Robust Face Recognition with Structural Binary Gradient Patterns

Weilin Huang\textsuperscript{a,c}, Hujun Yin\textsuperscript{b,*}

\textsuperscript{a}Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
\textsuperscript{b}School of Electrical and Electronic Engineering, The University of Manchester
\textsuperscript{c}Department of Engineering Science, University of Oxford

Abstract

This paper presents a compact and efficient yet powerful binary framework based on image gradients for robust facial representation. It is termed as Binary Gradient Patterns (BGP). To discover underlying local structures in the gradient domain, image gradients are computed from multiple directions and encoded into a set of binary strings. Certain types of these binary strings have meaningful local structures and textures, as they detect micro oriented edges and retain strong local orientation, thus enabling great discrimination. Face representations by these structural BGP histograms exhibit profound robustness against various facial image variations, in particular illumination. The binary strategy realized by local correlations substantially simplifies the computational complexity and achieves extremely efficient processing with only 0.0032s in Matlab for a typical image. Furthermore, the discrimination power of the BGP has been enhanced on a set of orientations of the image-gradient magnitudes. Extensive experimental results on various benchmarks demonstrate that the BGP-based representations significantly improve over the existing local descriptors and state-of-the-art methods in the terms of discrimination, robustness and complexity and in many cases the improvements are substantial. Combining with the deep networks, the proposed descriptors can further improve the performance of the deep networks on real-world datasets. Matlab codes for the

*Corresponding author

Email address: h.yin@manchester.ac.uk; Tel:+44(0)1613068714 (Hujun Yin)
1. Introduction

In the last two decades, face recognition has been one of the most active topics in image processing and pattern recognition due to much increased demands for its application in law enforcement, surveillance, human-computer interaction, and multimedia. Although tremendous progress has been made and deployments now exist in some areas, its reliability is still regarded as a serious issue in the real-world, unconstrained situations [1], where large within-class variations in facial appearance commonly occur due to illumination, expression and poses for instance. A key solution lies in facial representation and a great deal of effort has been devoted to it. Although learning especially deep learning mechanisms help to close the gap, well-defined, discriminative facial descriptors still play an dominant role in face recognition applications.

Appearance-based methods, one of widely adopted approaches, consider face images as holistic vectors of pixel intensities in high-dimensional space. Dimensionality reduction is typically applied to extract intrinsic features [2]. Typical early examples are Eigenfaces [3] and Fisherfaces [4] where linear PCA is used. They have been enhanced by nonlinear PCA or manifold methods [5] [6] [7] [8]. However, faces represented by pixel intensities are sensitive to illumination variation, noise and occlusion.

Searching for robust descriptors has been an important line of research in computer vision. Zhang et al. [9] have proposed a novel descriptor, termed Gradientfaces, to extract illumination insensitive features in the image gradient domain. Faces are described by Image Gradient Orientation (IGO) instead of intensity to achieve stronger invariance to illumination changes. To further take advantage of gradient features, Tzimiropoulos et al. [10] derived a similarity
measure based on cosine of IGO differences between images. They showed that this measure considerably mitigated the effect of variations. However, these measures, computed by pixel-wise correlations, are holistic representations and hence sensitive to local deformation, rotation and spatial scale, thus prone to facial variations such as expressions and poses [11, 12, 13].

Local feature descriptors have recently gained considerable attention due to their resilience to multiple variations by enforcing spatial locality in both pixel and patch levels. Two examples are Gabor wavelets [14] and Local Binary Patterns (LBP) [15, 16]. Local Gabor descriptor extracts micro textural details. Fusing these local features can provide certain global shape information, making the features robust to local distortions and leading to improvements in face recognition [17, 18, 19]. However, Gabor representations generate a large number of features with multiple convolution kernels, making them inefficient in real-time applications. Whilst, the LBP features are simple and efficient, and provide good invariance to illumination. They are also capable of detecting certain micro textures, such as spots, corners and edges [16]. However, their capability is severely affected by drastic changes of pixel intensities, such as extreme lighting. Most current local facial descriptors based on Gabor and LBP also suffer from such inherent limitations [20, 21, 18, 22].

Building on the properties gained from the IGO domain and local binary features, this paper presents a new local facial descriptor, termed Binary Gradient Patterns (BGP). It measures relationships between local pixels in the image gradient domain and effectively encodes the underlying local structures into a set of binary strings, not only increasing the discriminative power but also significantly simplifying computational complexity. We observe that the structural patterns of the descriptor are capable of detecting stable micro edges of various directions. Local features built on the histogram statistics of these oriented edge textures resemble the primary structural information extracted in biological vision systems and exhibit desirable characteristics of spatial locality, orientation and scale selectivity. They show stronger orientational power than the LBP and Gabor features, leading to improved discriminative representation. Fur-
thermore, an enhanced descriptor is devised by building BGP patterns on a set of orientations of the image-gradient magnitudes, termed as BGPM, to further enforce locality and orientation. Extensive experiments on several benchmark databases demonstrate the significant advantages of the BGP-based descriptors over existing face descriptors. By incorporating the state-of-the-art deep architectures such as the Convolutional Network Networks (CNN), the BGP with (RGB representation) obtains an accuracy of 99.32% (single model) on the widely-used LFW database, among the top performers on this benchmark.

A brief review on related work is given in Section 2. The proposed BGP descriptor is then presented in Section 3. Section 4 discusses favorable properties of the structural BGP and its connections and distinctions with BGP, LBP and Gabor representations. Section 5 describes the enhanced BGPM descriptor. Finally, experimental evaluations and extensive comparisons are provided in Section 6, followed by the concluding remarks.

2. Related work

Local histograms built on the IGO statistics have been considered as prominent visual features with favorable properties such as invariance against illumination changes. The Gradientfaces [9] use IGOs instead of intensities to obtain an illumination insensitive measure. It has been shown that features extracted from the gradient domain are more discriminative and robust than those from the intensity domain. Similarly, Tzimiropoulos et al. [10] presented a simple yet robust similarity measure based on IGO representation, termed IGOPCA, in which cosine of IGO differences between images (termed $IGO_{cos}$) was calculated, followed by a PCA subspace reduction.

Recently, a number of local facial descriptors have been derived from the Gabor or LBP features or their combinations. In [13], a Local Gabor Binary Pattern Histogram Sequence (LGBPHS) was proposed by running Gabor filters and building LBP histogram sequentially on face images. Similar methods include Histogram of Gabor Phase Patterns (HGPP) [22] and Gabor Volume
based LBP (GV-LBP) [23]. The advantages of these methods are built on the
virtues of both Gabor and LBP descriptors. However, they commonly suffer
from the drawbacks of Gabor-based representations, i.e. high computational
complexity and dimensionality.

As a simpler approach, Chen et al. [20] proposed a Weber Local Descriptor
(WLD) based on the Weber’s Law of human perception system, which states
that the size of a noticeable change is a constant proportion of the original
stimulus. In [21], Tan and Triggs presented Local Ternary Patterns (LTP) by
extending the LBP to 3-valued codes. Recently, Local Directional Number Pat-
tern (LDNP) was proposed to extract and encode directional information of
face textures [24]. These methods have been shown to be discriminative but
also resistant to illumination changes, exceeding the performance of LBP. How-
ever, similar to LBP, these descriptors build local relationships in the intensity
domain, which is severely affected by dramatic changes of pixel values.

In [25], gradient directions were computed around robust edges detected by
multi-step Canny edge detectors. In [26], multiple dissimilarity measures were
developed to improve robustness of the self-quotient image against large illumi-
nation changes. Dubey et al. [27] proposed the interleaved intensity order-based
local descriptor (IOLD) constructed on a set of local neighbors. Arandjelovic
and Cipolla [28] showed that the photometric model could be exploited and
a “re-illumination” algorithm was proposed for robust face recognition from
videos. A more comprehensive study on illumination-invariant representations
for face recognition can be found in [29].

Some recent efforts have been devoted to developing learning based ap-
proaches. For example, Wagner et al. [30] developed a sparse coding based
method for simultaneously handling variations in illumination, alignment and
occlusion. Lu et al. developed a Discriminative Multi-Manifold Analysis (DMMA)
method by learning discriminative features from image patches [31]. A general
framework for learning local image descriptors was introduced in [32].

The proposed BGP descriptor is closely related to the Center-Symmetric
Local Binary Patterns (CS-LBP) [33], in which local binary patterns are com-
computed from symmetric neighboring pixels. However, the BGP differs distinctly in three aspects. First, the BGP is defined in the IGO domain. Second, structural patterns and multiple spatial resolutions are used in the BGP. A number of theoretical insights are provided to show that structural BGPs work as oriented edge detectors, a key to a discriminative and compact representation. The multiple spatial resolution strategy increases the descriptor’s capability of covering considered neighbors of different radius. Third, similar to the multiple channel strategy of the invariant descriptors such as the SIFT [34] and POEM [12], the BGP descriptor can be computed on a set of oriented image gradient magnitudes; to further enhance its discriminative power. Their differences in discriminative ability and invariance to illumination will be demonstrated in Section 6.2.

