Multi-Directional Multi-Level Dual-Cross Patterns for Robust Face Recognition

Changxing Ding, *Student Member, IEEE*, Jonghyun Choi, *Student Member, IEEE*,
Dacheng Tao, *Fellow, IEEE*, and Larry S. Davis, *Fellow, IEEE*

Abstract—To perform unconstrained face recognition robust to variations in illumination, pose and expression, this paper presents a new scheme to extract "Multi-Directional Multi-Level Dual-Cross Patterns" (MDML-DCPs) from face images. Specifically, the MDML-DCPs scheme exploits the first derivative of Gaussian operator to reduce the impact of differences in illumination and then computes the DCP feature at both the holistic and component levels. DCP is a novel face image descriptor inspired by the unique textural structure of human faces. It is computationally efficient and only doubles the cost of computing local binary patterns, yet is extremely robust to pose and expression variations. MDML-DCPs comprehensively yet efficiently encodes the invariant characteristics of a face image from multiple levels into patterns that are highly discriminative of inter-personal differences but robust to intra-personal variations. Experimental results on the FERET, CAS-PERL-R1, FRGC 2.0, and LFW databases indicate that DCP outperforms the state-of-the-art local descriptors (e.g. LBP, LTP, LPQ, POEM, tLBP, and LGXP) for both face identification and face verification tasks. More impressively, the best performance is achieved on the challenging LFW and FRGC 2.0 databases by deploying MDML-DCPs in a simple recognition scheme.

Index Terms—Face recognition, face image descriptors, face image representation

✦

1 INTRODUCTION

F ACE recognition has been an active area of research due to both the scientific challenge and its potential use in a wide range of practical applications. Satisfactory performance has been achieved but often only in controlled environments. More recently, there has been increased demand for recognition of unconstrained face images, such as those collected from the internet [1] or captured by mobile devices and surveillance cameras [2]. However, recognition of unconstrained face images is a difficult problem due to degradation of face image quality and the wide variations of pose, illumination, expression, and occlusion often encountered in images [3].

A face recognition system usually consists of a face representation stage and a face matching stage. For face matching, multi-class classifiers have been used for face identification, such as the nearest neighbor (NN) classifier and the sparse representation classifier (SRC) [4], with two-class classifiers, such as support vector machine (SVM) and Bayesian analysis, being used for verification. In face representation, good representations discriminate inter-personal differences while being robust to intra-personal variations. Two major categories of face representation methods dominate recent research, namely face image descriptor-based methods [5], [6],

[7], [8] and deep learning-based methods [9], [10]. The former has the advantage of ease of use, data independence, and robustness to real-life challenges such as illumination and expression differences. Consequently, the design of effective face image descriptors is regarded as fundamental for face recognition. Based on design methodology, we can group existing face image descriptors into two groups: hand-crafted descriptors and learning-based descriptors.

Most face image descriptors are hand-crafted, of which Local Binary Patterns (LBP) and Gabor wavelets are two representative methods; Ahonen et al. [5] showed that the texture descriptor LBP is extremely effective for face recognition. LBP works by encoding the gray-value differences between each central pixel and its neighboring pixels into binary codes; the face image is then represented as the concatenated spatial histograms of the binary codes. Many variants of LBP have been proposed. For example, Local Ternary Patterns (LTP) [11] was proposed to enhance the robustness of LBP to noises. Transition LBP (tLBP) [12] and Direction coded LBP (dLBP) [12] were proposed to extract complementary information to LBP using novel encoding strategies. On the other hand, Gabor wavelets aim to encode multi-scale and multi-orientation information of face images [13], [14]. LBP-like descriptors are therefore better at encoding fine-scale information while Gabor-based descriptors extract information on larger scales. Other notable descriptors for face recognition include Local Phase Quantization (LPQ) [15] and Patterns of Oriented Edge Magnitudes (POEM) [16]. In LPQ, the blur-invariant information in face images is encoded and the method shows good performance in unconstrained

• *C. Ding and D. Tao are with the Centre for Quantum Computation & Intelligent Systems and the Faculty of Engineering and Information Technology, University of Technology, Sydney, 81 Broadway Street, Ultimo, NSW 2007, Australia. E-mail: changxing.ding@student.uts.edu.au, dacheng.tao@uts.edu.au.*
• *J. Choi and L.S.Davis are with the Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742. E-mail: {jhchoi,lsd}@umiacs.umd.edu.*

conditions [8]. POEM works by computing the LBP feature on oriented edge magnitudes.

Of the learning-based descriptors, LE [6], Local Quantized Patterns (LQP) [17], and Discriminant Face Descriptor (DFD) [18] have emerged in recent years, which rely on unsupervised or supervised learning techniques to optimize encoders. One important advantage of learning-based descriptors over hand-crafted descriptors is their greater diversity in sampling pattern shapes and the larger sampling size.

Despite the successful application of existing face image descriptors, the following three points are worth considering. First, the textual characteristics of human faces have mostly been overlooked in the design of existing descriptors. Second, it is generally prohibitive for hand-crafted descriptors to adopt a large sampling size due to the complication of the resulting encoding scheme and large feature size (i.e., the number of histogram bins) [17]. However, a large sampling size is desirable since it provides better discriminative power, as has been proved by learning-based descriptors; it is therefore reasonable to ask whether it is genuinely impossible for hand-crafted descriptors to exploit large sampling sizes. Third, the recently proposed descriptors achieve good performance but at the cost of using computationally expensive techniques such as Gabor filtering and codebook learning. It would therefore be desirable to obtain a face image descriptor with superior performance that retains a lower computational cost and feature size.

To address these three limitations of existing techniques, in this paper we present a novel face image descriptor named Dual-Cross Patterns (DCP). Inspired by the unique textural structure of human faces, DCP encodes second-order discriminative information in the directions of major facial components: the horizontal direction for eyes, eyebrows, and lips; the vertical direction for the nose; and the diagonal directions ($\pi/4$ and $3\pi/4$) for the end parts of facial components. The sampling strategy we adopt samples twice as many pixels as LBP. By appropriately grouping the sampled pixels from the perspective of maximum joint Shannon entropy, we keep the DCP feature size reasonable. DCP is very efficient - only twice the computational cost of LBP. Significantly better performance is achieved even when a sub-DCP (denoted herein as DCP-1 and DCP-2) of exactly the same time and memory costs as LBP is used.

A powerful face representation scheme is equally important for unconstrained face recognition as a good face image descriptor. To improve the discrimination of face representations and even the more general image representations, the usual approach is to fuse the information extracted by different descriptors [11], [19], [20] and at different scales [8], [21]. To enhance robustness to pose variation, Wright et al. [22] expanded image patch descriptors with their spatial locations and represented the face image as a histogram of quantized patch descriptors. Ding et al. [23] proposed to represent the face image using only the unoccluded facial texture that is automatically detected in the 3D pose normalized face image. Recently, face representation has benefited from rapid progress in face alignment techniques [24]. For example, Chen et al. [25] built a high dimensional face representation by extracting LBP features from multi-scale image patches around dense facial feature points. For other representative face representations, we direct readers to a recent survey [3].

In this paper we propose a highly robust and discriminative face representation scheme called Multi-Directional Multi-Level Dual-Cross Patterns (MDML-DCPs). Specifically, the MDML-DCPs scheme employs the first derivative of Gaussian operator to conduct multi-directional filtering to reduce the impact of differences in illumination. DCP features are then computed at two levels: 1) holistic-level features incorporating facial contours and facial components and their configuration, and 2) component-level features focusing on the description of a single facial component. Thus, MDML-DCPs comprehensively encodes multi-level invariant characteristics of face images.

Both the DCP descriptor and the MDML-DCPs scheme are extensively evaluated on four large-scale databases: FERET [26], CAS-PEAL-R1 [27], FRGC 2.0 [28], and LFW [1]. The proposed DCP descriptor consistently achieves superior performance for both face identification and face verification tasks. More impressively, the proposed MDML-DCPs exploits only a single descriptor but achieves the best performance on two challenging unconstrained databases, FRGC 2.0 and LFW. Besides, this paper provides a fair and systematic comparison between state-of-the-art facial descriptors, which has been rarely performed in the face recognition field [29].

The remainder of the paper is organized as follows: Section 2 details the DCP descriptor and the construction of the MDML-DCPs face representation scheme is discussed in Section 3. Face recognition pipelines using MDML-DCPs are introduced in Section 4. Experimental results are presented in Section 5, leading to conclusions in Section 6.

2 DUAL-CROSS PATTERNS

The design of a face image descriptor consists of three main parts: image filtering, local sampling, and pattern encoding. The implementation of image filtering is flexible: possible methods include Gabor wavelets [7], Difference of Gaussian (DoG), or the recently proposed discriminative image filter [18]. In this paper, we focus on local sampling and pattern encoding, which are the core components of a face image descriptor.

2.1 Local Sampling

The essence of DCP is to perform local sampling and pattern encoding in the most informative directions contained within face images. For face recognition, useful face image information consists of two parts: the configuration of facial components and the shape of each facial
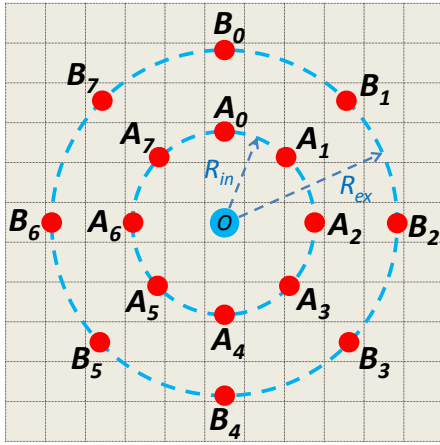
Fig. 1. Local sampling of Dual-Cross Patterns. Sixteen points are sampled around the central pixel O. The sampled points A_0 to A_7 are uniformly spaced on an inner circle of radius R_{in}, while B_0 to B_7 are evenly distributed on the exterior circle with radius R_{ex}.

component. The shape of facial components is, in fact, rather regular. After geometric normalization of the face image, the central parts of several facial components, i.e., the eyebrows, eyes, nose, and mouth, extend either horizontally or vertically, while their ends converge in approximately diagonal directions ($\pi/4$ and $3\pi/4$). In addition, wrinkles in the forehead lie flat, while those in the cheeks are either raised or inclined.

