



# Weighted statistical binary patterns for facial feature representation

Hung Phuoc Truong<sup>1</sup> · Thanh Phuong Nguyen<sup>2</sup> · Yong-Guk Kim<sup>1</sup>

Accepted: 26 April 2021 / Published online: 31 May 2021  
© The Author(s) 2021

## Abstract

We present a novel framework for efficient and robust facial feature representation based upon Local Binary Pattern (LBP), called Weighted Statistical Binary Pattern, wherein the descriptors utilize the straight-line topology along with different directions. The input image is initially divided into mean and variance moments. A new variance moment, which contains distinctive facial features, is prepared by extracting root  $k$ -th. Then, when Sign and Magnitude components along four different directions using the mean moment are constructed, a weighting approach according to the new variance is applied to each component. Finally, the weighted histograms of Sign and Magnitude components are concatenated to build a novel histogram of Complementary LBP along with different directions. A comprehensive evaluation using six public face datasets suggests that the present framework outperforms the state-of-the-art methods and achieves 98.51% for ORL, 98.72% for YALE, 98.83% for Caltech, 99.52% for AR, 94.78% for FERET, and 99.07% for KDEF in terms of accuracy, respectively. The influence of color spaces and the issue of degraded images are also analyzed with our descriptors. Such a result with theoretical underpinning confirms that our descriptors are robust against noise, illumination variation, diverse facial expressions, and head poses.

**Keywords** Local binary patterns · Completed LBP · Statistical moments · Facial feature representation

## 1 Introduction

Artificial intelligence has been developing rapidly with many real-world applications such as time series prediction [24], image classification [40, 46], and smart cities [28]. Among them, personal identification using biometric traits is a hot trend nowadays and has received increasing attention in the computer vision community. With biometric characteristics, the face image can be easily obtained from the camera as a non-invasive acquisition process. Therefore, face recognition can be widely applied to public

environments such as video surveillance, criminal detection, access control system, mobile device security, etc. [9]. Although diverse methods for face recognition have been introduced [18, 26, 36], they still have shortcomings. For this reason, face recognition is a challenging topic. Figure 1 shows several face recognition challenges, such as facial expression, head pose, illumination, and background complexity. Also, it has other difficulties, including occlusion, aging, makeup, image quality, etc. These challenges are formidable to deal with well.

A face recognition application typically consists of face detection, feature extraction, and classification. In general, the feature extraction stage plays a vital role because it will fail to achieve decent results when the employed feature descriptor is not adequate. Indeed, most well-known methods have robust feature descriptors, highly discriminative, and robust to extrinsic changes. In recent years, most face recognition algorithms, which have been studied extensively in addressing robust and discriminative descriptors, focus on three primary techniques: holistic, local, and hybrid models [23]. The holistic approach exploits the entire face and projects it into a small subspace such as Eigenfaces in manifold space [45], Fisherfaces [16, 33]. The local approach considers certain facial features

✉ Thanh Phuong Nguyen  
thanh-phuong.nguyen@univ-tln.fr

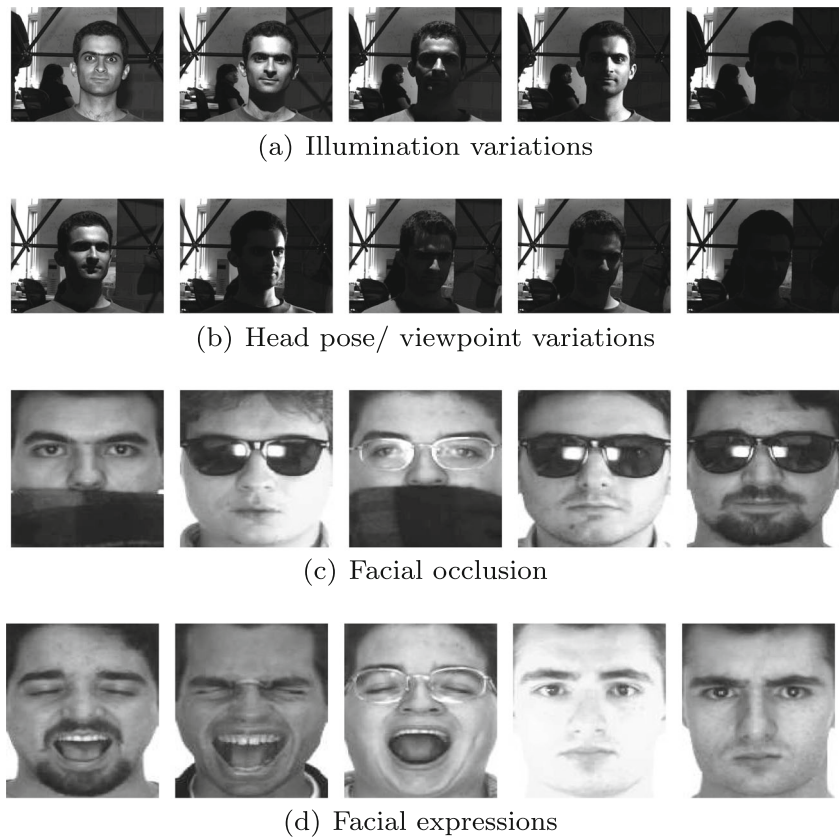
✉ Yong-Guk Kim  
ykim@sejong.ac.kr

Hung Phuoc Truong  
tphung@sju.ac.kr

<sup>1</sup> Department of Computer Engineering, Sejong University, Seoul, Korea

<sup>2</sup> CNRS, LIS, Université de Toulon, Aix Marseille Université, Marseille, France

**Fig. 1** Representative challenges in recognizing face. **a** illumination variations **b** head pose/viewpoint variations **c** facial occlusion **d** facial expressions



such as Speed-up robust features (SURF) [17], Local Binary Patterns (LBP) [22]. The local information combines with the holistic information to enrich feature descriptors for performance improvements in the hybrid approach: the fusion of 54 Gabor functions and fuzzy logic for facial expression recognition [15], two-color local descriptors, called Color ZigZag Binary Pattern (CZZBP) [19], or a fusion of Deep features [12].