3. Binary Gradient Patterns

With the profound properties of the IGO representation, we naturally consider to extract robust facial features from the IGO domain, and at the same time, to improve its robustness against local distortions by enforcing block-level locality. A straightforward approach is to directly compute histogram statistics of the IGO representation (HIGO) by dividing an image into a number of non-overlapped blocks. Each block is represented by an IGO histogram, whose bin number is determined by the segmentation between $[0, 2\pi)$.

To illustrate the advantage of HIGO, two simple experiments were con-
ducted. On the AR database, 100 subjects from the group of natural faces were used as gallery images, and two groups of faces with scream expressions and scarf occlusions (both cause large-scale local distortions) were presented as probe images. Each group included 100 images from the first session. On the Extended YaleB database, a subset of 10 subjects was used. The faces with the most natural light sources were used as galleries and two sets of medium and high illumination conditions (corresponding to sets 4 and 5 in [35]) were presented as the probes. We evaluated the IGO-based representations, such as Gradientfaces [9] and $IGO_{cos}$ [10], local histogram methods, e.g. LBP [36] and HIGO (with respect to bin numbers). The results are illustrated in Fig. 1, which evidently provides two observations. First, local features seem more robust to local deformations in expression and occlusion. Second, IGO methods are more capable than LBP of dealing with illumination changes. HIGO, taking advantages of both, appears the most robust to these effects. Notice that the experiments here aimed to show the experimental cue that motivated the derivation of the proposed descriptor. Systematical evaluations are reported in Section 6.

Our goal was to develop a descriptor that could effectively integrate the advantages of both approaches, while still being computationally efficient. To this end, we employ a compact representation, four-bin IGO histogram, as our basic model, based on the following considerations: (i) it is able to achieve good performance by using such a surprisingly compact representation, as shown in Fig. 1; (ii) it effectively balances computationally complexity and discrimination, by allowing acceptable loss of information; (iii) this four-bin IGO representation can be easily extended to a more powerful descriptor, with linear scalable computational increase, as described in Section 3 and 4; (iv) its capability for extracting meaningful image feature can be well explained theoretically, as discussed bellow.
3.1. Theoretical Analysis of Four-Bin HIGO

IGO computed by a four-quadrant inverse tangent, can be formulated as,

$$\Theta_{x,y} = \arctan\left(\frac{\text{sign}(G_{y,x,y})}{\text{sign}(G_{x,x,y})} \left|\frac{G_{y,x,y}}{G_{x,x,y}}\right|\right)$$  \hspace{1cm} (1)

where \(\text{sign}(G_{y,x,y})\) and \(\text{sign}(G_{x,x,y})\) return the signs of the gradients in the vertical and horizontal directions. \(|G_{y,x,y}|\) and \(|G_{x,x,y}|\) are the gradient contrasts. In a four-bin histogram, each bin counts the number of pixels whose IGO values located in one of four quadrants, e.g. \(\theta_{x,y} \in [0, \pi/2)\). Hence IGO values are quantified into four discrete values as \(\{0, \pi/2, \pi, 3\pi/2\}\), by discarding gradient contrast,

$$\hat{\Theta}_{x,y} = \arctan\left(\frac{\text{sign}(G_{y,x,y})}{\text{sign}(G_{x,x,y})}\right)$$  \hspace{1cm} (2)

In a four-bin HIGO, pattern labels can be directly computed by four different combinations of two gradient signs as [+ +], [+ -], [- +] and [- -], which are naturally applicable to binary strategy. Similar to LBP [16], the signs of the gradients are not affected by the changes of the mean intensities, yielding a distinct ability to resist gray-scale variations. Subsequently, we can use two binary bits to describe the patterns of four-bin HIGO as, 11, 10, 01 and 00. While LBP discards intensity contrast, two-bit HIGO discards gradient contrast to achieve illumination invariance as well as computational efficiency. To this end, we have derived the basic local binary features from the IGO domain, which serve as the basis of the proposed BGP descriptor.

3.2. Binary Image Gradients from Multiple Directions

The traditional IGO is computed on gradients of the horizontal and vertical directions. Its pixel-level locality is realized by using only four neighbors in two orthogonal directions. Most current high-performing local descriptors extract meaningful local information from at least eight neighbors and their discriminative power can be improved by suitably increasing the number of neighbors [15] [16] [17] [20] [23] [21] [37]. Similarly, it is expected that greater discrimination
can be achieved in the gradient domain by involving more local neighbors from multiple directions.

Following this intuition, we extend the four-bin HIGO to multiple directions, resulting a new facial descriptor, Binary Gradient Patterns (BGP). Specifically, the BGP computes binary correlations between symmetric neighbors of a central pixel from multiple \((k)\) directions. The number of neighbors is twice the number of directions. The computation is simple. The basic BGP operator of four directions is presented in Fig. 2 and detailed below:

1). A set of local neighbors of a central pixel is first given (e.g. eight neighbors in Fig. 2(a)).

2). Then, a pair of binary numbers, principal \((B^+_i)\) and associated \((B^-_i)\), are computed by correlating two symmetric neighbors in each direction based on Eq. 3: and totally four pairs of binary numbers are devised from four directions: G1, G2, G3 and G4, shown in Fig. 2(b) and (c):

\[
B^+_i = \begin{cases} 
1 & \text{if } G^+_i - G^-_i \geq 0 \\
0 & \text{if } G^+_i - G^-_i < 0 
\end{cases} \quad (3)
\]

\[
B^-_i = 1 - B^+_i \quad i = 1, 2, \ldots, k
\]

where \(G^+_i\) and \(G^-_i\) are the intensity values of the pixels corresponding to locations in Fig. 2(b).

3). Finally, label of the central pixel is computed from the resulting four
principal binary numbers,

\[ L = \sum_{i=1}^{k} 2^{i-1} B_i^+ \] (4)

Although eight binary numbers are obtained in four directions, the principal and associated binary numbers in each direction are always complementary. Hence, there are only two variances in each direction, requiring only a single binary bit to describe. For a compact representation, only the principal binary bits are needed for computing the labels by Eq. (4) and describing all possible variances of the BGP patterns. The number of BGP labels \( N_L \) is determined by the number of the principal binary bits, equal to the number of directions \( k \), \( N_L = 2^k \). Thus there are \( 2^k \in \{0, 1, 2, \ldots, 2^{k-1}\} \) possible labels for a \( k \)-directional BGP operator. Note that this number \( (2^k) \) is substantially smaller than the number of LBP labels \( (2^{2k}) \). In a typical model with sixteen neighbors, the numbers of labels for BGP and LBP are 256 and 65536, respectively.

3.3. Structural BGP

There are sixteen different labels for a four-directional BGP descriptor. The binary structures of these labels, ranging from 0 to 15, are shown in Fig. 3(a). As can be seen, each label is constructed by eight binary numbers/bits, including four bits of "1" and four bits of "0". The principal bits are presented in red or bold. It is interesting to investigate the distributions of "1"s and "0"s in these

![Figure 3: (a) Definition of structural (in red/bold boxes) and non-structural (in other boxes) BGP patterns. (b) List of the structural patterns.](image)
labels. It can be seen that certain labels have meaningful structures where four bits of "1" are located consecutively. There are eight such labels having consecutive bits of "1" (marked as red- or bold-lined boxes in Fig. 3(a)); while the "1"s in the other eight labels (marked as black or thin-lined boxes) are discontinuous. These continuous "1"s indicate more stable local changes in texture and essentially describe the orientations of local "edge" texture. An observation is that statistics on these patterns is highly stable and meaningful to characterize local structures. By contrast, labels with discontinuous "1"s contain arbitrary changes of local texture, likely to indicate noise or outliers. Furthermore, from the experimental statistics, patterns having continuous "1"s often take up a vast majority in a typical BGP face, e.g. 95%. The statistics of BGP patterns of various labels on the 2600 face images of the AR databases is presented in Fig. 4(a).

Based on these observations, we define the patterns having continuous "1"s as the structural BGP, while refer the others as non-structural patterns. This yields in total eight different labels for the structural patterns (as listed in Fig. 3(b)) while discarding all non-structural ones. Therefore, only eight bins are needed for the structural BGP histogram. This is an appealing property, not only helping to rule out noise and outliers in face images, but also further reducing feature dimensions. For instance, even with 24 neighbors, the BGP has 12 principal binary bits, and thus the structural BGP are the 12-bit pat-
tens computed by moving 12 continuous "1" through 24 neighbors' locations, similar to computation of the 8-neighbor BGP in Fig. 3. This results in only 24 structural patterns for histogram representation (needing only 24 bins), compared to $2^{12} = 4096$ bins in the CS-LBP histogram [33]. Further discussions and evaluations are presented in Section 4. Hence BGP is referred to structural BGP.

3.4. Spatial Resolutions

The basic BGP descriptor (Fig. 2) is computed from four directions ($k = 4$) in a square neighborhood of side length of two units. Similar to LBP-based descriptors, the capability of BGP can be further improved by increasing the number of gradient directions and by enlarging the neighborhood. To this end, we define the spatial resolution of BGP by the number of neighbors/directions and radius of the square, indicated as ($P,R$). Typically, the maximum number of neighbors is eight times of the radius, $P_{\text{max}} = 8R$, e.g. (8, 1), (16, 2) and (24, 3).