Based on the above observations, local sampling of DCP is conducted as shown in Fig. 1. For each pixel O in the image, we symmetrically sample in the local neighborhood in the 0, $\pi/4$, $\pi/2$, $3\pi/4$, π, $5\pi/4$, $3\pi/2$, and $7\pi/4$ directions, which are sufficient to summarize the extension directions of major facial textures. Two pixels are sampled in each direction. The resulting sampled points are denoted as $\{A_0, B_0; A_1, B_1; \cdots; A_7, B_7\}$. As illustrated in Fig. 1, $A_0, A_1, \cdots, A_7$ are uniformly spaced on an inner circle of radius R_{in}, while $B_0, B_1 \cdots, B_7$ are evenly distributed on the exterior circle with radius R_{ex}.

2.2 Pattern Encoding

Encoding of the sampled points is realized in two steps. First, textural information in each of the eight directions is independently encoded. Second, patterns in all eight directions are combined to form the DCP codes.

To quantize the textural information in each sampling direction, we assign each a unique decimal number:

$$DCP_i = S\left(I_{A_i} - I_O\right) \times 2 + S\left(I_{B_i} - I_{A_i}\right), 0 \le i \le 7, \quad (1)$$

where

$$S\left(x\right) = \begin{cases} 1, x \ge 0 \\ 0, x < 0 \end{cases}, \quad (2)$$

and I_O, I_{A_i}, and I_{B_i} are the gray value of points O, A_i, and B_i, respectively. Therefore, four patterns are defined to encode the second-order statistics in each direction

and each of the four patterns denotes one type of textural structure.

By simultaneously considering all eight directions, the total number of DCP codes is $4^8 = 65536$. This number is too large for practical face recognition applications; therefore, we adopt the following strategy. The eight directions are grouped into two subsets and each subset is further formulated as an encoder. In this way, the total number of local patterns is reduced to $4^4 \times 2 = 512$, which is computationally efficient. Although this strategy results in information loss, compactness and robustness of the descriptor are promoted. In the following subsection, we define the optimal grouping mode.

2.3 Dual-Cross Grouping

The grouping strategy introduced in the previous subsection produces $\frac{1}{2}\binom{8}{4} = 35$ combinations in total to partition all eight directions. To minimize information loss, we look for the optimal combination from the perspective of maximum joint Shannon entropy.

With the above analysis, $DCP_i \left(0 \le i \le 7\right)$ are discrete variables with four possible values: 0, 1, 2, and 3. Without loss of generality, the joint Shannon entropy for the subset $\{DCP_0, DCP_1, DCP_2, DCP_3\}$ is represented as

$$
\begin{aligned}
&H\left(DCP_0, DCP_1, DCP_2, DCP_3\right) \\
&= -\sum_{dcp_0} \cdots \sum_{dcp_3} P\left(dcp_0, \cdots, dcp_3\right) \log_2 P\left(dcp_0, \cdots, dcp_3\right),
\end{aligned}
$$
(3)

where dcp_0, dcp_1, dcp_2, and dcp_3 are particular values of DCP_0, DCP_1, DCP_2, and DCP_3, respectively. And $P\left(dcp_0, \cdots, dcp_3\right)$ is the probability of these values occurring simultaneously. The maximum joint Shannon entropy of the four variables is achieved when they are statistically independent.

In real images, the more sparsely the pixels are scattered, the more independent they are. Therefore, the maximum joint Shannon entropy for each subset is achieved when the distance between the sampled points is at its maximum. As a result, we define $\{DCP_0, DCP_2, DCP_4, DCP_6\}$ as the first subset and $\{DCP_1, DCP_3, DCP_5, DCP_7\}$ as the second subset. The resulting two subsets are illustrated in Fig. 2. Since each of the two subsets constructs the shape of a cross, the proposed descriptor is named Dual-Cross Patterns. In Section 5.1, we empirically validate that dual-cross grouping achieves the maximum joint Shannon entropy among all grouping modes on the FERET database.

2.4 DCP Face Image Descriptor

We name the two cross encoders DCP-1 and DCP-2, respectively. The codes produced by the two encoders at each pixel O are represented as

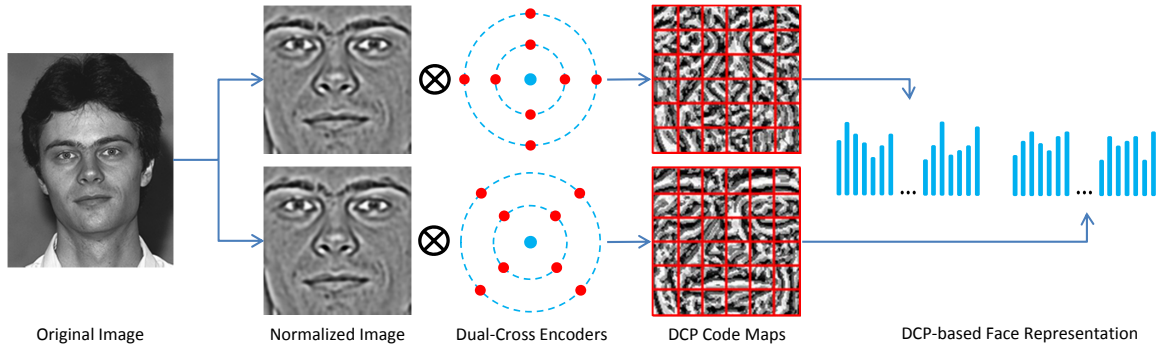$$DCP\text{-}1 = \sum_{i=0}^{3} DCP_{2i} \times 4^i, \quad (4)$$

Fig. 2. Face representation using Dual-Cross Patterns. The normalized face image is encoded by the two cross encoders, respectively. Concatenation of the regional DCP code histograms forms the DCP-based face representation.

$$DCP\text{-}2 = \sum_{i=0}^{3} DCP_{2i+1} \times 4^i. \qquad (5)$$

The DCP descriptor for each pixel O in an image is the concatenation of the two codes generated by the two cross encoders:

$$DCP = \left\{ \sum_{i=0}^{3} DCP_{2i} \times 4^i, \sum_{i=0}^{3} DCP_{2i+1} \times 4^i \right\}. \qquad (6)$$

After encoding each pixel in the face image using the dual-cross encoders, two code maps are produced that are respectively divided into a grid of non-overlapping regions. Histograms of DCP codes are computed in each region and all histograms are concatenated to form the holistic face representation. The overall framework of the above face representation approach is illustrated in Fig. 2. This face representation can be directly used to measure the similarity between a pair of face images using metrics such as the chi-squared distance or histogram intersection. The computation of the DCP descriptor, which doubles the feature size of LBP, is very efficient by only doubling the time cost of LBP.

We notice that two recently proposed descriptors DFD [18] and Center Symmetric-Pairs of Pixels (CCS-POP) [30] adopt similar sampling modes to DCP. However, they are essentially different from DCP. First, there is no clear motivation for the two descriptors that why such a sampling mode is suitable for face images. Second, the pattern encoding strategies are different: both the two descriptors rely on learning algorithms to handle the large sampling size problem mentioned in Section 1.

3 MULTI-DIRECTIONAL MULTI-LEVEL DUAL-CROSS PATTERNS

We now present a face representation scheme based on DCP named Multi-Directional Multi-Level Dual-Cross Patterns (MDML-DCPs) to explicitly handle the challenges encountered in unconstrained face recognition.

3.1 The MDML-DCPs Scheme

A major difficulty in unconstrained face recognition is that many factors produce significant intra-personal differences in the appearance of face images, in particular variations in illumination, image blur, occlusion, and pose and expression changes. We mitigate the influence of these factors using multi-directional gradient filtering and multi-level face representation.

In MDML-DCPs, the first derivative of Gaussian operator (FDG) is exploited to convert a gray-scale face image into multi-directional gradient images that are more robust to variations in illumination. The FDG gradient filter of orientation θ can be expressed as follows:

$$FDG\left(\theta\right) = \frac{\partial G}{\partial \mathbf{n}} = \mathbf{n} \cdot \nabla G, \qquad (7)$$

where $\mathbf{n} = (cos\theta, sin\theta)$ is the normal vector standing for the filtering direction and $G = \exp\left(-\frac{x^2+y^2}{\sigma^2}\right)$ is a two-dimensional Gaussian filter. The application of FDG is inspired by the classical work of Canny [31], which proved that FDG is the optimal gradient filter according to three criteria, namely signal-to-noise ratio (SNR) maximization, edge location accuracy preservation, and single response to single edge. These three criteria are also relevant for face recognition, where it is desirable to enhance the facial textural information while suppressing noise. FDG significantly saves computational cost compared to other gradient-like filters, such as Gabor wavelets [7]. We denote the concatenation of DCP descriptors extracted from the FDG-filtered images as MD-DCPs.

To build pose-robust face representation, MDML-DCPs normalizes the face image by two geometric rectifications based on a similarity transformation and an affine transformation, respectively. The similarity transformation retains the original information of facial contours and facial components and their configuration. The affine transformation reduces differences in intra-personal appearance caused by pose variation.