Thanks to the low computational cost and efficient feature extraction capability, the LBP-based methods have been studied and widely applied to many tasks such as face recognition, facial expression classification, or texture classification. A large number of the LBP variants and hybrid models based on LBPs have been introduced [1, 36] for face recognition. However, they still have some drawbacks, such as noise sensitivity, contrast information, or illumination variation. This paper proposes a weighting statistical binary pattern framework that can improve the local descriptor in terms of discriminative power and robust against noise and illumination variation.

This work is extended from our prior efforts where we consider neighborhoods in straight-line topology [44] to utilize more useful information for local feature descriptors by statistical binary patterns [14, 37]. In this way, the proposed framework firstly considers two statistical moments (mean and variance) for noise elimination and

obtain complementary information. Then, the proposed LBP variant is applied to the first-moment image for LBP representations. The second-statistical moment image is a complementary component for building the weighted histogram to incorporate each pattern contribution. This proposed framework can enrich local descriptors by utilizing both moments without increasing the fused histogram dimension. The present study addresses prior shortcomings and proposes an upgraded descriptor for face recognition. The contributions of it are given as follows:

- We present a straight-line topology approach with LBP by direction (known as  $LBP_{\alpha}$ ), which is robust against several visual challenges, such as noise, illumination, and facial expressions, as a base foundation.
- Then, we propose a novel complementary LBP variant (known as  $CLBP_{\alpha}$ ), which is inspired by the local difference magnitude-sign transform to complement information for the local descriptor.
- To extract more robust descriptors from salient information in statistical moments, we propose the fused histogram of  $CLBP_{\alpha}$ , that is constructed by using  $WSBP_{\alpha}$  to obtain enriched features.
- A comprehensive evaluation of six public datasets suggests that our proposed framework outperforms the state-of-the-art methods.

The paper is organized as follows. Section 2 prepares some background on LBP. Section 3 details the proposed framework. In section 4, we analyze the implementations through several parameter settings for evaluations. Experimental results are interpreted in Section 5. A discussion for our proposed framework is analyzed in Section 6, and the last one consists of our conclusion and future works.

## 2 Related works

Many methods based on basic LBP descriptors, that can encode the local appearance by the relation between neighborhoods, have been introduced. However, there exist several shortcomings, such as local information loss or sensitivity to noise. Diverse LBP variants have been proposed to address these shortcomings. Several neighborhood topologies or encoding operators have been introduced, such as Dominant Rotated Local Binary Patterns (DRLBP) [32] and Enhanced Line Local Binary Pattern (EL-LBP) [44].

Recently, several hybrid models based on LBP-like descriptors for face analysis have been examined and proved to have highly discriminative power [22]. Lin et al. [27] proposed a fast algorithm, called LBP edge-mapped descriptor, which was to fuse LBP and SIFT using the maxima of gradient magnitude points on the image to illustrate facial contours for face recognition. Ding et al. [11] introduced the Dual-Cross Patterns (DCPs) as a core algorithm to extract facial features at both the holistic and component levels of a human face, then applied the first derivative of Gaussian for eliminating the differences of illumination. The Multi-scale block Local Multiple Patterns (MB-LMP) [49] exploited multiple feature maps based on the modified Weber’s ratio, then fused the histograms of non-overlapping patches for more robust features. Kas et al. [21] addressed shortcomings of previous LBPs and proposed Mixed Neighborhood Topology Cross Decoded Patterns (MNTCDP) by considering multi-radial and multi-orientation information simultaneously to exploit the relationship between the referenced point and its neighbors on each  $5 \times 5$  pixel block. Inspired by LBP-like in face recognition, Shu et al. [43] proposed Equilibrium Difference LBP (ED-LBP) in multiple color channels (RGB, HSV, YCbCr) accompanied with an SVM classifier for face spoofing detection. Unlike the traditional LBP circle, the Local Diagonal Extrema Number Pattern (LDENP) [42] descriptor only encoded information within the local diagonal neighbors using the first-order local diagonal derivatives to obtain a compact description for face recognition. Deng et al. [10] proposed an accurate face recognition by exploiting the compressive binary patterns (CBP) on a set of first six random-field eigenfilters,

which reduced the bit error rate of LBP-like descriptor and were more robust against additive Gaussian noise. According to LBP, another approach encoded information by examining neighboring pixels at different distances across different derivative directions called Local Gradient Hexa Pattern (LGHP) [6] which generated discriminative inter-class facial images. Lu et al. [29] proposed an unsupervised feature learning to represent face images from raw pixels and jointly encoded codebook for small regions to obtain high discrimination in descriptors, called Simultaneous Local Binary Feature Learning and Encoding (SLBFLE).