The BGP descriptor with structural patterns in spatial resolution of ($P,R$) is referred as $\text{BGP}_{P,R}$. Assuming that the number of neighbors is maximized with respect to the radius, we present a generalized algorithm for computing the BGP operator from a given pixel in location $(i,j)$, with spatial resolution of ($P,R$), in Algorithm 1.
Algorithm 1 Computing structural BGP descriptor

Require: Location of a given pixel \((i, j)\), spatial resolution, \((P, R)\) and \(I(i,j)\), pixel intensity.

Ensure: Label of structural BGP descriptor, \(L(i,j)\).

1: **step one**: compute principal binary numbers \(B_k^t\) in \(k\) directions, \(k = P/2\).
2: \(t = 1\)
3: for \(n_1 = -R \rightarrow R\) do
4: \[
B_k^t = \begin{cases} 
1 & \text{if } I(i+n_1,j+R) - I(i-n_1,j-R) > 0 \\
0 & \text{if } I(i+n_1,j+R) - I(i-n_1,j-R) < 0 
\end{cases}
\] (5)
5: \(t = t + 1\)
6: end for
7: for \(n_2 = -(R-1) \rightarrow (R-1)\) do
8: \[
B_k^t = \begin{cases} 
1 & \text{if } I(i+R,j-n_2) - I(i-R,j+n_2) > 0 \\
0 & \text{if } I(i+R,j-n_2) - I(i-R,j+n_2) < 0 
\end{cases}
\] (6)
9: \(t = t + 1\)
10: end for
11: **step two**: compute \(L(i,j)\) by Eq. (4).
12: return Label of pattern, \(L(i,j)\).

Algorithm 2 Computing structural label

Require: Number of neighbors, \(P\).

Ensure: Labels of structural patterns, \(L^{sp}\).

1: Number of directions, \(k = P/2\).
2: for \(t = 1 \rightarrow P\) do
3: if \(t \leq k\) then
4: \(L^{sp}_t = 2^{t-1} - 1\)
5: else
6: \(L^{sp}_t = 2^{k} - L^{sp}_{2k-t+1} - 1\)
7: end if
8: end for
9: return \(\{L^{sp}_{t}\}_{t=1}^{P}\).

Algorithm 2 returns the label value of a pixel. To build histograms of the BGP structural patterns, one needs to know the number of the structural labels and their values, which is independent of face images and only determined by the given spatial resolution. From Fig. 3(a), we can find that four continuous \(^1\)s in eight structural labels run through all locations of eight neighbors, indicating that the number of structural labels \(N_{sp}\) is equal to the number of neighbors, \(N_{sp} = P\), compared to \(2^P\) of the LBP and \(2^{2P}\) of the CS-LBP. Based on the distributions of principal bits of the structural labels, we device Algorithm 2 for computing structural labels at resolution of \((P, R)\).
4. Analysis, Discussions and Comparisons

In this section, favorable characteristics of the BGP descriptor are discussed and compared with two fundamental descriptors, LBP and Gabor wavelets, in terms of discrimination, robustness and complexity. Insights can be gained by examining the underlying connections and distinctions among these descriptors, further supported by experimental studies.

Both BGP and LBP employ advantageous binary strategy for extracting pixel correlations in local neighborhoods. However, BGP differs from LBP in computing binary correlations, leading to distinctive properties between them.

4.1. Discrimination

Based on the definition of structural patterns in the previous section, the proposed BGP descriptor is primarily an orientated edge detector. Fig. 5 illustrates the outputs of $BGP_{8,1}$ and $LBP_{8,1}^{riu2}$ descriptors on a typical face image (from the AR database). It is evident that LBP detects various local textural features such as spots, corners and edges, while BGP extracts orientated edge features. The $BGP_{8,1}$ faces assemble more facial information than the $LBP_{8,1}^{riu2}$ faces. Location maps of the $BGP$ structural patterns are more informative and discriminative than those of LBP uniform patterns. The histogram statistics of the two descriptors in Fig. 4 show that distributions of BGP structural patterns are fairly even, while distributions of LBP uniform patterns mainly peaks at few patterns (UP04, UP05 and UP03). This means that all BGP structural patterns contribute fairly evenly, while LBP representation is dominated by few patterns.

As stated in [16], these three types of LBP patterns (UP03, UP04 or UP05) detect edge information from textural image and lead to the finding that edge information dominates local textural features of face images. In fact, the local structures described by these patterns of the LBP are guaranteed by the BGP structural labels, as shown in Fig. 6 determined by the orientation. In other words, LBP fuses all directions of the patterns into a single label, while
Figure 5: Demonstrations of Gabor, LBP and BGP faces. Top part: original AR face and Gabor magnitude faces of eight orientations and a fixed scale. Middle part: $LBPR_{8,1}$ face and location maps of eight uniform and one nonuniform patterns. Bottom part: $BGP_{8,1}$ face and location maps of eight structural and one nonstructural patterns.
Figure 6: Connections between LBP uniform patterns and BGP structural patterns. One of LBP UP04 can be transformed to BGP SP07, and one of LBP UP03 or UP05 can be transformed to BGP SP07 or SP15. But different structures of LBP UP04 may relate to different type of BGP structural patterns. Other types of LBP uniform patterns are not guaranteed to match BGP structural patterns.

structural BGP separately counts different oriented edges in eight labels to increase discrimination in orientation. Although HOG also computes histograms from multiple orientations, the gradient orientations used are computed only from four neighbors, thus cannot effectively detect all edge information and are insufficient to capture all the local structures.

Furthermore, the BGP resembles some essential properties of the human vision system, which is characterized by spatial locality, orientation and scale selectivity [38], and responds strongly to oriented lines or edges presented in the receptive fields [39]. Gabor wavelets are a well-known model for describing these properties and have had considerable successes in image feature representation [14, 17]. The inherent characteristics of the BGP possess these properties by enforcing its locality in both pixel and block levels, building on the statistics of edge orientations, and defining a tunable spatial resolution. Fig. 5 provides an intuitive view that BGP faces preserve better orientations than Gabor faces.

4.2. Robustness

LBP yields gray-scale invariance by discarding intensity contrast. BGP achieves invariance against illumination changes by employing a heuristic from the IGO representation to further take advantage of the gradient-domain local representation. The illumination invariance of IGO-based representations has
Figure 7: Histogram statistics of pixel intensity, LBP and BGP patterns on the two blocks (15 x 15) of two faces in (a) (same identity but significantly different illuminations).

been verified in the reflectance model by canceling out illumination of different directions when computing the ratio of gradients [9]. Benefiting from this merit, BGP gains stronger robustness to illumination variation than LBP by further discarding gradient contrast. Fig. 7 shows the histogram statistics of LBP and BGP patterns on two exemplar face blocks, illustrating the improved invariance of BGP against extreme illumination conditions.

Furthermore, BGP discards non-structural patterns, which contain non-smooth or discontinuous changes of local pixels. These patterns are often caused by noise or outliers, and contain little structural and meaningful information. The experimental statistics on 5032 face images from the AR and YaleB databases show that there is a very low proportion (only 5.8%) of these non-structural patterns in general face images, lower than the proportion (about 8.5%) of LBP nonuniform patterns. In addition, some LBP uniform labels have small numbers of patterns. For example, UP00 and UP08, shown in Fig. 4(b) and Fig. 5 (middle pan), detect bright and dark spots, respectively [16]. These types of patterns can easily include irregular appearances, such as noisy spots and corrupted pixels.

As shown in Fig. 8(b), BGP non-structural patterns contain or capture noise. BGP discards them to mitigate the effect of noise. Whilst LBP retains all of its nonuniform patterns and assign additional labels to them. Subsequently the numbers of LBP non-uniform, spot and corner patterns increase.
Figure 8: Noise on BGP non-structural and LBP non-uniform patterns. (a) Original face and with added Gaussian noise; (b) BGP non-structural and (c) LBP non-uniform patterns; (d) LBP UP00 and UP08 patterns (spots); and (e) LBP UP01 and UP07 patterns (corners).

dramatically when noise is present, as shown in Fig. 8(c)-(e).

4.3. Complexity

Complexity often refers to computational efficiency or speed and storage demand. For BGP and LBP, computational speed depends on the number of binary correlations and the number of resulting (principal) binary numbers (labels), both of which are determined by the number of neighbors/directions. The computation of Gabor features depends on the number of convolutions, which apply multiple Gabor kernels in various scales and orientations. So, the speed is determined by the numbers and sizes of the Gabor kernels. The storage demands of these three descriptors are measured mainly by the feature dimensions. The dimensions of local histogram-based features are the product of number of labels (bins) and number of blocks.