MDML-DCPs combines both holistic-level and component-level features, which are computed on the normalized images by the two transformations. Holistic-level features capture comprehensive information on both facial components and facial contour. However, it is also sensitive to changes in appearance of each component caused by occlusion, pose, and variations in expression. In contrast, component-level features
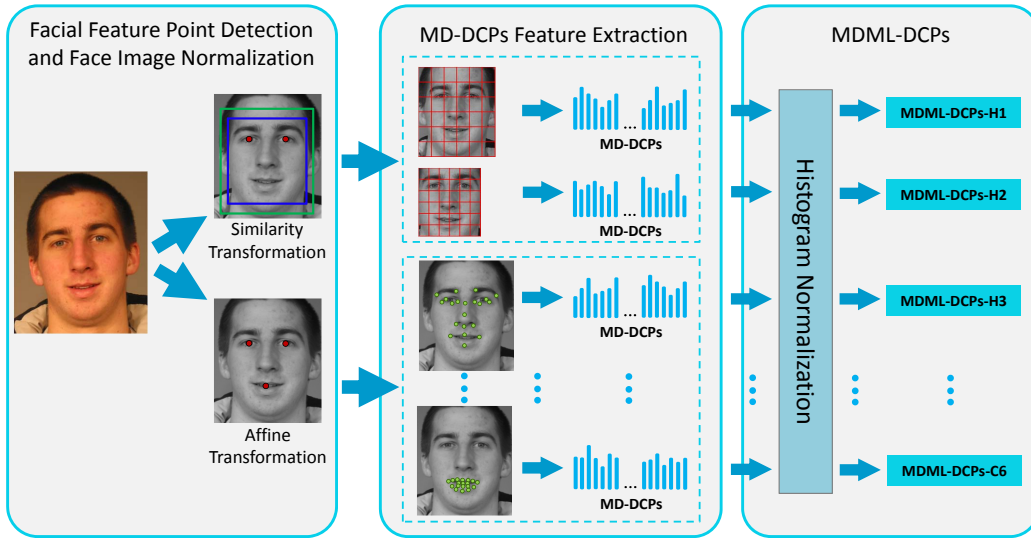
Fig. 3. Framework of the MDML-DCPs face representation scheme. MDML-DCPs-H1 and MDML-DCPs-H2 are extracted from the rectified image by similarity transformation. MDML-DCPs-H3, MDML-DCPs-C1 to C6 are extracted from the affine-transformed image. The MDML-DCPs face representation is the set of the above nine feature vectors.
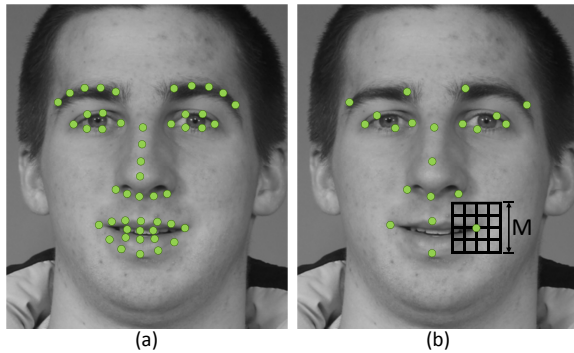

Fig. 4. (a) The 49 facial feature points detected by the face alignment algorithm. (b) MDML-DCPs-H3 employs 21 facial feature points over all facial components. MDML-DCPs-C1 to C6 respectively select 10 facial feature points on both eyebrows, 12 points on both eyes, 11 points on the left eye and left eyebrow, 11 points on the right eye and right eyebrow, 9 points on nose, and 18 points on mouth. Around each facial feature point, MD-DCPs are extracted from $J \times J$ (in this figure, $J = 4$) non-overlapping regions within the patch of size $M \times M$ pixels.

focus on the description of a single facial component, and thus are independent of changes in appearance of the other components. In this way the information generated by these two feature levels is complementary and appropriate fusion of the two promotes robustness to interference.

3.2 Implementation Details

Similar to most state-of-the-art face representation schemes, e.g., [25], [9], MDML-DCPs also benefits from recent progress in face alignment algorithms that accurately locate dense facial feature points in real-time.

The MDML-DCPs face representation scheme is shown in Fig. 3. In MDML-DCPs, a high performance face alignment algorithm based on [24] is first employed to locate 49 facial feature points (as shown in Fig. 4a), before applying the two geometric rectifications.

The similarity transformation is based on the two detected eye centers. In the rectified image, we compute MD-DCPs at two holistic levels: 1) MD-DCPs of non-overlapping regions of the external cropped face image (including facial contour), and 2) MD-DCPs of non-overlapping regions of the internal cropped face image (without facial contour). The first feature encodes both facial contour and facial components while the second feature focuses on encoding facial components only, which are free from background interference. For clarity, we denote the two features as MDML-DCPs-H1 and MDML-DCPs-H2.

The affine transformation is determined by the three detected facial feature points: the centers of the two eyes and the center of the mouth. In the rectified image, one holistic-level feature denoted as MDML-DCPs-H3, and six component-level features referred to MDML-DCPs-C1 to MDML-DCPs-C6, are computed based on the detected dense facial feature points. As shown in Fig. 4b, the method for feature extraction around each facial feature point is similar to the approaches used in [32], [25]: centered on each facial feature point, a patch of size $M \times M$ pixels is located and further divided into $J \times J$ non-overlapping regions. The concatenated MD-DCPs feature of the J^2 regions forms the description of the feature point.

MDML-DCPs-H3 and MDML-DCPs-C1 to C6 are formed by respectively concatenating the descriptions of different facial feature points. As shown in Fig. 4, MDML-DCPs-H3 selects 21 facial feature points over all

facial components, while MDML-DCPs-C1 to C6 select feature points on only one particular facial component. Elements of the three holistic-level features and six component-level features are normalized by the square root. Together, the set of the nine normalized feature vectors form the MDML-DCPs face representation.

4 FACE RECOGNITION ALGORITHM

In this section, the face matching problem is addressed using the proposed MDML-DCPs face representation scheme. First, one classifier is built for each of the nine feature vectors. Then, the similarity scores of the nine classifiers are fused by linear SVM or simple averaging.

Two algorithms are considered: Whitened Principal Component Analysis (WPCA; an unsupervised learning algorithm) and Probabilistic Linear Discriminant Analysis (PLDA; a supervised learning algorithm) [33], [34]. The choice of which of the two algorithms to use is dataset dependent: for datasets that have a training set with multiple face images for each subject, we choose PLDA; otherwise, WPCA is used. For more recent high-performance classifiers, readers can refer to [35], [36].

4.1 WPCA

Principal Component Analysis (PCA) learns an orthogonal projection matrix U from training data and projects high-dimensional feature vector x to the low-dimensional vector y,

$$y = U^T x. \tag{8}$$

The columns of U are composed of the leading eigenvectors of the covariance matrix of the training data. However, the first few eigenvectors in U encode mostly variations in illumination and expression, rather than information that discriminates identity. The whitening transformation tackles this problem by normalizing the contribution of each principal component

$$y = \left(U\Lambda^{-1/2}\right)^T x, \tag{9}$$

where $\Lambda = diag\{\lambda_1, \lambda_2, \cdots\}$ with λ_i being the ith leading eigenvalue. After projecting the facial feature vectors to the low-dimensional subspace using WPCA, the similarity score between two feature vectors y_1 and y_2 is measured by the cosine metric

$$Sim(y_1, y_2) = \frac{y_1^T y_2}{\|y_1\| \|y_2\|}. \tag{10}$$

4.2 PCA combined with PLDA

The feature vectors in MDML-DCPs are high-dimensional. To effectively apply PLDA, the dimensionality of the nine feature vectors is first reduced by PCA.

PLDA models the face data generation process

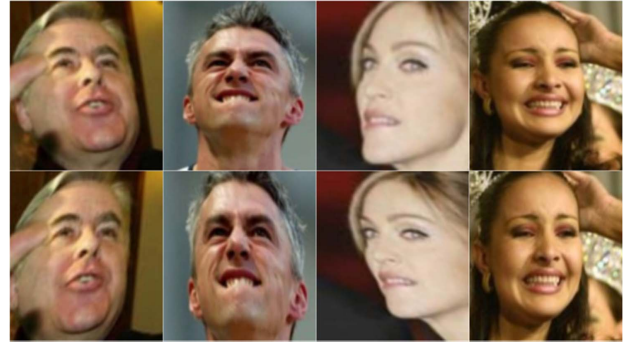$$x_{ij} = \mu + Fh_i + Gw_{ij} + \varepsilon_{ij}, \tag{11}$$

Fig. 6. Sample images from LFW. Images in the two rows are aligned by a similarity transformation and an affine transformation, respectively.

where x_{ij} denotes the jth face data of the ith individual. μ is the mean of all face data. F and G are factor matrices whose columns are the basis vectors of the between-individual subspace and the within-individual subspace, respectively. h_i is the latent identity variable that is constant for all images of the ith subject. w_{ij} and ε_{ij} are noise terms explaining intra-personal variance.

It is shown in [33] that the identification and verification problems can be solved by computing the log-likelihood ratio that whether two observed images share the same identity variable h or not. In this paper, we refer to the log-likelihood ratio as similarity score for consistency with the case of the WPCA classifier.

5 EXPERIMENTS

In this section, the proposed DCP and MDML-DCPs are extensively evaluated in both face identification and face verification tasks. Experiments are conducted on four publicly available large-scale face databases: FERET, CAS-PEAL-R1, FRGC 2.0, and LFW. Example images of the four databases are shown in Figs. 5 and 6.