The other aspect was to utilize more useful information for descriptors to overcome local information loss within images. For instance, the Completed LBP technique (CLBP) [14] described local difference Sign-Magnitude transform to obtain higher performance. Another improvement of CLBP, i.e. the statistical binary patterns model [37], was built on several statistical moments for robust descriptors and improved the performance.

### 2.1 LBP

LBP was first introduced by Ojala et al. [38]. The LBP feature describes the spatial relationship in an image by encoding the neighbor points of a given central point. Let  $f$  be an 2D discrete image in  $\mathbb{Z}^2$  space. Then, the LBP encoding of  $f$  can be considered as a mapping from  $\mathbb{Z}^2$  to  $\{0, 1\}^P$ :

$$LBP_{P,R}(f)(\mathbf{c}) = \sum_{p=0}^P s(f(\mathbf{g}_p) - f(\mathbf{c})), \quad \text{with } s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

and  $\mathbf{g}_p$  are intensity of  $P$  neighbors and are measured on the circle of central point  $\mathbf{c}$  and radius  $R$ .

The dimension of LBP descriptor can be reduced by considering its uniform patterns, whose values  $U(LBP_{P,R}) \leq 2$  and defined by the following equation:

$$U(LBP_{P,R}) = \sum_{p=1}^P |LBP_{P,R}^p - LBP_{P,R}^{p-1}| \tag{2}$$

where  $LBP_{P,R}^p$  is the  $p$ -th bit of  $LBP_{P,R}$ , and  $LBP_{P,R}^p = LBP_{P,R}^0 \cdot LBP_{P,R}^{u2}$  [38] was a very robust and reliable descriptor for face representation or texture classification. As a result, the mapping from  $LBP_{P,R}$  to  $LBP_{P,R}^{u2}$  produces  $L = P(P - 1) + 3$  distinct output values by building a lookup table of  $2^P$  patterns. Therefore, the local descriptor is described as follows:

$$\mathbf{H} = [H_0, H_1, \dots, H_{L-1}]^T \tag{3}$$

where

$$H_t = \sum T\{\text{LBP}_{P,R}(x, y) = t\}, \text{ and } T\{A\} = \begin{cases} 1, & \text{if } A \text{ is true} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

in which  $H_t$  is the occurrence of the  $t^{\text{th}}$   $\text{LBP}^{u2}$  code, where  $t \in [0..L-1]$ . Therefore, the length of histogram in uniform LBP representation is  $L = P(P-1) + 3$ .

### 2.2 Completed LBP

Guo et al. [14] considered a local difference sign-magnitude transform and proposed a completed model as a state-of-the-art variant of LBP. The transform  $d_p = s_p * m_p$  consists of two components, i.e. signs:  $s_p = \text{sign}(d_p)$  and magnitudes:  $m_p = |d_p| = |f(\mathbf{g}_p) - f(\mathbf{c})|$ . By adding components to them, three operators, called CLBP-Sign (CLBP\_S), CLBP-Magnitude (CLBP\_M), and CLBP-Center (CLBP\_C), were designed to code three features S, M, and C. The first operator CLBP\_S was the same as the original LBP operator and produced S component. The M component which expresses the local variance of magnitude should be consistent with S and defined as follows:

$$\text{CLBP\_M}_{P,R}(f)(\mathbf{c}) = (s(m_p - \bar{m}))_{0 \leq p < P} \quad (5)$$

where  $\bar{m}$  is the mean value of  $m_p$  from the whole image. Moreover, the last component C also carries discriminant information. Therefore, the CLBP\_C operator is formulated:

$$\text{CLBP\_C}(f)(\mathbf{c}) = s(f(\mathbf{c}) - \bar{f}) \quad (6)$$

where  $\bar{f}$  is set as the mean gray level of the whole image. Because of complementary relationship between these operators, it turns out that the Completed LBP descriptor is useful for the texture classification task.

### 2.3 Face representation based on LBPs

The face representation based on LBP descriptors has been first introduced by Ahonen et al. [1] by analyzing small local regions in the face instead of striving for a holistic facial texture representation. In such a local approach, a face image is partitioned into  $m$  non-overlapping patches  $R^{(j)}$  ( $j = 1..m$ ) where an LBP operator is independently applied to produce local histograms. It aims to fuse all LBP histograms as a single vector (also known as local LBP descriptors) for facial texture representation. The concatenation approach is a simple and efficient one for LBP description. Each LBP histogram  $H^{(j)}$  by each image patch  $R^{(j)}$  is computed by (3). Finally, the global LBP descriptor for all patches  $R^{(j)}$  is formulated as follows ( $T$  is the transpose operator):

$$\mathbf{H} = [(\mathbf{H}^{(1)})^T (\mathbf{H}^{(2)})^T \dots (\mathbf{H}^{(m)})^T]^T \quad (7)$$

The resulting feature vector has the size of  $m \times n$ , where  $n$  is the length of LBP histogram along with its topology. Therefore, this approach for face representation is more robust under variations such as poses or illumination. Notably, small patches within an image can be of different sizes or overlapping regions. Many face recognition works have followed the local approach and obtained significant LBP variants [5, 42, 47, 49].

### 2.4 Statistical moment images

Since we define that  $f$  is a 2D discrete image in  $\mathbb{Z}^2$  space, we can obtain a real-valued image in  $\mathbb{R}$  by a mapping technique. The spatial support, which is employed to compute the local statistics, is modeled as  $\mathcal{B} \subset \mathbb{Z}^2$ , such that  $\mathcal{O} \in \mathcal{B}$ , where  $\mathcal{O}$  is the origin of  $\mathbb{Z}^2$  [37]. Figure 2 illustrates how to construct a spatial support  $\mathcal{B}$ .