In this comparison, the numbers of neighbors were set to their maximum numbers with respect to the radii for both binary descriptors, \( P = 8R \), e.g. (8, 1), (16, 2) and (24, 3). The Gabor faces were run by using their typical parameter setting: kernel size of 31 × 31, eight orientations and five scales. We computed the average running time per face and the feature dimensions, together with the numbers of computational units for a given pixel and the numbers of labels generated by the descriptors. The results on the AR and YaleB face databases (5053 faces in total) are given in Table 1. The image size
Table 1: Complexity of BGP, LBP and Gabor features.

| Descriptors | ♯ comp. units | time(s) | ♯ labels | ♯ dimensions |
|-------------|---------------|---------|----------|--------------|
| Gabor       | 38440         | 0.9699  | -        | 400000       |
| LBP
(P,R) = (24,3) | 48            | 0.0189  | 555      | 19980        |
| BGP         | 24            | 0.0097  | 24       | 864          |
| (P,R) = (16,2) |               |         |          |              |
| LBP
(P,R) = (16,2) | 32            | 0.0128  | 243      | 8748         |
| BGP         | 16            | 0.0065  | 16       | 576          |
| (P,R) = (8,1)  |               |         |          |              |
| LBP
(P,R) = (8,1) | 16            | 0.0056  | 59       | 2124         |
| BGP         | 8             | 0.0032  | 8        | 288          |

was 100 × 100 and the numbers of blocks used by LBP and BGP were the same, 36. The experiments were run on a typical PC with AMD Dual Core processor of 2.2GHz and RAM of 4GB. BGP was run by our unoptimized MATLAB code. LBP code was from the authors of [16, 36] (also in MATLAB). Gabor representation was run based on the Matlab implementations of [40, 17].

As can be seen, the complexities of the two binary descriptors are significantly lower than that of Gabor. The ratios of computational cost, execution time and final feature dimension between Gabor feature and basic BGP descriptor are staggering 4800:1, 300:1 and 1400:1, respectively. The execution times of BGP are only about half of that of LBP in all resolutions. Running a basic BGP operator on a typical face image takes only 0.0032s, making it applicable to real-time applications. Furthermore, BGP uses much fewer pattern labels than the LBP descriptor, i.e. much lower dimensionality of the features. The dimensions of BGP features are only about 13.6%, 6.6% and 4.3% of $LBP^{u2}$ in three spatial resolutions. For CS-LBP [33], in the (24,3) case, the number of dimensions increases to $2^{12} \times 36 \approx 1.5 \times 10^5$, 170 times more than BGP. Hence, the BGP descriptor is extremely efficient and compact.

5. BGP on Orientations of Image-Gradient Magnitudes

Various extensions of Gabor and LBP representations have helped to yield some state-of-the-art local facial representations, such as combining both prop-
erties and enforcing spatial locality, orientation and robustness \[18\] \[22\] \[23\] \[21\] \[20\] \[12\]. Improving on these methods, we further propose a framework by applying BGP on orientational IGM (OIGM), abbreviated as BGPM, to enhance discriminative power by enforcing spatial locality and orientation. The framework of BGPM is depicted in Fig. 9 and its details given in Algorithm 3.

**Algorithm 3** Computing BGPM descriptor

**Require:** An input image.

**Ensure:** A set of BGPM images.

1. Compute IGO and IGM from input image.
2. Generate a set of OIGM images:
   - (a) Quantize the IGO into a number of dominant orientations, as \[20\].
   - (b) Generate an OIGM image for each \((t\text{-th})\) dominant orientation as:

\[
M_{i,j} = \frac{1}{n} \sum_{(u,v) \in \Omega_{i,j}^t} M_{u,v} \quad t = 1, 2, \ldots, s
\]

where \((i, j)\) is the location index, \(M_{u,v}\) is the IGM value in location \((u, v)\). \(\Omega_{i,j}^t\) is the local neighborhood of \((i, j)\) \((t\text{-th} \text{ dominant orientation})\), containing \(n\) pixels, e.g. \(n = 7 \times 7 = 49\).
3. Run BGP on the OIGM images to yield a set of BGPM images
4. **return** A set of BGPM images.

In the BGPM, the strength of edge information is enhanced by using the IGM image. It produces stronger orientation from different discrete dominant orientations and further enforces spatial locality by using the average IGM values. Effectively, BGPM gains greater discriminant ability from these enhancements, with a small increase in complexity. BGPM often achieves high performance in
generally low complexity, different from the Gabor representation that would require a larger number of Gabor faces (e.g. typically 40) and a larger convolution neighborhood (e.g. $31 \times 31$). Typical number of OIGM and its local resolution are 3 and $7 \times 7$, respectively, leading to only 147 additional computational units (less than 0.4% of Gabor faces) and 3 times of the dimensions (compared to 40 times in Gabor-based fusion models).

6. Robust Face Recognition

We systematically evaluated the performance of BGP-based descriptors for facial representation and their robustness against multiple variations in illumination, expression, occlusion and age. Two groups of experiments were conducted. First, the performance of the BGP was compared to the LBP and Gabor features, together with discussions on parameter selections. Second, the capability of the BGP and BGPM descriptors was further evaluated and compared with recent methods on five benchmark databases: the AR [41, 42], Extended YaleB [35, 43], CMU Multi-PIE [44], FERET [45] and Labeled Faces in the Wild (LFW) [46]. For unbiased and fair comparisons, all implemented methods operated directly on the raw face images without any pre-processing, such as DoG filtering, Gamma correction and lighting equalization, some of which could prominently affect the results.

1). The Extended YaleB database contains about 22000 face images of 38 subjects with 9 different poses and 64 illumination conditions for each subject. A widely used subset [35, 43], which includes all faces from the frontal pose ($64 \times 38 = 2432$), was employed in the experiments on illumination variations.
The dataset was divided into five different groups with increasing effect of illumination according to [43]. Exemplar faces are shown in Fig. 10 (a)-(c). Note that the Extended YaleB is different from the original Yale B dataset [35], which only has 10 subjects.

2). The AR database consists of over 4000 images of 126 subjects, each having 26 facial images taken in two different sessions separated by two weeks. Each session has 13 images with multiple variations in expression, lighting and occlusion (sun glasses and/or scarf). A subset of cropped faces (by its original authors [42]) of 50 male and 50 female subjects was used in the experiments. Examples of these variations are shown in Fig. 10 (d)-(h).

3). The CMU Multi-PIE face database [44] contains 337 subjects with over 750,000 images captured in four sessions. The images vary significantly and have 20 illumination conditions, 6 facial expressions and 15 different view points.

4). The FERET database [45] has five subsets, including a gallery set (Fa) and four probe sets (Fb, Fc, DupI and DupII). The gallery set contains one frontal image for each of 1196 subjects. The DupI and Dup II sets, including 722 and 234 face images respectively, have been proven extremely challenging due to large appearance variations caused by aging. The experiments were conducted on both challenging sets. Following the existing approaches, we cropped the original images into smaller sizes (140 × 120) according to the available eye’s coordinates, but without any further pre-processing.

5). The LFW dataset [46] contains 13233 face images of 5749 people, collected in unconstrained environments from the web. They contain large real-world variations in expression, lighting, pose, age, and also in image scale and quality. The evaluation followed the standard image-restricted test model, verifying whether a pair of faces are from a same person. Our methods were evaluated on the widely used View 2 set, containing ten non-overlapping subsets, each having 600 pairs of images (300 matching and 300 non-matching pairs). Following the previous work [37, 12], we cropped out the main face area of size 150 × 80 from the images provided by Wolf et al [48].
6.1. BGP, LBP and Gabor Representations

We investigated the performance of these three fundamental descriptors for face recognition. The proposed BGP and the rotational invariant uniform LBP \cite{16} were implemented in local histogram model as \cite{36}. There are only two parameters for both methods, spatial resolution, \((P,R)\), and number of blocks, \(N_{blk} \times N_{blk}\). The nearest neighbor (NN) classifier was used to assign the probe to the most similar subject in the gallery. Similarities between feature vectors were computed by histogram intersection for LBP and BGP, and by Euclidean distance for Gabor representation \cite{17}.

Their performances were evaluated on the YaleB and the AR databases. On the YaleB, a single face with natural illumination condition (“A+000E+00”) per subject was used as the gallery image. All five groups with different levels of illumination effects were tested. For the AR, the gallery images were the natural faces from the session one (also a single image per subject), and the probe images were grouped as expression, lighting, sunglass & lighting and scarf & lighting, each of 600 images. The results are presented in Fig. 11.

The recognition rates of all three methods reached 100% in Group one and Group two of the YaleB database. Fig. 11 (top row) shows that, LBP and Gabor
descriptors had similar performances, reasonable for Group three with medium illumination effect, but deteriorated drastically with increased illumination effects in Groups four and five. In contrast, BGP were consistently excellent (with recognition rate above 90%) even with extreme lighting conditions (Group five). The improvements of BGP in Groups four and five were significant, outperforming LBP and Gabor by over 30% and 70%, respectively.

Similarly, BGP was the best performer in all test groups on the AR database (shown in Fig. 11 (bottom row)). It yielded 90% or higher in recognition rates for the groups of expression, lighting, and sunglass & lighting, and above 80% for the groups severely affected by both large-scale occlusions and illuminations. The performances of LBP and Gabor were substantially affected by multiple variations in the last two groups. The excellent results of BGP demonstrate its enhanced discrimination and robustness to multiple facial variations.

It has also been found that the BGP descriptor was fairly insensitive to the choice of its parameters. The overall performance of BGP was stable in different spatial resolutions, especially for (16, 2) and (24, 3). By contrast, changes in spatial resolution caused large differences in recognition rate of the LBP. The performance of both LBP and BGP descriptors can be improved by increasing the number of blocks. As can be seen, in most test groups, the recognition rates of BGP became stable when the number of blocks was equal to or greater than 12 x 12, except for Groups four and five of the YaleB, which required larger numbers of blocks to alleviate the effect of severe illumination conditions. Therefore, by trading off performance and computational complexity, the spatial resolution of the BGP was set to (16, 2) in all our experiments, while the number of blocks depends on the size of images.