The FERET database contains one gallery set Fa and four probe sets, i.e., Fb, Fc, Dup1, and Dup2. In this paper, the standard face identification protocol specified in [26] is employed.

The CAS-PEAL-R1 database includes one training set, one gallery set, and nine probe sets. Each of the nine probe sets is restricted to one type of variations. In detail, the PE, PA, PL, PT, PB, and PS probe sets correspond to variations in expression, accessory, lighting, time, background, and distance of frontal faces, respectively. The PU, PM, and PD probe sets correspond to one type of pose variation. The standard identification protocol defined in [27] is followed.

The FRGC 2.0 database incorporates data for six face verification experiments [28]. We focus on Experiments 1 and 4. The shared target set of Experiments 1 and 4 is collected from frontal facial images taken under controlled illumination, while images in their query sets are captured under controlled and uncontrolled conditions, respectively. Verification rates at 0.1% FAR are reported.
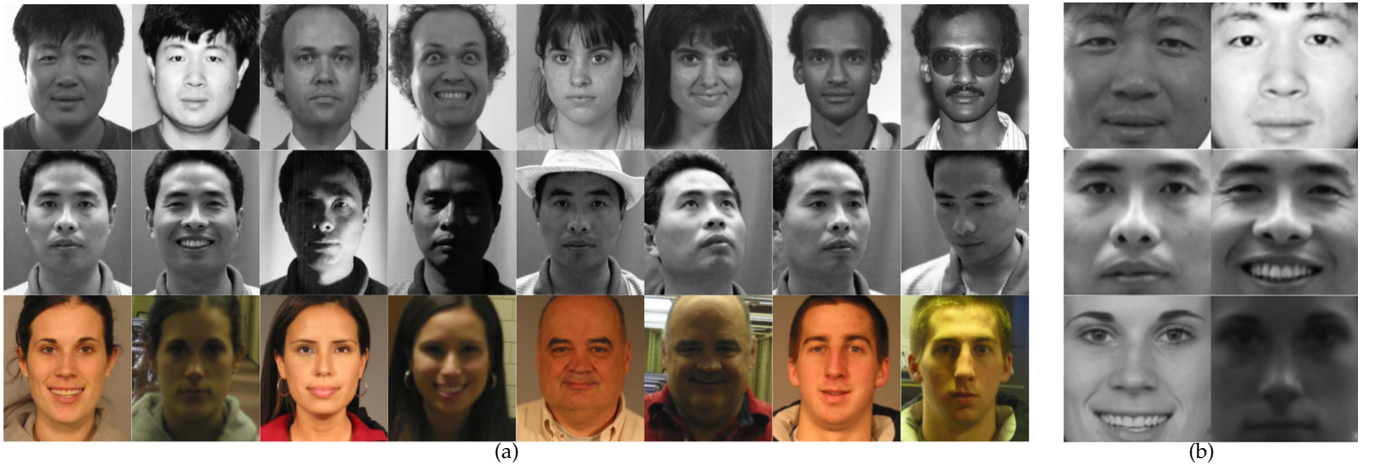
(a)

(b)

Fig. 5. (a) Sample images from FERET (first row), CAS-PEAL-R1 (second row) and FRGC 2.0 (third row) containing typical variations in each database. (b) Samples of normalized images for experiments from Section 5.1 to 5.3.

The LFW database [1] contains 13,233 unconstrained face images that are organized into two "Views". View 1 is for model selection and parameter tuning while View 2 is for performance reporting. Two paradigms are used to exploit the training set in View 2, an image-restricted paradigm and an image-unrestricted paradigm. In the first paradigm, only the officially defined image pairs are available for training. In the second paradigm, the identity information of the training images can be used. We report the mean verification accuracy and standard error of the mean (S_E) on the View 2 data.

Four experiments are conducted. First, the dual-cross grouping mode for DCP is empirically proved to achieve the maximum joint Shannon entropy. Second, the performance of DCP is compared with eleven state-of-the-art face image descriptors. Third, the performance of MD-DCPs is evaluated to determine the contribution of multi-directional filtering by FDG. Finally, the power of the MDML-DCPs face representation scheme is presented. More experimental results, e.g., parameter selection of DCP, are available in the appendix.

In the first three experiments, we focus on the evaluation of face image descriptors. In these three experiments, all face images are cropped and resized to 128×128 ($rows \times columns$) pixels with the eye centers located at $(34, 31)$ and $(34, 98)$, as shown in Fig. 5b. For LFW, we use the aligned version (LFW-A) [19], while for the other three databases, images are cropped according to the eye coordinates provided by the databases. The cropped images are photometrically normalized using a simple operator (denoted as TT) developed by Tan & Triggs [11] and then encoded by one face image descriptor. Each encoded face image is divided into $N \times N$ non-overlapping regions. Concatenating the code histograms of these regions forms the representation for the image.

In the fourth experiment, we aim to demonstrate that the proposed MDML-DCPs face representation scheme
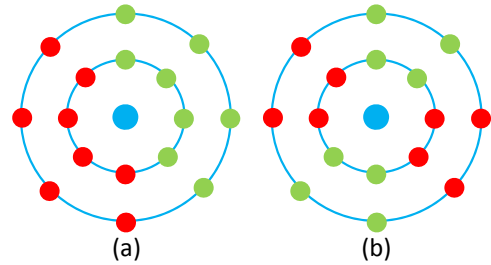

(a)

(b)

Fig. 7. Another two representative grouping modes for the eight sampling directions of DCP. Sampled points of the same colour belong to the same subset.

has excellent performance characteristics. In this experiment, all face images are normalized by two geometric transformations, as illustrated in Fig. 6. In both transformations, face images are resampled to 180×162 pixels with the eye centers mapped to $(66, 59)$ and $(66, 103)$. For the affine-transformed face images, the centers of the mouths are unified to $(116, 81)$. TT is applied to the resampled images for photometric normalization.

5.1 Empirical Justification for Dual-Cross Grouping

Section 2.3 intuitively suggests that the dual-cross grouping mode is optimal from the perspective of the joint Shannon entropy maximization. In this experiment, we empirically validate this point on the FERET database. For each of the 35 possible grouping modes, $DCP_i \, (0 \leq i \leq 7)$ are divided into two subsets. Then, the joint Shannon entropy for each of the two subsets are calculated on one image using (3) and are summed together. The above process is repeated on the 1,196 gallery images of the FERET database. The mean value of the summed joint Shannon entropy is recorded.

Experimental results show that dual-cross grouping mode achieves the highest joint Shannon entropy among all 35 grouping modes. Fig. 8 characterizes the joint Shannon entropy as a function of R_{in} and R_{ex} and
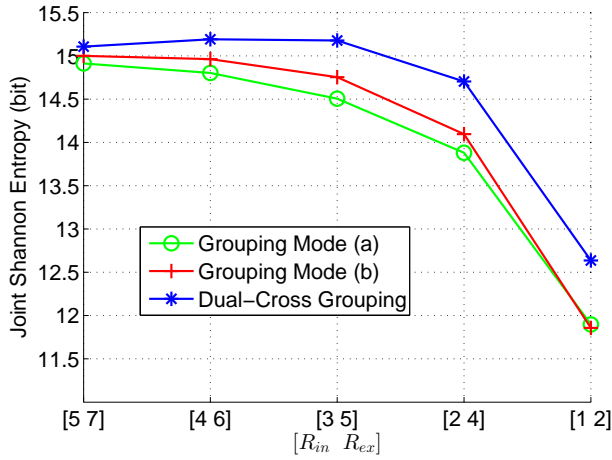
Fig. 8. Joint Shannon entropy as a function of R_{in} and R_{ex}. Three grouping modes are evaluated in this figure: modes (a) and (b) in Fig. 7 and the dual-cross grouping.

illustrate the superiority of dual-cross grouping mode by comparing it with two representative grouping modes, exampled in Fig. 7. Note that the joint Shannon entropy is related to the sampling radii of DCP: smaller sampling radii mean stronger dependence among the sampled points, which results in a smaller joint Shannon entropy. The dual-cross grouping mode achieves the highest entropy under all sets of the radii values. Therefore, the dual-cross grouping mode is empirically optimal.

5.2 Evaluation of the Performance of DCP

In this section, the performance of DCP, DCP-1, and DCP-2 for both face identification and face verification tasks are evaluated. To better illustrate the advantages of DCP, the performance of eleven state-of-the-art face image descriptors, i.e., LBP [5], LTP [11], LPQ [15], POEM [16], Local Gabor XOR Patterns (LGXP) [7], Multi-scale LBP (MsLBP), Multi-scale tLBP (MsTLBP) [12], Multi-scale dLBP (MsDLBP) [12], LQP [37], DFD [18], and CHG [30] are also presented. [1]

The first eight descriptors are hand-crafted and they are easy to implement. We therefore test them together with DCP. It is worth noting that the cropped face image data for these descriptors are exactly the same. All of them extract features from photometrically normalized images by TT. The parameters for each of the descriptors and TT are carefully tuned on each database and different distance metrics (chi-squared or histogram intersection) are tested. Finally, the best results for each descriptor are reported. Therefore, the experimental results directly compare the recognition capabilities of the descriptors.

On the other hand, LQP, DFD, and CHG are learning-based descriptors, which are complicated to implement.

1. For Table 2 to Table 6, a superscript * means that the results are cited from the original papers. A suffix '-Flip' means that the descriptor adopts the 'flip' trick [37].