The  $r$ -order moment image associated to  $f$  and  $\mathcal{B}$  is also a mapping from  $\mathbb{Z}^2$  to  $\mathbb{R}$ , defined as

$$m_{(f,\mathcal{B})}^r(\mathbf{c}) = \frac{1}{|\mathcal{B}|} \sum_{\mathbf{b} \in \mathcal{B}} (f(\mathbf{c} + \mathbf{b}))^r \quad (8)$$

where  $\mathbf{c}$  is a pixel from  $\mathbb{Z}^2$ , and  $|\mathcal{B}|$  is the cardinality of the structuring element  $\mathcal{B}$ . Accordingly, the  $r$ -order centered moment image ( $r > 1$ ) is defined as

$$\mu_{(f,\mathcal{B})}^r(\mathbf{c}) = \frac{1}{|\mathcal{B}|} \sum_{\mathbf{b} \in \mathcal{B}} (f(\mathbf{c} + \mathbf{b}) - m_{(f,\mathcal{B})}^1(\mathbf{c}))^r \quad (9)$$

where  $m_{(f,\mathcal{B})}^1(\mathbf{c})$  is the average value (1-order moment) calculated around  $\mathbf{c}$ . Finally, the  $r$ -order normalized centered moment image ( $r > 2$ ) is defined as

$$\beta_{(f,\mathcal{B})}^r(\mathbf{c}) = \frac{1}{|\mathcal{B}|} \sum_{\mathbf{b} \in \mathcal{B}} \left( \frac{f(\mathbf{c} + \mathbf{b}) - m_{(f,\mathcal{B})}^1(\mathbf{c})}{\sqrt{\mu_{(f,\mathcal{B})}^2(\mathbf{c})}} \right)^r \quad (10)$$

where  $\mu_{(f,\mathcal{B})}^2(\mathbf{c})$  is the variance (2-order centered moment) calculated around  $\mathbf{c}$ .

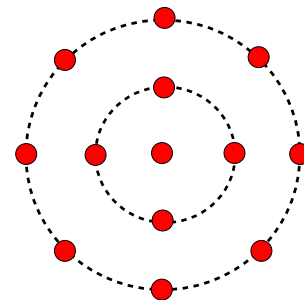


Fig. 2 Illustration of the spatial support with an example  $\mathcal{B} = \{(1, 4); (2, 8)\}$  designed as a collection of neighbor points sampled on different circles having the same center

### 3 Weighted Statistical Binary Patterns by direction $\alpha$ (WSBP $_{\alpha}$ )

We propose the Weighted Statistical Binary Patterns by direction  $\alpha$  (WSBP $_{\alpha}$ ) descriptor to enhance the discriminant capability of LBPs for face recognition while reducing its sensitivity to representative challenges, such as facial emotions, noise, or illumination. Such a descriptor can encode spatial information in a set of local statistical moment images and maps this coding to uniform LBP “u2” to produce a more compact descriptor. Because of complementary and consistent characteristics, two crucial components CLBP $_{S_{\alpha}}$  and CLBP $_{M_{\alpha}}$  (Here,  $\alpha$  is a given direction) are computed on a mean image and weighted by a variance image to improve the performance. The detail is given as follows.

#### 3.1 Local Binary Patterns by direction (LBP $_{\alpha}$ )

In the original LBP and several variants, the neighbors  $\mathbf{g}_p$  have the coordinate  $(R \cos(2\pi p/P), R \sin(2\pi p/P))$  lying on a circle of radius  $R$ . In the proposed LBP $_{\alpha}$ , we consider the relationship between pixels by a straight-line topology on a direction  $\alpha$ , given that the coordinate of  $\mathbf{c}$  is  $(0,0)$ . The neighbors of straight-line topology are defined as follows:

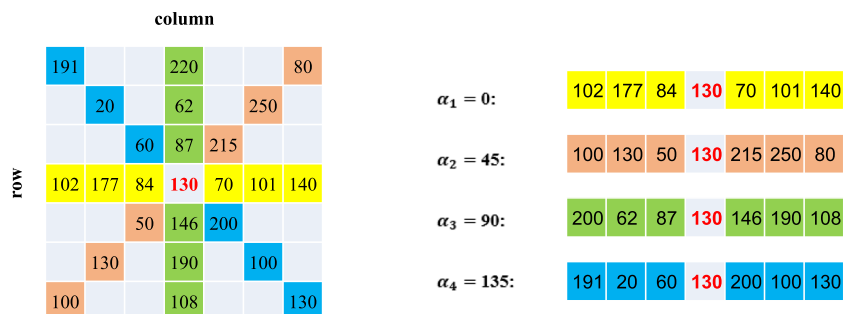
$$\mathbf{g}_{\alpha}^p = \left( \frac{2pR \cos \alpha}{P}, \frac{2pR \sin \alpha}{P} \right)_{-P/2 \leq p \leq P/2} \quad (11)$$

When considering a line topology, the number of neighbors should be an even number, and neighbors are bilateral symmetry with a central point  $\mathbf{c}$ . Figure 3 illustrates four LBP $_{\alpha_i}$  by considering 6 neighbors along with a line topology.