6.2. Lighting and Multiple Variations

The efficiency of BGP and its enhanced BGPM descriptors was further evaluated by comparing with recent methods, including IGO-based methods (Gradientfaces [9] and IGOPCA [10]), local feature methods (CS-LBP [33], WLD [20, 49], LTP [21], POEM [12] and Volterrafaces [37]), and fusion of both
For BGPM, the number of OIGM and local resolution were optimally set to 3 and $7 \times 7$ in all experiments. For a fair and unbiased comparison, all implemented methods employed their optimal parameters and similarity measures suggested by the original authors. IGOPCA verified the number of reduced dimensions from 10 to its maximum number. The implementation of WLD was suggested in [49], with the number of quantized orientations set to 8 and differential excitation value varied among $\{32, 48, 64\}$. The CS-LBP was implemented with 8 neighbors with radius of 2 and binary threshold of 0.01, as suggested in [33]. The number of bins for local IGO-based methods (PHOG and LGOBP) was verified among $\{4, 40\}$. All local histogram methods (BGP, BGPM, CS-LBP, WLD, LTP, POEM, and LGOBP) were run by varying the numbers of blocks from $8 \times 8$ to $36 \times 36$. The pyramid level for the PHOG was optimized from 1 to 5. Finally, the best performance of each method, computed on three widely-used similarity measures: histogram intersection, $\chi^2$ [21] and Euclidean distance, was reported. The BGP methods used histogram intersection.

### 6.2.1. Illumination Variation

The illumination invariance was evaluated on the Extended YaleB with the similar schemes as in section 6.1. To provide more comprehensive results, the comparisons also included a group of methods based on the reflectance model specially developed to address illumination effect. These methods include logarithm total variation (LTV) model [52], logarithmic wavelet transform (LWT) [53], multi-linear Principal Component Analysis on tensors-CT histograms (TCT-MPCA) [54] and reconstruction with normalized large- and small-scale feature images (RLS) [55]. The results are presented in Table 2.

As one can see, IGO-based methods yielded better overall performance than the intensity-based methods (local features and reflectance models). As expected, BGP-based methods achieved the best performance in all groups. Even the basic BGP outperformed all other methods, while the BGPM had the low-
Table 2: Performance of single training sample per person on YaleB database.

| Method     | Publication & Year | Group 3 | Group 4 | Group 5 | Avg. |
|------------|--------------------|---------|---------|---------|------|
| LTV        | TPAMI 2006         | 21.5    | 24.2    | 17.6    | 20.7 |
| LWT        | PR 2009            | 18.0    | 18.0    | 29.2    | 22.7 |
| TCT-MPCA [54] | BMVC 2009     | 5.8     | 39.9    | -       | 23.9 |
| RL55       | TIP 2011           | 14.0    | 14.7    | 15.2    | 14.7 |
| LGOBP      | ICIP 2009          | 13.4    | 48.3    | 67.9    | 47.3 |
| WLD        | TPAMI 2010         | 1.1     | 15.3    | 60.5    | 30.6 |
| LTP        | TIP 2011           | 2.4     | 16.2    | 39.5    | 22.4 |
| CS-LBP     | PR 2009            | 4.4     | 39.9    | 85.7    | 49.3 |
| PHOG       | AVHH 2007          | 6.1     | 54.2    | 72.9    | 51.5 |
| POEM       | TIP 2012           | 4.8     | 11.3    | 40.4    | 21.9 |
| Volterrafaces | TPAMI 2012       | 6.6     | 32.3    | 17.4    | 19.3 |
| Gradientfoaces | TIP 2009   | 8.4     | 12.6    | 17.2    | 13.4 |
| IGOPCA     | TPAMI 2012         | 10.6    | 12.0    | 28.2    | 18.5 |
| BGP        | Proposed           | 2.8     | 9.2     | 12.9    | 9.1  |
| BGPM       | Proposed           | 1.3     | 1.9     | 2.7     | 2.1  |

The average error rate at only 2.1%, which is about one tenth of the errors of other methods. Substantial improvements were gained in the severe illumination conditions, Groups four and five, with only 1.9% and 2.7% errors, respectively.

We also compared the BGPM with the IGOLDA, which generally outperforms the IGOPCA but requires at least two images per subject from training [10]. Therefore, we use the first group of images as the training and gallery images and test the remained groups. The parameters of the IGOLDA were optimized and error rates of 0.22%, 5.83% and 16.34% were obtained for Groups three, four and five, compared to the 0%, 0.19% and 0% error rates achieved by the BGPM. The improvements of the BGP-based representations over these existing methods are exceptional and significant.

The errors of LBP-based methods (e.g. WLD and LTP and CS-LBP) increased drastically under the extreme lighting conditions of Group five, where the proposed BGP methods consistently excelled. The results suggest that discarding gradient contrast yields stronger gray-scale invariance than discarding intensity contrast. Our methods were further compared with a recent approach based on gradients and edges [25], developed for dealing with extreme illuminations. Following the same experimental setting conducted on the original Yale B dataset (using a single face per subject as the gallery), BGP achieved accuracy of 97.5% while BGPM 100%, favorably against the best result (0.8% error rate).
Table 3: Performance of single training sample per person on AR database.

| Method     | Publication & Year | Error Rate (%) | N | E | L | GL | SL | Ave. |
|------------|--------------------|----------------|---|---|---|----|----|------|
| UP [57]    | PR 2010            |                | 23 | 39.7 | - | - | - | -   |
| DMMA [31]  | TPAMI 2013         |                | 12 | 36.3 | - | - | - | -   |
| LSRC-G [56]| TPAMI 2013         |                | 12 | - | - | - | - | -   |
| LGBP       | ICIP 2009          |                | 12 | 31.7 | 37.7 | 49.0 | 85.3 | 51.1 |
| WLD        | TPAMI 2010         |                | 7  | 20.7 | 24.3 | 27.3 | 21.9 |
| LTP        | TIP 2011           |                | 12 | 18.7 | 10.3 | 16.0 | 30.0 | 17.5 |
| CS-LBP     | PR 2009            |                | 3  | 18.3 | 11.3 | 16.3 | 25.7 | 16.8 |
| PHOG       | AVIR 2007          |                | 2  | 22.3 | 8.3 | 17.0 | 25.3 | 16.5 |
| Gradientfaces | TIP 2009      |                | 8  | 36.7 | 14.0 | 25.3 | 45.0 | 29.2 |
| IGOPCA     | TPAMI 2012         |                | 8  | 28.7 | 10.3 | 22.7 | 34.7 | 22.5 |
| BGP        | Proposed           |                | 3  | 17.7 | 9.4 | 13.3 | 25.0 | 15.3 |
| BGPM       | Proposed           |                | 2  | 14.7 | 3.3 | 13.0 | 10.3 | 9.7  |

* Only 80 subjects for test and the other 20 for training.

6.2.2. Multiple Variations

The robustness against multiple variations was analyzed on the AR database. The experiments were divided into two groups of different training schemes: a single training sample per person and multiple training samples per person.

*(1): A Single Training Sample Per Person*

We used one neutral face per subject (N) in the first session as the gallery image and tested all other faces, 4 remaining groups in session one (expression (E), lighting (L), sunglass & lighting (GL) and scarf & lighting (SL)) and 5 groups in session two (N, E, L, GL and SL). Results are presented in Table. 3

Recently published results achieved by the same experimental scheme are also included for comparison, such as DMMA [31] and ESRC-Gabor [56].
The local feature methods outperformed IGO-based holistic representations. Again, BGP-based methods had the best overall performance in all implementations. BGPM achieved the lowest error rates in all tests and the average error rates were less than 1% and 10% for sessions one and two, respectively, significantly surpassing the closest performances 3.5% by POEM and 16.5% by PHOG in sessions one and two, respectively.

It can be seen from the table that the main gains of the local feature methods (e.g. POEM and LTP) over IGO methods lie in the cases of Expression (E), Scarf & Lighting (SL), both of which cause large-scale local distortions and can lead to significant differences in performance of local features and holistic methods. By integrating two approaches, the BGPM has not only yielded substantially improved performances in single variations, but also maintained low error rates in cases affected by multiple variations.

We also compared BGPM with IGOLDA [10] by using two natural faces from both sessions as gallery/training images and the remaining as the test. The IGOLDA obtained 10.0% and 9.9% error rates, while BGPM had only 0.6% and 1.3% errors.

(2): Multiple Training Samples Per Person

The BGP-based methods were further compared with recently proposed local fusion models and SRC-based methods on their latest published results on the AR database, under the four different implementations used by these publications, which are described as follows. The results are shown in Table 4.