TABLE 1
Feature Size of the Investigated Face Image Descriptors

Descriptor	Feature Size	Descriptor	Feature Size
LBP [5]	256	MsDLBP [12]	512
MsLBP [38]	512	LQP [37]	300
LTP [11]	512	DFD [18]	1024
LPQ [15]	256	DCP-1	256
POEM [16]	1024	DCP-2	256
LGXP [7]	640	DCP	512
MsTLBP [12]	512		

In this paper, their performance on FERET and LFW is directly cited from the original papers, even though the experimental settings are different from that of DCP. For LQP, we cite its performance with the image filtering step by Gabor wavelets, which is more robust to illumination variation. Before presenting the detailed experimental results, the feature size (number of histogram bins) for each descriptor except CHG (CHG is not a histogram-based descriptor) is listed in Table 1. In the following experiments, LBP-based descriptors are implemented without using uniform coding [38] (using uniform coding actually degrades the performance of LBP, LTP, and POEM, et al.). For POEM, we compute LBP codes on four-orientation gradient magnitude maps, so its feature size is 1024 in this paper.

5.2.1 Face Identification: FERET

Face identification experiments are conducted on the FERET and CAS-PEAL-R1 databases. The rank-1 identification rates on the four probe sets of FERET are presented in Table 2. We make four observations:

1) While high performance is achieved by all descriptors on the well controlled Fb and Fc probe sets, DCP still outperforms MsLBP by over 1%. In particular, DCP achieves a perfect identification rate on the Fc set.

2) There is a substantial performance drop for all descriptors on Dup1 and Dup2 probe sets, in which images contain moderate expression and illumination variations. DCP performs best on both sets, with a margin of 1.28% on Dup2 by comparing with the second best descriptor. The performance of MsTLBP ranks second on Dup1 and LGXP ranks second on Dup2. It is worth pointing out that LGXP depends on 80 expensive convolutions and produces a larger feature size, while both DCP and MsTLBP have low computational complexity.

3) Both DCP-1 and DCP-2 perform better than most of the other descriptors at lower computational cost. Their time and memory costs are exactly the same as those of LBP, suggesting the sampling and encoding strategies of DCP are highly effective.

4) As expected, the performance of DCP is better than both DCP-1 and DCP-2, which means that DCP-1 and DCP-2 contain complementary information.

TABLE 2
Identification Rates for Different Descriptors on FERET

	Fb	Fc	Dup1	Dup2	Mean
LBP	96.90	98.45	83.93	82.48	90.44
LTP	96.90	98.97	83.93	83.76	90.89
LPQ	97.41	99.48	82.69	81.62	90.30
POEM	98.24	99.48	82.83	82.05	90.65
LGXP	97.32	99.48	85.46	85.47	91.93
MsLBP	97.07	98.97	83.38	83.33	90.69
MsTLBP	98.16	99.48	85.73	85.04	92.10
MsDLBP	95.65	97.94	79.09	79.49	88.04
LQP-Flip*	**99.50**	99.50	81.20	79.90	90.03
DFD*	99.20	98.50	85.00	82.90	91.40
CHG*	97.50	98.50	85.60	84.60	91.55
DCP-1	97.91	98.45	84.49	84.19	91.26
DCP-2	97.99	99.48	84.35	85.04	91.72
DCP	98.16	**100.0**	**86.29**	**86.75**	**92.80**

When DCP-1 and DCP-2 are combined, the mean identification rate increases by over 1%.

5.2.2 Face Identification: CAS-PEAL-R1

Identification results on the nine probe sets of CAS-PEAL-R1 are shown in Table 3. For each descriptor, results are reported with the set of parameters that achieves the highest mean identification rate over all nine probe sets. We make the following observations:

1) In general, the performance of all the descriptors is good on the PE, PA, PT, PB, and PS probe sets, with DCP producing the highest mean identification rate and MsTLBP the second highest. The performance of all descriptors is poor on the remaining PL, PU, PM, and PD probe sets due to serious illumination and pose variations.

2) DCP outperforms MsLBP and LTP by 2.5% and 3.08% on the PL set, respectively. However, the performance of DCP is lower than LPQ, POEM, and LGXP. The PL images contain not only rich illumination variation, but also serious image blur, which explains the superior performance of the blur-invariant descriptor LPQ. LGXP and POEM benefit from image filtering steps that turn the gray-scale image into gradient images. We show in Section 5.3 that by introducing multi-directional gradient filtering by FDG, the performance of DCP is significantly improved.

3) DCP exhibits excellent robustness to pose variations on the PU, PM, and PD probe sets, with superior mean recognition rate by as much as 3.44%. This result is encouraging, since pose variation is a major challenge in unconstrained face recognition [3].

5.2.3 Face Verification: FRGC 2.0

Face verification experiments are conducted on the FRGC 2.0 and LFW databases. Verification rates at 0.1%

TABLE 4
Verification Results on the FRGC 2.0 Experiment 1

	ROC I	ROC II	ROC III	Mean
LBP	89.93	87.53	84.83	87.43
LTP	90.98	88.63	86.07	88.56
LPQ	89.96	87.38	84.48	87.27
POEM	90.60	88.26	85.63	88.16
LGXP	91.89	88.80	85.33	88.67
MsLBP	90.78	88.45	85.87	88.37
MsTLBP	**93.24**	91.16	88.83	91.08
MsDLBP	87.65	84.93	81.96	84.85
DCP-1	91.40	89.14	86.65	89.06
DCP-2	92.96	90.78	88.42	90.72
DCP	93.22	**91.21**	**88.93**	**91.12**

TABLE 5
Verification Results on the FRGC 2.0 Experiment 4

	ROC I	ROC II	ROC III	Mean
LBP	18.62	19.12	20.07	19.27
LTP	20.21	21.39	22.99	21.53
LPQ	22.53	23.25	24.39	23.39
POEM	29.99	30.60	**31.46**	30.68
LGXP	**32.44**	**31.81**	31.13	**31.79**
MsLBP	19.34	19.80	20.81	19.98
MsTLBP	20.04	20.25	21.02	20.44
MsDLBP	18.34	18.85	19.60	18.93
DCP-1	20.10	20.57	21.51	20.73
DCP-2	18.22	18.74	19.58	18.85
DCP	21.78	22.49	23.59	22.62

FAR on Experiments 1 and 4 of FRGC 2.0 are listed in Tables 4 and 5, respectively. As mentioned above, the query sets of the two experiments are composed of controlled and uncontrolled images, respectively. The query images in Experiment 4 are degraded by serious image blur and significant illumination variation, making this dataset very challenging. We make the following observations:

1) In Experiment 1, DCP shows the best performance. Both DCP-1 and DCP-2 achieve better performance than most of the existing descriptors. These observations are consistent with the results on FERET.

2) In Experiment 4, the mean verification rate of DCP is higher than MsLBP and LTP by 2.64% and 1.09%, respectively. Due to the lack of an image filtering step in DCP, its performance is lower than that of LGXP and POEM. This result is consistent with that seen for the PL set of CAS-PEAL-R1, whose images also feature serious illumination variation and image blur.

5.2.4 Face Verification: LFW

LFW is a very challenging dataset since its images contain large pose, expression, and illumination variations. Experiments in this section follow the image-restricted

TABLE 3
Rank-1 Identification Rates for Different Face Image Descriptors on the Nine Probe Sets of PEAL

	PE	PA	PT	PB	PS	PL	PU	PM	PD	Mean (PE-PS)	Mean (PE-PL)	Mean (PU-PD)
LBP	94.27	91.82	100.0	**99.46**	**99.64**	46.90	60.32	83.66	44.60	97.04	88.68	62.86
LTP	94.39	91.77	100.0	**99.46**	**99.64**	47.17	61.32	84.46	44.68	97.05	88.74	63.49
LPQ	93.95	92.39	100.0	99.28	99.27	57.16	57.78	86.46	44.76	96.98	90.34	63.00
POEM	95.54	92.39	100.0	**99.46**	**99.64**	54.66	58.14	85.12	42.52	97.41	90.28	61.93
LGXP	94.97	91.33	100.0	99.28	**99.64**	**63.26**	39.46	66.83	22.91	97.04	**91.41**	43.07
MsLBP	95.16	92.04	100.0	**99.46**	**99.64**	47.75	61.76	85.16	44.88	97.26	89.01	63.93
MsTLBP	95.41	92.74	100.0	**99.46**	**99.64**	48.06	60.98	87.34	45.48	97.45	89.22	64.60
MsDLBP	92.42	90.63	100.0	99.28	99.27	48.11	55.00	82.03	37.03	96.32	88.29	58.02
DFD*	**99.30**	**94.40**	-	-	-	59.00	-	-	-	-	-	-
DCP-1	95.99	92.60	100.0	98.73	99.27	46.37	60.68	85.74	48.14	97.32	88.83	64.85
DCP-2	95.54	91.90	100.0	98.92	**99.64**	43.91	60.64	83.08	43.62	97.20	88.32	62.45
DCP	96.11	92.82	100.0	99.10	**99.64**	50.25	**65.39**	**87.44**	**51.30**	**97.53**	89.65	**68.04**

TABLE 6
Mean Verification Accuracy on the LFW View 2 Data

	Accuracy(%)$\pm S_E$		Accuracy(%)$\pm S_E$
LBP	72.43 ± 0.49	MsDLBP	72.17 ± 0.59
LTP	72.65 ± 0.52	LQP-Flip*	75.30 ± 0.26
LPQ	72.68 ± 0.46	**DFD-Flip***	**80.02 ± 0.50**
POEM	73.98 ± 0.56	DCP-1	74.50 ± 0.48
LGXP	70.58 ± 0.43	DCP-2	73.28 ± 0.48
MsLBP	72.88 ± 0.50	DCP	75.00 ± 0.64
MsTLBP	74.12 ± 0.57	DCP-Flip	76.37 ± 0.71

Fig. 9. Performance comparison between DCP and MsLBP on the four face datasets.

paradigm. For each descriptor, its parameters are tuned on the View 1 data, and performance on the View 2 data is reported in Table 6. We observe that:

1) Among the manually-designed descriptors, DCP achieves the best performance while DCP-1 ranks second. This result is consistent with the results on FERET and Experiment 1 on FRGC 2.0. The robustness of DCP to pose variations is consistent with the results on CAS-PEAL-R1.