Similar to traditional LBP, we encode an image by LBP $_{\alpha}$  operator as defined in (1). Therefore, it can be expressed as follows:

$$\text{LBP}_{\alpha(P,R)}(f)(\mathbf{c}) = \sum_{p=0}^P s(f(\mathbf{g}_{\alpha}^p) - f(\mathbf{c})), \quad \text{with } s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

**Fig. 3** An example of several Local Binary Patterns by directions  $\{\alpha_i\} = \{0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}\}$ , where  $\{\text{line}_i\} = \{\text{yellow, orange, green, blue}\}$ , respectively



where  $\mathbf{g}_{\alpha}^p$  is defined in (11), and remaining variables such as  $f, c, P,$  and  $R$  are defined in (1). As a result, LBP $_{\alpha}$  operator, which produces  $2^P$  distinct patterns, leads to a huge descriptor. Inspired by the LBP uniform principle in Section 2.1, we reduce the number of patterns by considering the uniform patterns concept for LBP $_{\alpha}$ . After this process, the LBP $_{\alpha}$  “uniform patterns” have  $P(P-1)+3$  distinct output values from a lookup table of  $2^P$  values.

The main difference between the circle LBP and the LBP $_{\alpha}$  is that the LBP considers spatial relationship by a circle, whereas the LBP $_{\alpha}$  exploits spatial information in the straight-line of neighbors along with the given directions. Although several primary factors such as the direction of exposure, illumination, or facial expressions are given as challenges in face recognition, it turns out that the LBP $_{\alpha}$ -based representation is robust against changes of illumination and scale since it examines micro-patterns in a line topology. Moreover, by taking advantage of traditional LBP, the proposed LBP $_{\alpha}$  can characterize the distribution of local pixels by a direction, and the frequency of occurrences of LBP $_{\alpha}$  values can be used to represent various facial structures.

#### 3.2 Complementary Local Binary Patterns by direction $\alpha$ (CLBP $_{\alpha}$ )

The CLBP [14] had been used for texture classification by combining three operators CLBP $_{S}$ , CLBP $_{M}$ , and CLBP $_{C}$  in a joint or hybrid way. Similar to CLBP, we propose Complementary Local Binary Patterns by direction  $\alpha$  (CLBP $_{\alpha}$ ) which considers neighbors  $\mathbf{g}_{\alpha}^p$  by direction  $\alpha$  for the face recognition task. The proposed CLBP $_{\alpha}$  consists of two operators: CLBP $_{\alpha}$ -Sign (CLBP $_{S_{\alpha}}$ ) and CLBP $_{\alpha}$ -Magnitude (CLBP $_{M_{\alpha}}$ ). In general, the CLBP $_{S_{\alpha}}$  is similar to the proposed LBP $_{\alpha}$  described in (12). The CLBP $_{S_{\alpha}}$  operator describes the structure of image  $f$  with respect to the local relationship, whereas the CLBP $_{M_{\alpha}}$  complements local difference Magnitude and is in a consistent format with that of the CLBP $_{S_{\alpha}}$ . This operator is defined as follows:

$$\begin{aligned} \text{CLBP\_M}_{\alpha(P,R)}(f)(\mathbf{c}) &= (s(m_{\alpha}^p - \bar{m}_{\alpha}))_{0 \leq p < P}, \\ m_{\alpha}^p &= |d_p| = |f(\mathbf{g}_{\alpha}^p) - f(\mathbf{c})| \end{aligned} \quad (13)$$

where  $\bar{m}_{\alpha}$  is the mean value of  $m_{\alpha}^p$  for the whole discrete image  $f$ . Each component S and M has  $P(P - 1) + 3$  distinct values corresponding to the “uniform” LBP $_{\alpha}$  coding of discrete image  $f$ . Inspired by forming CLBP descriptors [14], we have two ways to combine different components for enhanced descriptors. The first descriptor CLBP\_S/M $_{\alpha}$ , which forms a joint 2D histogram from the CLBP\_S $_{\alpha}$  and CLBP\_M $_{\alpha}$  codes, has  $[P(P - 1) + 3]^2$  values. The second descriptor CLBP\_S.M $_{\alpha}$ , which concatenates two histograms together, has  $2[P(P - 1) + 3]$  values. The distribution of the first one can become too sparse when the dimension (i.e., the number of neighbors  $P$ ) increases. However, the marginal histogram of the second one obtains a reasonable size of  $2[P(P - 1) + 3]$ . As a trade-off between the performance and computational cost, the marginal histogram approach is utilized in our experiments. Note that component C, which expresses the local gray level in the image, is ignored in our proposed model. The proposed CLBP\_S.M $_{\alpha}$  produces more reliable and significant expressiveness for the facial feature representation.

### 3.3 Weighted Statistical CLBP by directions $\alpha_i$ (WSBP $_{\alpha_i}$ )

An introduction of two first-order moments (mean and variance moments) into an LBP-based operator was proposed as Statistical Binary Patterns (SBP) [37]. The first order, known as mean-valued moment  $m_1$ , gives the contribution of individual pixel intensity for the entire image. The second order, known as variance-valued moment  $\mu_2$ , is to find how each pixel varies from its neighboring pixels and represents salient regions in an image. Our proposed WSBP can build a novel histogram from CLBP $_{\alpha_i}$  descriptors by computing CLBP $_{\alpha_i}$  image on the first-order moment  $m_1$  and counting occurrences of every pattern on that CLBP $_{\alpha_i}$  image by a significance index corresponding to the salient regions using the new second-order moment  $\mu_2'$ . The proposed descriptor can discard the noise, illumination, or near-uniform regions. Figure 4 illustrates the flow diagram of our descriptor based on the WSBP $_{\alpha_i}$  descriptors. With the mean image  $m_1$ , the spatial relationship between local structures is represented using CLBP $_{\alpha_i}$  operator to obtain two essential components S $_{\alpha_i}$  and M $_{\alpha_i}$ . Then, each component (S $_{\alpha_i}$  and M $_{\alpha_i}$ ) obtained by CLBP $_{\alpha_i}$  operator is weighted by the contribution of every local pattern according to the new variance image  $\mu_2'$  for the weighting histogram.