**Implementation A** scheme of [23] used two neutral faces from both sessions as gallery images and tested on all four variations including expression (E), lighting (L), sunglass & lighting (GL) and scarf & lighting (SL), each having six faces per subject. The BGP methods were evaluated against two fusion models integrating LBP and Gabor representations. The BGPM achieved perfect performance in the group of lighting, which was not reported in [23]. The largest improvement was in the group of sunglass & lighting, resulting in only 0.3% error for BGPM compared to 46.1% for the best of the compared methods.
**Implementation B** evaluated the performance on variations of expression & lighting (E&L) and occlusions. For E&L, seven faces per subject were used for training: one neutral, three expressions and three lighting faces from session one; and then the corresponding seven faces from session two were tested. For occlusions, eight faces per subject including two neutral and six expression faces from both sessions were used as gallery images, two faces of sunglasses or scarves in both sessions were tested. Although the SRC-based approaches achieved low error rates for variations in expression or lighting, their performances suffered seriously in large-scale occlusions such as scarves. A common remedy for mitigating this effect is to manually partition a face image into a number of regions, and discard the occluded parts. The GRRC with partitions improved the performance substantially with very low error rates of 2.7%, 0.0% and 1.0% \[58\]. The BGPM further exceeded these and yielded almost perfect performance with 0.3%, 0.0% and 0.0% error rates. Note that the performance of SRC-based approaches depends on a manual partition scheme, while the BGP-based methods are automatic.

**Implementation C** trained on seven non-occluded faces from session one (as in Implementation B) and tested on four sets of occluded faces. Each set contained three faces per subject, with sunglasses or scarves, including multiple effects by lighting, in session one or two. Four sets are indicated as GL[S1], GL[S2], SL[S1] and SL[S2] in the table.

**Implementation D** conducted three separate experiments according to \[59\]. The first one trained on seven non-occluded faces and one sunglass face (randomly chosen from three in session one) and tested on seven non-occluded faces from session two and the remaining five sunglass faces in both sessions. The second experiment applied the similar training/test scheme for the faces with scarf occlusions. The last one evaluated both sunglass and scarf occlusions by using nine faces for training (seven non-occluded faces plus one random sunglass and one scarf faces from session one) and seventeen faces for testing (including seven non-occluded faces from session two and the remaining five sunglass and five scarf faces). The listed results of the proposed methods were
Table 4: Comparisons with recent local fusion models (Implementation A[23]) and SRC-based methods (Implementation B[58], C[58] and D[59]) from published results.

| Method | Publication & Year | Error Rate (%) |
|--------|--------------------|----------------|
| Implement. A | | |
| LGBP-M [18] | ICCV 2005 | 13.9 - - 62.4 17.4 |
| LGBP-P [54] | BMVC 2009 | 14.1 - - 63.0 16.5 |
| GVLBP-M [23] | TIP 2010 | 9.4 - - 46.1 12.6 |
| GVLBP-P [23] | TIP 2011 | 8.9 - - 53.9 9.6 |
| BGP | Proposed | 2.5 0.5 1.0 4.7 |
| BGPM | Proposed | 2.2 0.0 0.3 1.2 |

| Implement. B | | |
| SRC [60] | TPAMI 2009 | 5.3 13.0(2.5) 40.5(6.5) |
| LRC [61] | TPAMI 2010 | 23.3 4.0 (- -) 74.0(4.5) |
| CESR [52] | TPAMI 2011 | - - 30.0(- -) - - (1.7) |
| CRCRLS [63] | ICCV 2011 | 6.3 31.5(8.5) 9.5 (5.0) |
| RSC [64] | ICCV 2011 | 10.0 - - - - |
| CRC [63] | CVPR 2012 | 4.1 - (1.5) - - (3.5) |
| GRRC | PR 2012 | 2.7 7.0 (0.0) 21.0(1.0) |
| BGP | Proposed | 2.0 0.3 1.0 |
| BGPM | Proposed | 0.3 0.0 0.0 |

| Implement. C | | |
| SRC [60] | TPAMI 2009 | 16.7 51.3 51.0 71.0 |
| CRCRLS [63] | ICCV 2011 | 22.0 47.7 55.3 70.7 |
| GRRC [58] | PR 2012 | 7.7 48.3 5.0 15.7 |
| BGP | Proposed | 0.0 6.0 2.3 11.0 |
| BGPM | Proposed | 0.3 2.3 0.3 4.3 |

| Implement. D | | |
| SRC [60] | TPAMI 2009 | 15.8 23.7 22.0 |
| LLC [56] | CVPR 2010 | 15.5 23.4 21.0 |
| CVPR [53] | CVPR 2012 | 14.6 15.6 18.4 |
| BGP | Proposed | 2.6 3.7 3.6 |
| BGPM | Proposed | 1.6 1.2 1.9 |

a The error rates presented in parentheses were achieved by using manually partition scheme.
b SIFT [34], and its extension, Partial-Descriptor-SIFT, were tested on this group with error rate of 6.1% and 4.5% in [67].

The BGP methods were tested on Implementations C and D with more complex variations such as sunglass & lighting, or scarf & lighting, which had been rarely evaluated by the SRC-based methods. These variations did not seem to hinder the capabilities of the BGP and BGPM, while the performance of SRC-based methods suffered badly. BGPM again achieved extremely low average error rates, 1.8% and 1.6% for implementations C and D, respectively, only a fraction of the average error rate of the best of the compared methods (19.2% by GRRC and 16.2% by LR).

We further investigated the performance of BGPM on the Multi-PIE dataset
Table 5: Performance on Multi-PIE with illumination, expression and pose variations (accuracy rate (%)). Compared results are directly cited from [68]

| Method     | 2     | 3     | 4     | 5     | 2     | 3     | 4     | 5     | 2     | 3     | 4     | 5     |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| NN         | 19.8  | 27.1  | 33.2  | 37.9  | 40.3  | 48.6  | 54.9  | 58.8  | 11.4  | 13.7  | 17.0  |
| LRC [61]   | 32.9  | 52.7  | 67.6  | 78.4  | 47.4  | 59.8  | 68.4  | 74.1  | 18.4  | 29.5  | 41.9  |
| SRC [60]   | 69.0  | 82.6  | 88.9  | 93.3  | 64.8  | 76.3  | 83.9  | 87.9  | 45.2  | 57.8  | 65.8  |
| ESRC [59]  | 66.5  | 79.9  | 86.6  | 91.5  | 78.2  | 87.5  | 92.6  | 93.6  | 57.3  | 69.0  | 73.0  |
| SSRC [69]  | 66.6  | 78.5  | 83.9  | 88.8  | 79.9  | 89.3  | 93.3  | 95.0  | 61.2  | 73.8  | 79.3  |
| LR [59]    | 66.3  | 79.4  | 86.1  | 91.2  | 61.0  | 75.7  | 84.3  | 88.6  | 46.6  | 59.4  | 66.2  |
| LRSI [59]  | 68.0  | 80.5  | 88.1  | 92.7  | 60.5  | 76.3  | 84.6  | 87.3  | 45.2  | 58.6  | 66.6  |
| SDR-SLR [68]| 72.4  | 85.1  | 89.8  | 93.8  | 85.3  | 93.6  | 96.6  | 98.0  | 67.1  | 81.4  | 87.2  |
| BGPM       | 100   | 100   | 100   | 100   | 90.2  | 94.4  | 97.4  | 99.6  | 69.1  | 91.1  | 95.2  |

with severe illumination, expression and pose variations [44]. Each image was divided into $16 \times 16$ blocks. By following exactly the experiments conducted in [68] for a fair comparison, we used all 105 subjects of all four sessions. Each image was cropped into $160 \times 130$, based on the eye locations provided by [70]. The cropped set was provided by the original authors of [68], and we followed their experimental settings as follows. 1). **Illumination**: for each subject, 2, 3, 4, 5 of 20 different illuminations were randomly selected for training, and the rest were used for testing. 2). **Expression**: 2, 3, 4, 5 of 6 different expressions (neutral, smile, surprise, squint, disgust and scream) were used for training and the remaining one was tested. 3). **Pose**: we used five different poses captured by the camera moving from +30 to -30 degree. 2, 3, 4 poses per subject were used for training and the remaining were tested. For the proposed methods, training images were directly used as the gallery without any further learning process. Also as in [68], for each experiment, the average value of ten different implementations was presented. Results are given in Table 5.

As can be seen, the proposed BGPM suppresses the best performing SDR-SLR [68] by a large margin in almost all cases. Again, BGPM showed strong capability for handling substantial illumination changes, consistent with the previous results on the Extended Yale B. We also tested with single illumination as gallery, and tests were carried out cross all 20 different illumination conditions. The BGP and BGPM consistently yielded high average accuracies of 98.2% and 100% respectively, favorable to that of the SDR-SLR [68], even though our methods were not specifically designed for handling pose variations.
6.3. Aging and Unconstrained Variations

The performance of the BGPM descriptor was further investigated on age changes (DupI and DupII of the FERET database) and unrestricted real-world images (the LFW database). The number of blocks was set to 16×16. Our experiments followed most of the previous work by using squared root of the features for representation and cosine distance for similarity measure [47, 12]. For an unbiased evaluation, our descriptors were fairly compared to a set of manually-designed features. In our implementation, we applied the Fisher’s Linear Discriminant Analysis (FDA) [4] for classification, trained on the gallery set of the FERET. The number of reduced dimensions by FDA was set to 1000. The recognition rates by the BGPM on two databases are compared to the recent published results in Table 6.

The results show that the BGPM achieved competitive performance over the recent descriptors with correct rates reaching 94.3% and 89.7% on DupI and DupII, respectively. The margins between the BGPM and the closest methods on the list are about 4% on both subsets, significant for this challenging dataset. Similarly, the proposed descriptor obtained 78.7% correct rate for face verification on the LFW database, exceeding the closest single descriptor (POEM) by 5% and the best multiple or fusion descriptors by about 4%.

