Lei, M. Pietikainen, S.Z. Li, Learning discriminant face descriptor, IEEE Trans. Pattern Anal. Mach. Intell. (2013).
[24] J. Lu, V. Liong, X. Zhou, J. Zhou, Learning compact binary face descriptor for face recognition, IEEE Trans. on Patt. Analysis and Machine Intell. (99) (2015). 1–1
[25] R.M. Haralick, Digital step edges from zero crossing of second directional derivatives, IEEE Trans. Pattern Anal. Mach. Intell. (1) (1984) 58–68.
[26] J. Prewitt, Object Enhancement and Extraction, 75–149, Academic Press, New York, 1970.
[27] R.M. Parry, I. Essa, Feature weighting for segmentation, in: ISMIR, 4, 2004, pp. 10–14.
[28] M. Swain, D. Ballard, Color indexing, Int. J. Comput. Vis. 7 (1) (1991) 11–32.
[29] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The feret evaluation methodology for face-recognition algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 22 (10) (2000) 1090–1104.
[30] G.B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, Technical Report, University of Massachusetts, Amherst, 2007.
[31] S. Xie, S. Shan, X. Chen, X. Meng, W. Gao, Learned local gabor patterns for face representation and recognition, Signal Process. 89 (12) (2009) 2333–2344.
[32] D. Chen, X. Cao, L. Wang, J. Sun, Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification, in: Computer Vision and Pattern Recognition (CVPR), 2013, pp. 57–64.
[33] D. Maturana, D. Mery, A. Soto, Face recognition with decision tree-based local binary patterns, in: Asian Conference on Computer Vision (ACCV), Springer, 2010, pp. 618–629.
[34] D. Maturana, D. Mery, A. Soto, Learning discriminative local binary patterns for face recognition, in: International Conference Automatic Face and Gesture Recognition and Workshops (FG), 2011, pp. 470–475.
[35] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, in: Analysis and Modeling of Faces and Gestures, 2007, pp. 168–182.
[36] J. Zou, Q. Ji, G. Nagy, A comparative study of local matching approach for face recognition, IEEE Trans. Image Process. 16 (10) (2007) 2617–2628.
[37] Z. Chai, Z. Sun, H. Mendez-Vazquez, R. He, T. Tan, Gabor ordinal measures for face recognition, IEEE Trans. Inf. Forensics Secur., 9 (1) (2014) 14–26.
[38] Y. Sun, X. Wang, X. Tang, Deep convolutional network cascade for facial point detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3476–3483.
[39] L. Wolf, T. Hassner, Y. Taigman, Similarity scores based on background samples, in: Asian Conference on Computer Vision (ACCV), 2009, pp. 88–97.
[40] M. Guillaumin, J. Verbeek, C. Schmid, Is that you? metric learning approaches for face identification, in: International Conference on Computer Vision (ICCV), 2009, pp. 498–505.
[41] Y. Taigman, L. Wolf, T. Hassner, Multiple one-shots for utilizing class label information, in: The British Machine Vision Conference (BMVC), 2009, pp. 1–12.
[42] P. Li, Y. Fu, U. Mohammed, J.H. Elder, S.J. Prince, Probabilistic models for inference about identity, IEEE Trans. Pattern Anal. Mach. Intell. 34 (1) (2012) 144–157.
[43] X. Wu, R. He, Z. Sun, T. Tan, A light cnn for deep face representation with noisy labels, arXiv:1511.02683 (2015).
[44] D. Chen, X. Cao, L. Wang, F. Wen, J. Sun, Bayesian face revisited: A joint formulation, in: European Conference Computer Vision (ECCV), 2012, pp. 566–579.
[45] Q. Cao, Y. Ying, P. Li, Similarity metric learning for face recognition, in: International Conference on Computer Vision (ICCV), 2013, pp. 2408–2415.
[46] O. Barkan, J. Weill, L. Wolf, H. Aronowitz, Fast high dimensional vector multiplication face recognition, in: International Conference on Computer Vision (ICCV), 2013, pp. 1960–1967.
[47] K. Simonyan, O.M. Parkhi, A. Vedaldi, A. Zisserman, Fisher vector faces in the wild, in: British Machine Vision Conference (BMVC), 2013, pp. 1–12.

Zhen Xu received the B. E. degree from Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan, China, in 2015. He is currently working towards the Master degree in electronics engineering at Tsinghua University, Beijing, China. His research interests include applications of image processing and pattern recognition.

Yinyan Jiang received the B.E. and M.E. degrees in Electronic Engineering from Tsinghua University, Beijing, China, in 2012 and 2015, respectively. Her research interests include image processing and pattern recognition in biometrics and super-Resolution.

Yichuan Wang received the B.E. degree in Electronic Engineering from Tsinghua University, Beijing, China, in 2014. He is currently working towards the Master degree in electronics engineering at Tsinghua University. His research interests include applications of image processing and pattern recognition in biometrics.

Yicong Zhou received the B.S. degree in electrical engineering from Hunan University, Changsha, China, and the M.S. and Ph.D. degrees in electrical engineering from Tufts University, Medford, MA, USA. He is currently an Associate Professor and the Director of the Vision and Image Processing Laboratory, Department of Computer and Information Science, University of Macau, Macau, China. His research interests include chaotic systems, multimedia security, image processing and understanding, and machine learning.Dr. Zhou was a recipient of the Third Price of the Macau Natural Science Award in 2014. He serves as an Associate Editor for the Neurocomputing, the textitJournal of Visual Communication and Image Representation, and the textitSignal Processing: Image Communication. He is the Co-Chair of Technical Committee on Cognitive Computing in the IEEE Systems, Man, and Cybernetics Society.

Weifeng Li received the M.E. and Ph.D. degrees in Information Electronics at Nagoya University, Japan in 2003 and 2006, respectively. He joined the Idiap Research Institute, Switzerland in 2006, and in 2008 he moved to Swiss Federal Institute of Technology, Lausanne (EPFL), Switzerland as a research scientist. Since 2010 he has been an associate professor in the Department of Electronic Engineering / Graduate School at Shenzhen, Tsinghua University, China. He has published more than 100 research papers in the areas of audio and visual signal processing, Biometrics, Human–Computer Interactions (HCI), and machine learning techniques. He is a member of the IEEE and IEICE.

Qingmin Liao received the B.S. degree in radio technology from the University of Electronic Science and Technology of China, Chengdu, China, in 1984, and the M.S. and Ph.D. degrees in signal processing and telecommunications from the University of Rennes 1, Rennes, France, in 1990 and 1994, respectively. Since 1995, he has been joining with Tsinghua University, Beijing, China. In 2002 he became Professor in the Department of Electronic Engineering of Tsinghua University. Since 2010, he has been the director of the Division of Information Science and Technology in the Graduate School at Shenzhen, Tsinghua University. He is also affiliated with the Shenzhen Key Laboratory of Information Science and Technology (Director), China. Over the last 30 years, he has published over 100 peer-reviewed journal and conference papers. His research interests include image/video processing, transmission and analysis; biometrics; and their applications to teledetection, medicine, industry, and sports.