Let H be the histogram vector of each component, and  $(x, y)$  be location of pixel in each component

of CLBP $_{\alpha_i(P,R)}$  image. Then, the histogram for each component is based on the contribution of every location (pixel) in new variance moment  $\mu_2'$ . Equation (4) defines the occurrence of every CLBP $_{\alpha_i(P,R)}$  code  $t^{th}$  as follows:

$$H_t = \begin{cases} \sum_{\forall(x,y)} \mu_2'(x, y), & \text{if CLBP}_{\alpha_i(P,R)}(x, y) = t \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

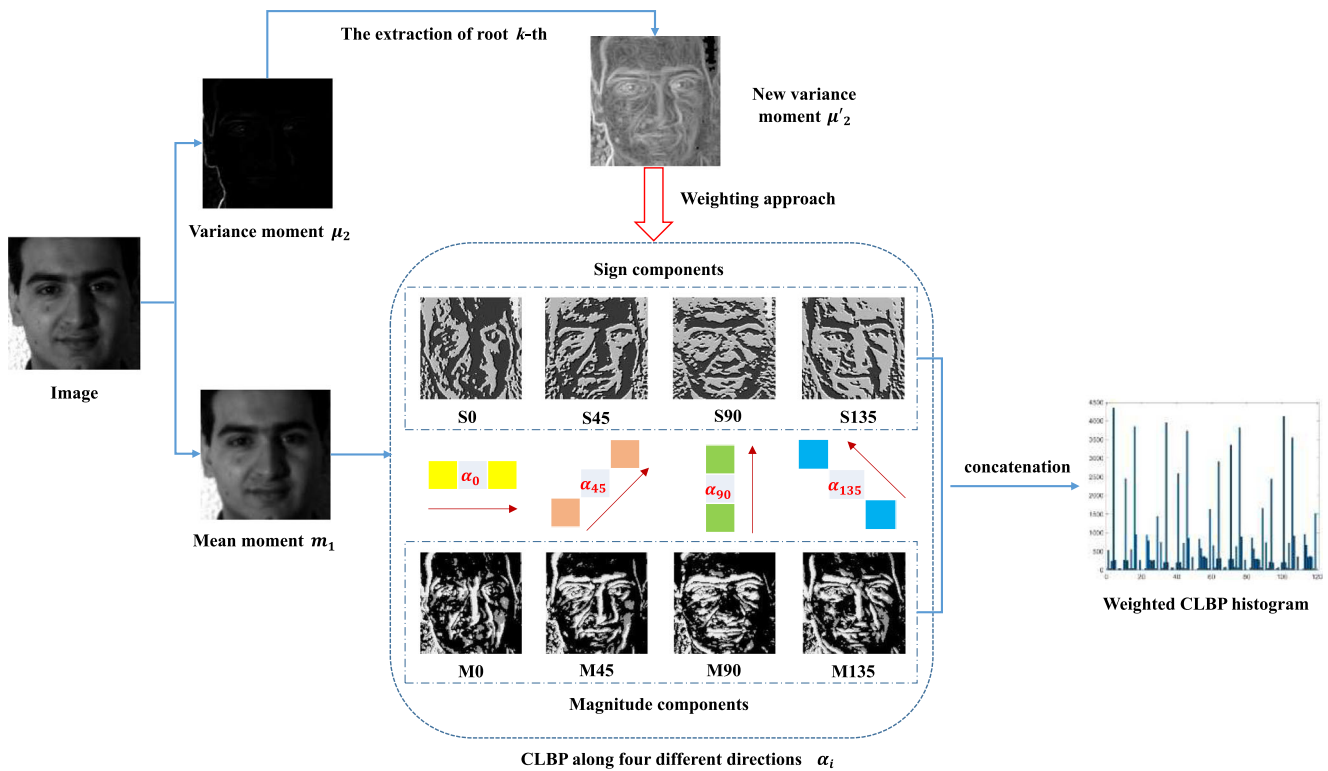
The SBP descriptor [37] produces enhanced descriptors and only considers all patterns having the same weights and ignoring their significance. In this paper, the WSBP $_{\alpha_i}$  descriptors capture the local relationships within images corresponding to the mean moment, and exploit contrast and gradient magnitude information through variance moment to enhance the local relationship description. Equation (14) describes how every pixel occurrence is weighted by its contribution corresponding to those pixels in a new variance moment  $\mu_2'$ . Therefore, the histogram of each component S $_{\alpha_i}$  and M $_{\alpha_i}$  has  $P(P - 1) + 3$  values. Finally, the dimensionality of WSBP $_{\alpha_i}$  descriptor is  $2[P(P - 1) + 3]$  because of the concatenation of histograms. As a result, the WSBP $_{\alpha_i}$  descriptor is not only compact but also robust to noise, illumination and other variations.

### 3.4 The computational complexity

In this section, we address the computational complexity of WSBP descriptor for an input image of size  $N \times N$ . Suppose that the pre-defined spatial support  $\mathcal{B}$  is defined as  $(R_1, P_1), (R_2, P_2)$ ; WSBP $_{\alpha}$  is calculated by considering  $P$  neighbors. The computational complexity of WSBP descriptor depends on the following factors.