6.4. Deep Learning Models

It is worth noting that the goal of the work was to develop an efficient descriptor for robust face representation and little attention was paid to classifiers. Recent high-performance systems often include multiple, complex pre- and/or post-processing steps, as well as certain machine learning models, which are particularly useful for improving performance in unconstrained conditions. For example, Chen et al. [81] exploited an face alignment algorithm to detect a number of facial landmarks in unconstrained face images. Then they developed multiple learning approaches and rotated sparse regression to extract a compact face representation, along with the use of a large amount of additional training data. Furthermore, it has been shown recently that deep learning models
Table 6: Performance on LFW for ageing and unconstrained variations (CR-Correct Rate).

| Method        | CR (%) | Method        | CR (%) |
|---------------|--------|---------------|--------|
| PS-SIFT [71]  | 61.0   | V1-like [73]  | 64.2   |
| Gabor-WPCA [72] | 78.8  | V1-like+ [73] | 68.1   |
| LBP-WPCA [12] | 79.4   | Gabor (C1) [74] | 68.4 |
| LGBP [18]     | 74.0   | LGBP-WPCA [75] | 83.8 |
| WHGPP [22]    | 79.5   | G-LDP [74]b  | 78.8  |
| G-LDP [74]b   | 78.8   | FPLBP [12]   | 67.5  |
| LGBP-WPCA [75] | 83.8  | TPLBP [47]   | 69.0  |
| Zou’s Result [19] | 85.0 | Comb. [47]a  | 74.5  |
| Tan’s Result [10] | 90.0 | SIFT [47]   | 69.9  |
| POEM-WPCA [12]c | 88.8  | POEM [12]   | 73.7  |
| IGOFCA [10]   | 88.9   | –             | –     |
| BGPM          | 94.3   | BGPM          | 78.7  |

Note:
- Fusion of multiple features: Gabor, LBP, FPLBP and TPLBP.
- The highest approximated results reported by curve in [74], using Gabor pre-processing.
- A pre-processing step was applied for getting higher performance, i.e. Gamma correction and DoG filter were used in [70], and Retina filtering was processed before POEM [12].

Table 7: Comparisons with state-of-the-art deep learning based approaches on the LFW. Only single model performances are compared.

| State-Of-The-Art Results | Our Results |
|--------------------------|-------------|
| DeepFace [77]            | 97.35%      |
| Wen et. al. [78]         | 97.37%      |
| DeepID2+ [79]            | 98.70%      |
| VGGface [80]             | 98.95%      |
| RGB                      | 98.43%      |
| BGP                      | 98.82%      |
| RGB+BGP                  | 99.32%      |

Trained on huge numbers of additional data can further boost the performance, surpassing human-level face verification on the LFW database [82, 77, 79, 78].

To verify the efficiency of the proposed BGP descriptor with such a deep learning model, we followed the experimental settings of [78] by using the same Convolutional Neural Network (CNN) architecture with single softmax loss. We also used the additional training data, including face images from CASIA-Webface [83], CACD2000 [84], and Celebrity+ [85] databases. After removing the images with same identities included in the LFW test set, we gathered about 700 thousands images of 17189 unique persons. As most commonly-used CNN architectures for face recognition are designed for three-channel image input, we followed this common setting by using three same BGP maps as the input to the CNN. The results are compared in Table 7. Our single model obtained an accuracy of 98.82% on the LFW. By combining the BGP and
RGB as the input, our four-channel representation further improved the accuracy to 99.32%, markedly improved over the result of using only RGB input, 98.43%. Our final result compares favorably against a number of current top-ranked performances on the LFW, including 97.35% of DeepFace [77], 97.37% of [78] (using softmax loss), and 98.70% of DeepID2+ (single model) [79], and 98.95% of VGGface [80]. This shows that a structured, robust descriptor can add additional benefits to deep learning on its great ability to extract more discriminant features. Note that developing advanced deep learning model is beyond the scope of this work. The focus here is the development of an efficient and robust descriptor to improve on the previous counterparts such as POEM [12], Gradientfaces [9] and those proposed in [47].

7. Conclusion

In this paper a novel framework for robust facial representation has been introduced. The proposed structural Binary Gradient Pattern (BGP) effectively enforces spatial locality in the gradient domain to achieve invariance against both illumination and local distortions. By encoding local structures in a set of binary patterns, the BGP descriptor is compact and computationally efficient. Analysis shows that the defined structural patterns work proficiently as oriented micro edge detectors and possess strong spatial locality and orientation properties, leading to effective discrimination. Furthermore, the BGP is generic and suitable for building fusion models. As an example, the enhanced BGPM descriptor has also been presented as the result of combining BGP and orientations of image gradients. Extensive justifications and experimental verifications demonstrate the efficiency of the BGP and BGPM descriptors, and their significant performance improvements in face recognition over the existing methods on a variety of robustness tests against variations in lighting, expression, occlusion and aging. Initial experiments on combining them with deep learning models also show that they can benefit the deep learning framework to further elevate the recognition performance. Future work will look into how well-defined
descriptors and deep learned features can work best in real applications.

References

[1] L. Wolf, L. Hassner, I. Maoz, Face recognition in unconstrained videos with matched background similarity, in: IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2011.

[2] W. Huang, H. Yin, On nonlinear dimensionality reduction for face recognition, Image and Vision Computing 30 (2012) 355–366.

[3] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3 (1991) 71–86.

[4] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman, Eigenfaces vs. fisherfaces: Recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (1997) 711–720.

[5] B. Schölkopf, A. Smola, K.-R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Computation 10 (1998) 1299–1319.

[6] S. T. Roweis, K. S. Laurence, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (2000) 2323–2326.

[7] H. Yin, W. Huang, Adaptive nonlinear manifolds and their applications to pattern recognition, Information Sciences 180 (2010) 2649–2662.

[8] W. Huang, H. Yin, Linear and nonlinear dimensionality reduction for face recognition, in: IEEE Intl Conf. on Image Processing (ICIP), 2009, pp. 3337–3340.

[9] T. Zhang, Y. Y. Tang, B. Fang, Z. Shang, X. Liu, Face recognition under varying illumination using gradientfaces, IEEE Trans. on Image Processing 18 (2009) 2599–2606.

[10] G. Tzimiropoulos, S. Zafeiriou, M. Pantic, Subspace learning from image gradient orientations, IEEE Trans. Pattern Anal. Mach. Intell. 34.

[11] B. Heisele, P. Ho, J. Wu, T. Poggiob, Face recognition: component-based versus global approaches, Computer Vision and Image Understanding 91 (2003) 6–21.

[12] N.-S. Vu, A. Caplier, Enhanced patterns of oriented edge magnitudes for face recognition and image matching, IEEE Trans. on Image Processing 21 (2012) 1352–1365.

[13] W. Huang, H. Yin, A dissimilarity kernel with local features for robust facial recognition, in: IEEE Intl Conf. on Image Processing (ICIP), 2010, pp. 3785–3788.

[14] B. Manjunath, W. Ma, Texture features for browsing and retrieval of image data, IEEE Trans. Pattern Anal. Mach. Intell. 18 (1996) 837–842.

[15] T. Ojala, M. Pietikäinen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, Pattern Recognition 29 (1996) 51–59.
[16] T. Ojala, M. Pietikäinen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 971–987.

[17] C. Liu, H. Wechsler, Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition, IEEE Trans. on Image Processing 11 (2002) 467–476.

[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: IEEE Intl Conf. Computer Vision (ICCV), 2005, pp. 786–791.

[19] J. Zou, Q. Ji, G. Nagy, A comparative study of local matching approach for face recognition, IEEE Trans. on Image Processing 16 (2007) 2617–2628.

[20] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikäinen, X. Chen, W. Gao, Wld: A robust local image descriptor, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 1705–1720.

[21] X. Tan, B. Bill Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, IEEE Trans. on Image Processing 20 (2011) 1635–1650.

[22] B. Zhang, S. Shan, X. Chen, W. Gao, Histogram of gabor phase patterns (hgpp): A novel object representation approach for face recognition, IEEE Trans. on Image Processing 16 (2007) 57–67.

[23] Z. Lei, S. Liao, M. Pietikäinen, S. Z. Li, Face recognition by exploring information jointly in space, scale and orientation, IEEE Trans. on Image Processing 20 (2011) 247–256.

[24] A. R. Rivera, J. R. Castillo, O. Chae, Local directional number pattern for face analysis: Face and expression recognition, IEEE Trans. on Image Processing 22 (2013) 1740–1752.

[25] O. Arandjelović, Gradient edge map features for frontal face recognition under extreme illumination changes, in: British Machine Vision Conference (BMVC), 2012.

[26] O. Arandjelović, Making the most of the self-quotient image in face recognition, in: IEEE Conf. Automatic Face and Gesture Recognition (FG), 2013.

[27] S. R. Dubey, S. K. Singh, R. K. Singh, Rotation and illumination invariant interleaved intensity order-based local descriptor, IEEE Trans. on Image Processing 23 (2014) 5323–5333.

[28] O. Arandjelović, R. Cipolla, Achieving robust face recognition from video by combining a weak photometric model and a learnt generic face invariant, Pattern Recognition 46 (2013) 9–23.