2) LGXP does not perform well on LFW because LFW images contain significant pose variations. Although LGXP is robust to serious illumination variations, it is sensitive to pose variations, consistent with the results seen from experimentation on the non-frontal probe sets of CAS-PEAL-R1.

3) Compared with the two learning-based descriptors, the performance of DCP is better than that of LQP, but is worse than that of DFD. However, DCP has clear advantage in both time and memory costs. Besides, DFD adopts supervised learning algorithms to enhance its discriminative power, while DCP is independent of any learning algorithms. We show in Section 5.4 that by applying supervised learning algorithms to the extracted DCP feature, superior performance can be achieved.
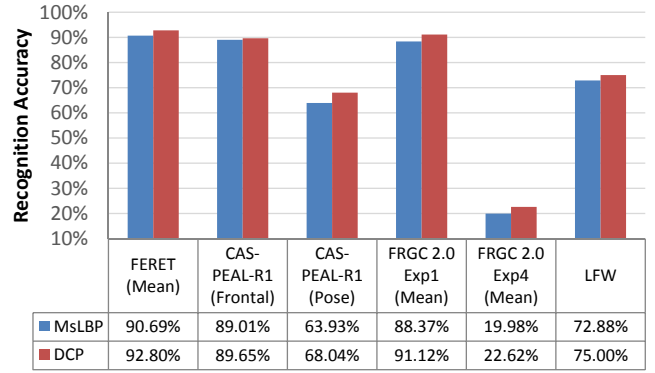
5.2.5 Discussion

The experimental results in Section 5.2 reveal that DCP performs quite well across a range of evaluations. In particular, DCP significantly outperforms MsLBP, which is of exactly the same time and memory costs as DCP, as shown in Fig. 9. The performance of MsTLBP is usually slightly worse than DCP on well-controlled datasets. However, DCP considerably outperforms MsTLBP on the challenging PU and PD probe sets of CAS-PEAL-R1, and the Dup2 probe set of FERET. This indicates DCP is a more robust descriptor to pose and aging factors.

The excellent performance of DCP can be explained as follows. It has large sampling size to encode more discriminative information; it encodes the second-order statistics in the most informative directions on human face; and the dual-cross grouping strategy ensures good complementarity between the divided two encoders.

5.3 The Contribution of Multi-directional Filtering

In Section 5.2, DCP is shown to have strong discriminative power and excellent robustness to expression, pose,

and moderate illumination variations. It has also been shown that when there is serious illumination variation and image blur, the performance of DCP is degraded due to the lack of an image filtering step. In this section, we show that MD-DCPs that extracts the DCP feature on FDG-filtered images is robust to illumination variation. The roles of FDG are two-fold: first, it converts the face image into gradient maps, which are more robust to illumination variation, and second, FDG is proven to achieve high SNR [31], which is vital for the performance of face image descriptors for low quality images.

Throughout this paper, the number of filtering orientations by FDG is fixed to 4 ($\theta = 0, \pi/4, \pi/2$, and $3\pi/4$) and the parameter σ of the Gaussian kernel is set at 1.0. Experiments are conducted on three datasets:

- A new FERET probe set is constructed by collecting the 146 most challenging FERET probe images. The FERET Fa set is still utilized as the gallery set. The identification rate of DCP on this probe set is only 17.12%. In contrast, the identification rate when MD-DCPs is used increases to 33.56%.
- For the PL probe set of CAS-PEAL-R1, the identification rates of DCP and MD-DCPs are 50.25% and 65.23%, respectively. When considered along with the previously best results shown in Table 3, it is evident that the performance of MD-DCPs is superior to that of LPQ (57.16%), POEM (54.66%), and LGXP (63.26%).
- For the face verification task on FRGC 2.0 Experiment 4, the mean verification rate of MD-DCPs at 0.1% FAR for ROC I, ROC II, and ROC III is 30.74%. Together with the previously best results shown in Table 5, it is clear that MD-DCPs outperforms LPQ (23.39%) and POEM (30.68%). LGXP outperforms MD-DCPs with the mean verification rate of 31.79%, but at the cost of 80 expensive convolution operations.

In conclusion, MD-DCPs consistently achieves excellent performance, which means that multi-directional filtering by FDG is indeed helpful for removing illumination variation and enhancing the robustness of face image descriptors on low SNR images.

5.4 Performance Evaluation of MDML-DCPs

Before presenting the performance of MDML-DCPs, we first unify most of its parameters for testing the four databases. For the holistic-level feature MDML-DCPs-H1, we compute it within the area whose top left corner is located at $(33, 27)$, while the bottom right corner is located at $(154, 136)$ in the normalized image by similarity transformation. The holistic-level feature MDML-DCPs-H2 is computed within the area where the two corners are located at $(36, 41)$ and $(140, 122)$, respectively. Both areas are divided into 9×9 non-overlapping regions. For MDML-DCPs-H3 and MDML-DCPs-C1 to C6, we consistently use a patch size $M \times M$ of 40×40 and the number of regions $J \times J$ within the patch to be 4×4.

TABLE 7
Identification Rates for Different Methods on FERET

	Fb	Fc	Dup1	Dup2
AMFG07' (L,G) + KDCV [40]	98.00	98.00	90.00	85.00
TIP10' LGBP + LGXP + LDA [7]	99.00	99.00	94.00	93.00
BMVC12' G-LQP + WPCA [37]	**99.90**	**100.0**	93.20	91.00
PAMI14' DFD + WPCA [18]	99.40	**100.0**	91.80	92.30
MDML-DCPs + WPCA	99.75	**100.0**	**96.12**	**95.73**

The two parameters σ_1 and σ_2 for TT are set to 1.4 and 2.0 on FERET, CAS-PEAL-R1, and FGRC 2.0. Since TT in fact slightly degrades the performance of MDML-DCPs on the LFW View 1 data, the photometric normalization step is omitted for the experiment on LFW. The subspace dimensions of WPCA and PCA for all nine features are set to 600. The dimensions of both the between-individual subspace and within-individual subspace of PLDA are set to 100. There are only two remaining sets of parameters to tune over the four databases: the first is the sampling radii R_{in} and R_{ex} of DCP, and the other is the cost parameter c of the linear SVM [39]. The optimal values of R_{in} and R_{ex} for FERET, CAS-PEAL-R1, and FGRC 2.0 are $R_{in} = 2$ and $R_{ex} = 3$. c is set at 10^{-4} for FRGC 2.0. For LFW, R_{in}, R_{ex}, and c are set to 4, 6, and 1.0, respectively, based on the results on the View 1 data.

5.4.1 Face Identification: FERET

As there is no officially defined training set for FERET, we test MDML-DCPs with the WPCA-based classifiers. The nine WPCA projection matrices are trained on the 1,196 FERET gallery images. For simplicity, the similarity scores computed by the nine classifiers are fused by averaging with equal weights. Performance comparison between MDML-DCPs with the state-of-the-art approaches is shown in Table 7. The first two approaches use supervised learning algorithms. The others utilize the unsupervised learning algorithm WPCA. The performance of MDML-DCPs is superior even to those that employ a supervised learning algorithm. In particular, MDML-DCPs achieves the best results on the Fc, Dup1, and Dup2 probe sets. On the Fb set, three images are misclassified by MDML-DCPs, two of which are due to the mislabeling of subjects in the database itself.

5.4.2 Face Identification: CAS-PEAL-R1

For experiments on CAS-PEAL-R1, we present MDML-DCPs results with both WPCA and PLDA, with DCP set as the baseline. Both WPCA and PLDA are trained on all 1,200 training images. In both approaches, the similarity scores are fused by simple averaging without weighting. A comparison between the performance of MDML-DCPs with other state-of-the-art methods is presented in Table 8. We make the following observations:

1) "MDML-DCPs + WPCA" achieves the best performance for eight of the nine probe sets of CAS-

TABLE 8
Rank-1 Identification Rates for Different Methods on the Nine Probe Sets of PEAL

	PE	PA	PL	PT	PB	PS	PU	PM	PD
TIP07' LGBPHS [14], [13]	95.20	86.80	51.00	100.0	98.70	98.90	-	-	-
TIP07' HGPP [13]	96.80	92.50	62.90	98.40	99.80	99.60	-	-	-
SP09' Weighted LLGP [41]	98.00	92.00	55.00	-	-	-	-	-	-
PAMI14' DFD + WPCA [18]	99.00	96.90	63.90	-	-	-	-	-	-
DCP + χ^2	96.11	92.82	50.25	100.0	99.10	99.64	65.39	87.44	51.30
MDML-DCPs + PLDA	98.22	97.51	63.26	100.0	**100.0**	100.0	34.31	70.24	33.59
MDML-DCPs + WPCA	**99.62**	**99.21**	**82.92**	**100.0**	99.82	**100.0**	**68.29**	**98.10**	**53.92**

PEAL-R1. In particular, it outperforms DCP by 32.67% on the challenging PL probe set.

2) Compared with DCP, there is limited performance promotion made by "MDML-DCPs + WPCA" on the PU and PD sets. The reason is that the training set contains only frontal images while the PU and PD images possess large pose variations. Therefore the two probe sets benefit little from training.