- *Construction of moment images:* At each pixel, the mean value can be obtained after  $O(P_1 + P_2)$  operations, while the variance value requires  $O((P_1 + P_2)^2)$  operations. Therefore, the construction of moment images can be done in  $O((P_1 + P_2)^2 N^2) = O(N^2)$ .
- *Construction of CBLP $_{\alpha}$ :* CBLP $_{\alpha}$  consists of 2 components CBLP\_S $_{\alpha}$  and CBLP\_M $_{\alpha}$ . The first one is calculated in  $O(PN^2)$ . The second one has the same complexity of  $O(PN^2)$ . As a result, the complexity of CBLP $_{\alpha}$  is  $O(2PN^2) = O(N^2)$ .
- *Construction of WSBP $_{\alpha}$ :* WSBP $_{\alpha}$  addresses CBLP $_{\alpha}$  on mean image and considers variance image for constructing the weighted histogram. As mentioned above, each component can be done in  $O(N^2)$ .

Therefore, the computational complexity of WSBP is  $O(N^2)$ . It is evident that WSBP requires more calculation than LBP, but both are in the same computational



**Fig. 4** The flow diagram of our  $WSBP_{\alpha_i}$  descriptors. The input image is initially divided into mean ( $m_1$ ) and variance ( $\mu_2$ ) moments. A new variance moment ( $\mu'_2$ ), which contains distinctive facial features, is prepared by extracting root  $k$ -th. Then, when Sign and Magnitude components along four different directions using the mean moment

are constructed, a weighting approach according to the new variance is applied to each component. Finally, the weighted histograms of Sign and Magnitude components are concatenated to build a weighted CLBP histogram

complexity order. Such a constraint guarantees that our operator is effective as the non-LBP methods in terms of computation time.

or head pose variations. Such a WSBP can be expressed as follows:

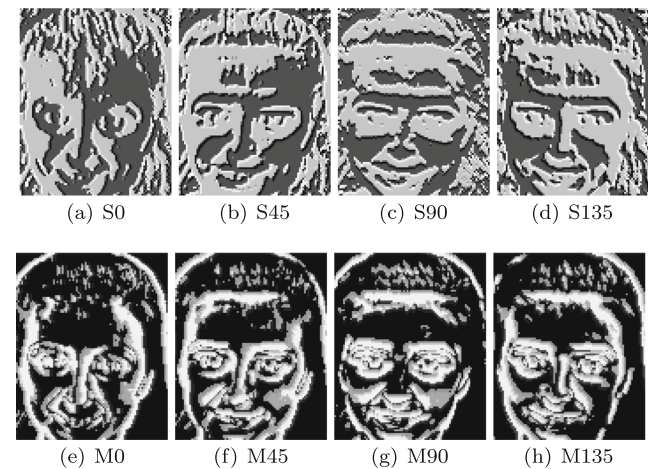
$$WSBP = WSBP_{\alpha_1} - WSBP_{\alpha_2} - WSBP_{\alpha_3} - WSBP_{\alpha_4}$$

### 4 Implementation

In this section, we detail the configuration of the WSBP descriptor.

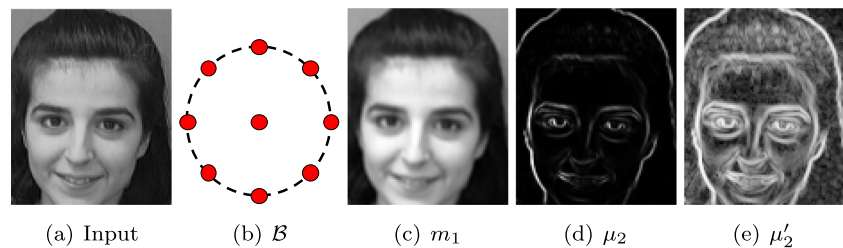
#### 4.1 The fusion of different descriptors $WSBP_{\alpha_i}$

Suppose that  $WSBP_{\alpha}$  considers only one direction ( $\alpha$  is a given direction), it could lead to an inadequate description simply because such a descriptor would exploit only the local relationship along that direction. What we aim here is that this descriptor should utilize every useful surrounded features. Inspired by LBP operators in a circle topology (with scale of  $(P, R) = (8, 1)$ ), we propose to consider at least four directions for the fused histogram,  $\alpha_i \in \{0^0, 45^0, 90^0, 135^0\}$  (see Section 5). Figure 5 shows components S and M of CLBP at four directions  $\{\alpha_i\}$  as four views of a given image. The fusion of four views could be an adequate descriptor in recognizing face against illumination



**Fig. 5** Illustration of  $CLBP_{\alpha_i}$  for component S (a, b, c, d) and M (e, f, g, h) with four different directions  $\{0^0, 45^0, 90^0, 135^0\}$ , respectively. Each  $CLBP_{\alpha_i}$  operator, consisting of two components, is computed on mean moment with a structuring element  $B = \{(1, 8)\}$

**Fig. 6** Illustration of **c** Mean ( $m_1$ ) and **d** Variance ( $\mu_2$ ) moments for **a** the given input using **b** the structuring element  $\mathcal{B} = \{(1, 8)\}$ . **e** A new variance ( $\mu'_2$ ) can be obtained by extracting root 9-th from  $\mu_2$



## 4.2 Moment parameters

For a successful implementation of our descriptor, a proper parameter setting has to be made. As a pre-processing step, the mean ( $m_1$ ) and variance ( $\mu_2$ ) moments obtained by computing the spatial support  $\mathcal{B}$  are used to reduce the noise sensitivity. Thus, moment parameters should be in the optimal settings for this purpose.