[29] H. Han, S. Shan, X. Chen, W. Gao, A comparative study on illumination preprocessing in face recognition, Pattern Recognition 46 (2013) 1691–1699.

[30] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, Y. Ma, Toward a practical face recognition system: Robust alignment and illumination by sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2012) 372–386.

[31] J. Lu, Y.-P. Tan, G. Wang, Discriminative multi-manifold analysis for face recognition from a single training sample per person, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013) 39–51.
[32] S. A. J. Winder, M. Brown, Learning local image descriptors, in: IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2007.

[33] M. Heikkilö, M. Pietikäinen, C. Schmid, Description of interest regions with local binary patterns, Pattern Recognition 42 (2009) 425–436.

[34] D. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2004) 91–110.

[35] A. Georgiades, P. Belhumeur, D. Kriegman, From few to many: Illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2001) 643–660.

[36] A. Ahonen, A. Hadid, M. Pietikäinen, Face description with local binary patterns: Application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (2006) 2037–2041.

[37] R. Kumar, A. Banerjee, B. C. Venuri, P. Hanspeter, Trainable convolution filters and their application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2012) 1423–1436.

[38] B. A. Olshausen, D. J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, Nature 381 (1996) 607–609.

[39] D. Marr, E. Hildreth, Theory of edge detection, The Royal Society of London. Series B, Biological Sciences (1934-1990) 207 (1980) 187–217.

[40] M. Yang, L. Zhang, Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition, in: Euro. Conf. Computer Vision (ECCV), 2010, pp. 448–461.

[41] A. M. Martinez, R. Benavente, The AR Face Database, CVC Technical Report No.24, 1998.

[42] A. M. Martinez, A. C. Kak, Pca versus lda, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2001) 228–233.

[43] K. C. Lee, J. Ho, D. Kriegman, Acquiring linear subspaces for face recognition under variable lighting, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 684–698.

[44] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, Multipie, Image Vis. Comput. 28 (2010) 807–813.

[45] P. J. Phillips, H. Moon, P. J. Rauss, S. Rizvi, The feret evaluation methodology for face recognition algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 1090–1104.

[46] G. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, Technical report, 07-49, University of Massachusetts, Amherst, 2007.

[47] L. Wolf, T. Hassner, Y. Taigman, Effective unconstrained face recognition by combining multiple descriptors and learned background statistics, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2011) 1978–1990.
[48] L. Wolf, T. Hassner, Y. Taigman, The one-shot similarity kernel, in: IEEE Intl Conf. Computer Vision (ICCV), 2009.

[49] G. Hermosilla, J. Ruiz-del Solar, R. Verschae, M. Correa, A comparative study of thermal face recognition methods in unconstrained environments, Pattern Recognition 45 (2012) 2445–2459.

[50] S. Liao, A. Chung, Face recognition with salient local gradient orientation binary patterns, in: IEEE Intl Conf. on Image Processing (ICIP), 2009, pp. 3317–3320.

[51] A. Bosch, A. Zisserman, X. Munoz, Representing shape with a spatial pyramid kernel, in: ACM International Conference on Image and Video Retrieval, 2007.

[52] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, T. S. Huang, Total variation models for variable lighting face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (2006) 1519–1524.

[53] T. Zhang, B. Fang, Y. Yuan, Y. Y. Tang, Z. Shang, D. Li, F. Lang, Multiscale facial structure representation for face recognition under varying illumination, Pattern Recognition 42 (2009) 251–258.

[54] J. A. Ruiz-Hernandez, J. L. Crowley, A. Meler, A. Lux, Face recognition using tensors of census transform histograms from gaussian features maps, in: British Machine Vision Conference (BMVC), 2009, pp. 1–8.

[55] X. Xie, W. S. Zheng, J. Lai, P. C. Yuen, C. Y. Suen, Normalization of face illumination based on large-and small-scale features, IEEE Trans. Pattern Anal. Mach. Intell. 20 (2011) 1807–1821.

[56] W. Deng, J. Hu, J. Guo, Extended src: Undersampled face recognition via intra-class variant dictionary, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2012) 1864–1870.

[57] W. Deng, J. Hua, J. Guo, W. Cai, D. Feng, Robust, accurate and efficient face recognition from a single training image: A uniform pursuit approach, Pattern Recognition 43 (2010) 1748–1762.

[58] M. Yang, L. Zhang, S. Shiu, D. Zhang, Gabor feature based robust representation and classification for face recognition with gabor occlusion dictionary, Pattern Recognition 45 (2012) 2445–2459.

[59] C.-F. Chen, W. C.-P., Y.-C. Wang, Low-rank matrix recovery with structural incoherence for robust face recognition, in: IEEE Intl Conf. Computer Vision and Pattern Recognition (CVPR), 2012.

[60] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2009) 210–227.

[61] I. Naseem, R. Togneri, M. Bennamoun, Linear regression for face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 2106–2112.

[62] R. He, W.-S. Zheng, B.-G. Hu, Maximum correntropy criterion for robust face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2011) 1561–1576.
[63] L. Zhang, M. Yang, X. Feng, Sparse representation or collaborative representation: Which helps face recognition?, in: IEEE Intl Conf. Computer Vision (ICCV), 2011.

[64] I. Theodorakopoulos, I. Rigas, G. Economou, S. Fotopoulos, Face recognition via local sparse coding, in: IEEE Intl Conf. Computer Vision (ICCV), 2011.

[65] M. Yang, L. Zhang, D. Zhang, S. Wang, Relaxed collaborative representation for pattern classification, in: IEEE Intl Conf. Computer Vision and Pattern Recognition (CVPR), 2012.

[66] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, Y. Gong, Locality-constrained linear coding for image classification, in: IEEE Intl Conf. Computer Vision and Pattern Recognition (CVPR), 2010, pp. 3360–3367.

[67] C. Geng, X. Jiang, Face recognition using sift features, in: IEEE Intl Conf. on Image Processing (ICIP), 2009, pp. 3313–3316.

[68] X. Jiang, J. Lai, Sparse and dense hybrid representation via dictionary decomposition for face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 37 (2015) 1067–1079.

[69] W. H. Deng, J. N. Hu, J. Guo, In defense of sparsity based face recognition, in: IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2013.

[70] L. E. Shafey, C. McCool, R. Wallace, S. Marcel, A scalable formulation of probabilistic linear discriminant analysis: Applied to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013) 1788–1794.

[71] J. Luo, Y. Ma, E. Takikawa, S. Lao, M. Kawade, L. L.-B., Person-specific sift features for face recognition, in: IEEE Intl Conf. on Acoustics, Speech and Signal Processing (ICASSP), 2007, pp. 593–596.

[72] W. Deng, J. Hu, J. Guo, Gabor-eigen-whiten-cosine: A robust scheme for face recognition, in: Analysis and Modelling of Faces and Gestures (AMFG), 2005, pp. 336–349.

[73] N. Pinto, J. DiCarlo, D. Cox, Establishing good benchmarks and baselines for face recognition, in: Faces in Real-Life Images Workshop in European Conference on Computer Vision, 2008.

[74] B. Zhang, Y. Gao, S. Zhao, J. Liu, Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor, IEEE Trans. on Image Processing 19 (2010) 533–544.

[75] H. V. Nguyen, L. Bai, L. Shen, Local gabor binary pattern whitened pca: A novel approach for face recognition from single image per person, in: Advances in Biometrics, 2009, pp. 269–278.

[76] X. Tan, B. Triggs, Fusing gabor and lbp feature sets for kernel-based face recognition, in: Analysis and Modelling of Faces and Gestures (AMFG), 2007, pp. 235–249.

[77] Y. M. R. M. Taigman, Y., L. Wolf, Deepface: Closing the gap to humanlevel performance in face verification, in: IEEE Intl Conf. Computer Vision and Pattern Recognition (CVPR), 2014, p. 1701–1708.
[78] Y. Wen, K. Zhang, Z. Li, Y. Qiao, A discriminative feature learning approach for deep face recognition, in: Euro. Conf. Computer Vision (ECCV), 2016.

[79] Y. Sun, X. Wang, X. Tang, Deeply learned face representations are sparse, selective, and robust, in: IEEE Intl Conf. Computer Vision and Pattern Recognition (CVPR), 2015.

[80] V. A. Parkhi, O.M., A. Zisserman, Deep face recognition, in: 27th British Machine Vision Conference (BMVC), 2015.

[81] D. Chen, X. Cao, F. Wen, J. Sun, Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification, in: IEEE Computer Vision and Pattern Recognition (CVPR), 2013.

[82] C. Lu, X. Tang, Surpassing human-level face verification performance on lfw with gaussianface, in: The 30th AAAI Conference on Artificial Intelligence (AAAI), 2015.

[83] D. Yi, Z. Lei, S. Liao, S. Z. Li, Learning face representation from scratch, arXiv preprint arXiv:1411.7923.

[84] C. C. Chen, B.C., W. Hsu, Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset, IEEE Trans. on Multimedia 17 (2015) 804815.

[85] Z. Liu, P. Luo, X. Wang, X. Tang, Deep learning face attributes in the wild, in: IEEE Intl Conf. Computer Vision and Pattern Recognition (CVPR), 2015, p. 3730–3738.