3) The identification rates by "MDML-DCPs + PLDA" are lower than "MDML-DCPs + WPCA", which suggests that there might be a gap between the distribution of training and testing data. In particular, the discriminative subspace learnt by PLDA on the frontal images does not generalize to the three non-frontal probe sets.

5.4.3 Face Verification: FRGC 2.0

The performance of MDML-DCPs on Experiments 1 and 4 of the FRGC 2.0 database is discussed. The FRGC 2.0 database provides a large training set, on which PLDA and linear SVM are trained. For each subject in the training set, 12 images are randomly selected and there are 2,664 images in total for SVM training, for which 13,320 intra-personal images pairs and 13,320 inter-personal image pairs are generated according to the method described in [1]. All the remaining images are used to train the PLDA model. The performance of the proposed algorithm, together with other state-of-the-art approaches, is shown in Table 9. We observe that:

1) The MDML-DCPs approach achieves superior performance on both Experiment 1 and 4 with the single descriptor DCP. Score fusion by linear SVM also works slightly better than score averaging.

2) On Experiment 1, the performance of MDML-DCPs is nearly perfect. On Experiment 4, it outperforms the current state-of-the-art method [21] by 0.9%. Moreover, MDML-DCPs utilizes only a single face image descriptor, while the method in [21] relies on color information and makes use of three descriptors: LBP, Gabor, and Fourier features.

Different from the approaches in Table 9, it is shown in [44] that by conducting normalization on the similarity matrices which contain similarity scores of each pair of target and query images, verification rates can be improved significantly. However, this operation may

not be suitable for the general face verification problems, where only a pair of target and query images is available each time. Therefore, we provide results of MDML-DCPs without score normalization in this paper.

5.4.4 Face Verification: LFW

Experiments in this section follow the image-unrestricted paradigm of the LFW database. Both the PLDA classifier and the more recently proposed Joint Bayesian (JB) classifier [35] are tested for face matching, and both linear SVM and score averaging are tested for the score fusion. For each of the 10-fold cross validations, the remaining nine subsets are partitioned into two: the first eight subsets are used for the training of PCA, PLDA, and JB models, while the last subset is used to train the linear SVM or learn the optimal threshold of similarity scores fused by averaging. According to the description in [1], 2,500 intra-personal image pairs and 2,500 inter-personal image pairs are generated using the images in the last subset. No outside training data are employed. In addition, following [19], [33], all images with right profile faces are flipped to the left profile faces.

A comparison of the classification accuracies is presented in Table 10. All the methods in Table 10 follow the image-unrestricted training paradigm and do not use outside training data. The ROC curves are plotted in Fig. 10. We make the following observations:

1) MDML-DCPs outperforms the current state-of-the-art method [25] by over 2.2%. It is worth noting that [25] also employs dense facial landmarks for face representation. MDML-DCPs outperforms [33], which also employs the PLDA-based classifier, by over 5%. The above facts suggest that the MDML-DCPs face representation scheme is more effective than previous approaches.

2) Score fusion by linear SVM results in better performance than using score averaging. Moreover, with either linear SVM or score averaging, MDML-DCPs is able to achieve better performance than existing methods, and using both fusion methods the standard error of classification accuracy is consistently smaller than the other approaches.

3) The JB-based classifiers slightly outperform PLDA-based classifiers. By fusing the 18 similarity scores produced by the two kinds of classifiers with linear

TABLE 9
Verification Rates at 0.1% FAR for Different Methods on the FRGC 2.0 Experiments 1 and 4

| | | Experiment 1 | | | Experiment 4 | | |
		ROC I	ROC II	ROC III	ROC I	ROC II	ROC III
Single Descriptor	TIP09' Gabor + LDA [42]	-	-	97.00	-	-	83.00
	TIP10' Gabor + Kernel LDA [11]	-	-	-	-	-	80.00
	PAMI13' Multiscale LPQ (MLPQ) + Kernel fusion [8]	98.99	98.68	98.37	89.26	89.80	90.36
	PAMI13' Multiscale LBP (MLBP) + Kernel fusion [8]	98.79	98.50	98.21	86.77	87.54	88.21
	MDML-DCPs + PLDA + Score averaging	99.59	99.41	99.22	93.40	93.18	92.89
	MDML-DCPs + PLDA + Linear SVM	**99.61**	**99.43**	**99.25**	**93.91**	**93.64**	**93.39**
Multiple Descriptors	AMFG07' Gabor + LBP + KDCV [40]	-	-	-	-	-	83.60
	TIP09' Gabor + DCT + LDA [42]	-	-	98.00	-	-	89.00
	ICB09' Hybrid RCrQ with Gabor + MLBP + DCT [21]	-	-	-	-	-	92.43
	TIP10' Gabor + LBP + Kernel LDA [11]	-	-	-	-	-	88.10
	TIP10' LGBP + LGXP + Block-based LDA [7]	98.60	98.00	97.30	83.90	84.70	85.20
	TIP11' Hybrid Fourier feature + LDA [43]	96.38	95.11	93.90	81.82	81.50	81.14
	PAMI13' MLPQ + MLBP + Kernel fusion [8]	99.04	98.77	98.49	90.30	90.94	91.59

TABLE 10
Mean Verification Accuracy on the LFW View 2 Data

	Accuracy(%) $\pm S_E$
PAMI12' Combined PLDA [33]	90.07 ± 0.51
ECCV12' Combined Joint Bayesian [35]	90.90 ± 1.48
ICCV13' VMRS [45]	92.05 ± 0.45
BMVC13' Fisher vector faces [46]	93.03 ± 1.05
CVPR13' high-dim LBP + JB [25]	93.18 ± 1.07
MDML-DCPs + PLDA + Score averaging	**94.57 ± 0.30**
MDML-DCPs + PLDA + Linear SVM	**95.13 ± 0.33**
MDML-DCPs + JB + Linear SVM	**95.40 ± 0.33**
MDML-DCPs + PLDA + JB + Linear SVM	**95.58 ± 0.34**

Fig. 10. ROC curves of the MDML-DCPs method and other state-of-the-art methods in the unrestricted paradigm.

SVM, we achieve a better mean verification rate of 95.58%.

6 CONCLUSION

Due to the degradation of face image quality and large variations of illumination, pose, and expression, the recognition of unconstrained face images is a challenging task. Solving this problem demands work in at least two areas: development of an effective face image descriptor and a comprehensive face representation scheme. To achieve this goal, we make the following contributions:

1) We presented a novel face image descriptor named Dual-Cross Patterns. DCP encodes second-order statistics in the most informative directions within a face image. Experimentation on four large-scale face databases shows that DCP has superior discriminative power and is robust to pose, expression, and moderate variations in illumination.

2) A comprehensive and systematic comparison of fourteen state-of-the-art face image descriptors is conducted on the four face databases. Detailed analysis is provided. Conclusions about the advantages and disadvantages of the descriptors can be drawn from these results.
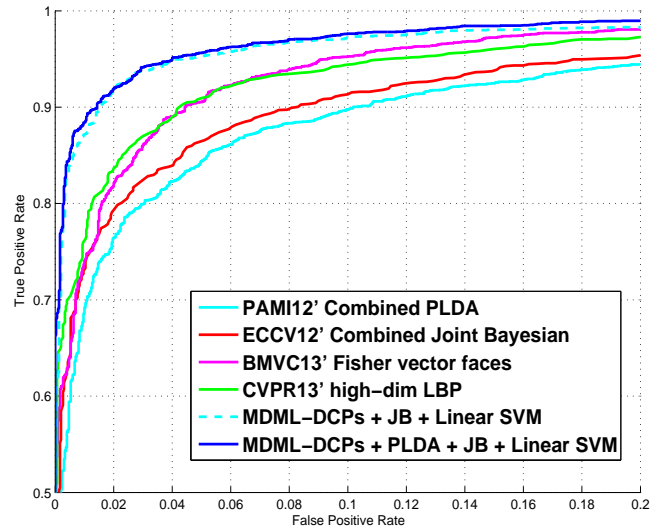
3) The proposed MDML-DCPs face representation scheme comprehensively incorporates both holistic-level DCP features and component-level DCP features. Exploiting the single descriptor DCP, MDML-DCPs consistently achieves the best results on the four databases. In particular, MDML-DCPs improves the verification rate of ROC III in Experiment 4 of FRGC 2.0 to 93.39% and outperforms the state-of-the-art result on LFW by 2.4%.

This work helps to expedite the design of practical face image descriptors and face representation schemes.
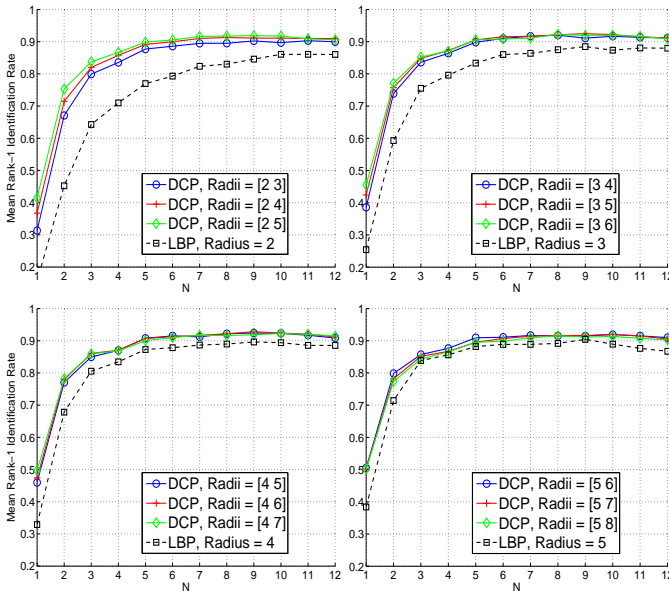
Fig. 11. The mean rank-1 identification rates of DCP and LBP on four FERET probe sets as a function of N.

APPENDIX A
PARAMETER SELECTION OF DCP

In this experiment, we take the FERET database [26] for example to illustrate the influence of the DCP parameters on its performance. The face images are normalized to 128×128 pixels, as described in Section 5 of the paper. There are three parameters in DCP, the radii R_{in} and R_{ex} (with different values of R_{in} and R_{ex} capturing information on different scales) and the region number N. A larger value of N helps to preserve spatial information but makes the descriptor more sensitive to misalignment errors. The chi-squared metric is used to measure the similarity of two face images. In this experiment, LBP is also evaluated for comparison with a sampled points number of eight and exploiting all 256 LBP codes. The radius of LBP is restricted to the value of R_{in} to highlight the importance of including more sampling points for effective pattern encoding.

The mean rank-1 identification rates against N for the four FERET probe sets using DCP and LBP are plotted in Fig. 11. Optimal performance of DCP is achieved when $N = 9$, $R_{in} = 4$ and $R_{ex} = 6$. DCP consistently achieves better performance than LBP with all sets of parameters. In particular, DCP outperforms LBP by 10% to 30% when $N \leq 2$, which justifies the motivation of DCP to include more sampling points for pattern encoding.

The chi-squared metric has been used to measure the similarity of two face images, but it is noted that DCP works equally well when assessed using the histogram intersection metric. The mean identification rates using DCP measured with the chi-squared and histogram intersection metrics are 92.80% and 92.66%, respectively.

TABLE 11
AUC Values of Face Image Descriptors on the Morph Database

	Mean$\pm$ Std		Mean$\pm$ Std
LBP	0.9479 ± 0.0045	MsTLBP	0.9506 ± 0.0037
LTP	0.9499 ± 0.0048	MsDLBP	0.9393 ± 0.0044
LPQ	0.9454 ± 0.0044	DCP-1	0.9499 ± 0.0035
POEM	0.9500 ± 0.0045	DCP-2	0.9529 ± 0.0034
LGXP	0.9369 ± 0.0040	**DCP**	**0.9557 ± 0.0044**
MsLBP	0.9494 ± 0.0037		

APPENDIX B
PERFORMANCE OF DCP ON FACIAL AGING DATABASE

We have also tested the performance of the face image descriptors on the Morph database [47]. The face images in Morph are geometrically and photometrically normalized as described in Section 5 of the paper. We follow the protocol defined in [29] and report the performance of the face image descriptors in Table 11. The same as [29], a subset of 1,700 images in the Morph database is randomly selected for experiment. It is shown that the proposed DCP descriptor achieves the best performance on the Morph database, indicating that DCP is also robust to the aging factor.

ACKNOWLEDGMENTS

REFERENCES

[1] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, Oct. 2007.

[2] R. Beveridge, H. Zhang, B. Draper *et al.*, "Report on the fg 2015 video person recognition evaluation," in *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognit.*, 2015.

[3] C. Ding and D. Tao, "A comprehensive survey on pose-invariant face recognition," *arXiv preprint arXiv:1502.04383*, 2015.

[4] J. Wright, A. Y. Yang, A. Ganesh, S. 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.

[5] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[6] Z. Cao, Q. Yin, X. Tang, and J. Sun, "Face recognition with learning-based descriptor," in *Proc. Comput. Vis. Pattern Recognit.*, 2010, pp. 2707–2714.

[7] S. Xie, S. Shan, X. 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.

[8] C. Chan, M. Tahir, J. Kittler, and M. Pietikainen, "Multiscale local phase quantisation for robust component-based face recognition using kernel fusion of multiple descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 5, pp. 1164–1177, May 2013.

[9] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 1701–1708.

[10] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 1891–1898.

[11] 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.

[12] J. Trefnỳ and J. Matas, "Extended set of local binary patterns for rapid object detection," in *Proc. Comput. Vis. Winter Workshop*, 2010.

[13] B. Zhang, S. Shan, X. Chen, and W. Gao, "Histogram of gabor phase patterns (hgpp): A novel object representation approach for face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 57–68, Jan. 2007.

[14] 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. Int. Conf. Comput. Vis.*, vol. 1, 2005, pp. 786–791.

[15] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkila, "Recognition of blurred faces using local phase quantization," in *Proc. Int'l Conf. Pattern Recognit.*, 2008, pp. 1–4.

[16] N.-S. Vu and A. Caplier, "Enhanced patterns of oriented edge magnitudes for face recognition and image matching," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1352–1365, Mar. 2012.

[17] S. ul Hussain and B. Triggs, "Visual recognition using local quantized patterns," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 716–729.

[18] Z. Lei, M. Pietikainen, and S. Z. Li, "Learning discriminant face descriptor." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 289–302, Feb. 2014.

[19] L. Wolf, T. Hassner, and Y. Taigman, "Effective unconstrained face recognition by combining multiple descriptors and learned background statistics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 1978–1990, Oct. 2011.

[20] C. Xu, D. Tao, and C. Xu, "Multi-view intact space learning," *IEEE Trans. Pattern Anal. Mach. Intell., doi: 10.1109/TPAMI.2015.2417578*, 2015.

[21] Z. Liu and C. Liu, "Robust face recognition using color information," in *Proc. Int'l Conf. Biometrics*, 2009, pp. 122–131.

[22] J. Wright and G. Hua, "Implicit elastic matching with random projections for pose-variant face recognition," in *Proc. Comput. Vis. Pattern Recognit.*, 2009, pp. 1502–1509.

[23] C. Ding, C. Xu, and D. Tao, "Multi-task pose-invariant face recognition," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 980–993, 2015.

[24] X. Xiong and F. De la Torre, "Supervised descent method and its applications to face alignment," in *Proc. Comput. Vis. Pattern Recognit.*, 2013, pp. 532–539.

[25] D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification," in *Proc. Comput. Vis. Pattern Recognit.*, 2013, pp. 3025–3032.

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

[27] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao, "The cas-peal large-scale chinese face database and baseline evaluations," *IEEE Trans. Syst., Man, Cybern. A*, vol. 38, no. 1, pp. 149–161, Jan. 2008.

[28] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proc. Comput. Vis. Pattern Recognit.*, vol. 1, 2005, pp. 947–954.

[29] Z. Akhtar, A. Rattani, A. Hadid, and M. Tistarelli, "Face recognition under ageing effect: A comparative analysis," in *Proc. Int. Conf. Image Anal. Process.*, 2013, pp. 309–318.

[30] J. Choi, W. R. Schwartz, H. Guo, and L. S. Davis, "A complementary local feature descriptor for face identification," in *Proc. IEEE Workshop Applications of Comput. Vis.*, 2012, pp. 121–128.

[31] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, Nov. 1986.

[32] S. J. Prince, J. Warrell, J. Elder, and F. M. Felisberti, "Tied factor analysis for face recognition across large pose differences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 970–984, 2008.

[33] P. Li, Y. Fu, U. Mohammed, J. H. Elder, and S. J. Prince, "Probabilistic models for inference about identity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 144–157, Jan. 2012.

[34] S. J. Prince and J. H. Elder, "Probabilistic linear discriminant analysis for inferences about identity," in *Proc. Int. Conf. Comput. Vis.*, 2007, pp. 1–8.

[35] D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun, "Bayesian face revisited: A joint formulation," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 566–579.

[36] T. Liu and D. Tao, "Classification with noisy labels by importance reweighting," *IEEE Trans. Pattern Anal. Mach. Intell., doi: 10.1109/TPAMI.2015.2456899*, 2015.

[37] S. U. Hussain, T. Napoléon, F. Jurie *et al.*, "Face recognition using local quantized patterns," in *Proc. Int'l Conf. Biometrics*, 2012.

[38] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[39] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, pp. 27:1–27:27, Apr. 2011.

[40] X. Tan and B. Triggs, "Fusing gabor and lbp feature sets for kernel-based face recognition," in *Proc. Int'l Conf. Analysis and Modeling of Faces and Gestures*, 2007, pp. 235–249.

[41] S. Xie, S. Shan, X. Chen, X. Meng, and W. Gao, "Learned local gabor patterns for face representation and recognition," *Signal Process.*, vol. 89, no. 12, pp. 2333–2344, Dec. 2009.

[42] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical ensemble of global and local classifiers for face recognition," *IEEE Trans. Image Process.*, vol. 18, no. 8, pp. 1885–1896, Aug. 2009.

[43] W. Hwang, H. Wang, H. Kim, S.-C. Kee, and J. Kim, "Face recognition system using multiple face model of hybrid fourier feature under uncontrolled illumination variation," *IEEE Trans. Image Process.*, vol. 20, no. 4, pp. 1152–1165, Apr. 2011.

[44] Y. Li, S. Shan, H. Zhang, S. Lao, and X. Chen, "Fusing magnitude and phase features for robust face recognition," in *Proc. Asian. Conf. Comput. Vis.*, 2013, pp. 601–612.

[45] O. Barkan, J. Weill, L. Wolf, and H. Aronowitz, "Fast high dimensional vector multiplication face recognition," in *Proc. Int. Conf. Comput. Vis.*, 2013.

[46] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Fisher vector faces in the wild," in *Proc. Brit. Mach. Vis. Conf.*, vol. 1, no. 2, 2013, p. 7.

[47] K. Ricanek and T. Tesafaye, "Morph: A longitudinal image database of normal adult age-progression," in *Proc. Seventh Int'l Conf. Automatic Face and Gesture Recognit.*, 2006, pp. 341–345.