We define the structuring elements as a circle spatial support  $\mathcal{B} = \{(R_i, P_i)\}$ , such that ( $P_i$ ) is the number of neighbors and ( $R_i$ ) is its radii. Figure 6 shows an example of two-moment images using  $\mathcal{B} = \{(1, 8)\}$ . Given that the second-order moment (variance moment) tends to emphasize only dominant edges, some potentially important information could be discarded. To handle this problem, we propose to perform an extraction of root  $k$ -th for the variance moment as  $\mu'_2 = \sqrt[k]{\mu_2}$  ( $k \in [2, 16]$ ). For example, Fig. 6e shows the new variance moment ( $\mu'_2$ ), built by extracting root 9-th from the original one. With this method, more useful facial features, such as eye, nose, and mouth, can be enhanced as salient regions. Thus, the weighted histogram can enrich the essential areas by exploiting the contribution of every statistical pattern in the variance image. In the next section, we show how  $\mu'_2 = \sqrt[9]{\mu_2}$  under the  $\mathcal{B} = \{(1, 6)\}$  for structuring element makes a huge difference through a series of experiments with six public face datasets.

## 5 Experiments

This section describes experiments with six face datasets, such as ORL, YALE, AR, Caltech, FERET, and KDEF. Our statistical feature descriptors were processed with the algorithm mentioned above. Below features, that were the concatenation of 4 directions in exploiting  $\text{CLBP}_{\alpha_i}$  operators, were used in our experiments:

- $\text{CLBP\_S}(m_1) = \text{CLBP\_S}_0\text{-CLBP\_S}_{45}\text{-CLBP\_S}_{90}\text{-CLBP\_S}_{135}$
- $\text{CLBP\_M}(m_1) = \text{CLBP\_M}_0\text{-CLBP\_M}_{45}\text{-CLBP\_M}_{90}\text{-CLBP\_M}_{135}$
- $\text{CLBP\_S}(m_1, \mu'_2) = \text{CLBP\_S}(m_1)\text{-CLBP\_S}(\mu'_2)$
- $\text{CLBP\_M}(m_1, \mu'_2) = \text{CLBP\_M}(m_1)\text{-CLBP\_M}(\mu'_2)$
- $\text{CLBP\_S\_M}(m_1) = \text{CLBP\_S}(m_1)\text{-CLBP\_M}(m_1)$

- $\text{CLBP\_S\_M}(m_1, \mu'_2) = \text{CLBP\_S\_M}(m_1)\text{-CLBP\_S\_M}(\mu'_2)$
- $\text{WSBP\_S}$ ,  $\text{WSBP\_M}$ , and  $\text{WSBP}$  were weighted-statistical CLBPs applied to S, M, and a fusion of S and M components as described in Section 3, respectively. Note that each descriptor had the concatenated histogram by 4 directions of  $\text{CLBP}_{\alpha_i}$ ,  $\{\alpha_i\} = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ .

The fusion of different directions and components (S, M,  $m_1$ ,  $\mu'_2$ ) would lead to a very long descriptor as the concatenation of histograms. To handle this problem, the Principal Component Analysis (PCA) with the percentage of cumulative sum of eigenvalues of 95%, was adopted for the dimension reduction purpose. For the classification task, the Linear SVMs were utilized.

### 5.1 Databases and experimental protocols

**The ORL dataset**<sup>1</sup> had 40 subjects and 10 different grayscale images with a size of  $92 \times 112$  were collected from each subject. All ORL images were collected under various conditions such as facial expression, illumination changes, occlusion (sun glasses), see Fig. 7.

**The YALE Face dataset**<sup>2</sup> included 165 images from 15 individuals and 11 different images with the size of  $243 \times 320$  were collected from each subject. The dataset had various expressions and lighting conditions, see Fig. 8.

**The Caltech 1999 dataset**<sup>3</sup> produced by California Institute of Technology had 447 images from 26 persons, yet the number of images for each person was different and collected under the unconstrained background. The dataset had various conditions such as different expression, illuminations and occlusions, see Fig. 9.

**The KDEF dataset**<sup>4</sup> [3] is a set of 4900 photographs of facial expressions. The set had 70 persons (35 males and 35 females) displaying seven facial expressions under five

<sup>1</sup><https://cam-orl.co.uk/facedatabase.html>

<sup>2</sup><http://vision.ucsd.edu/content/yale-face-database>

<sup>3</sup><http://www.vision.caltech.edu/html-files/archive.html>

<sup>4</sup><https://www.kdef.se/index.html>



**Fig. 7** The ORL dataset samples which are decomposed into pairs of moment images (mean  $m_1$ , variance  $\mu'_2$ ) for each two-row pair



different viewing angles. For the present evaluation, the frontal view was considered for each facial expression. This subset contained 490 color images for 70 individuals as shown in Fig. 10 where each subject expressed seven different emotions.

**The AR dataset**<sup>5</sup> [31] had 3016 face images for 116 persons (63 men and 53 women), each having 26 color images ( $768 \times 576$ ) under severe illumination conditions (left-light, right-light or all sidelights), 7 basic emotions (happy, sad, neutral, sleepy, anger, surprised and wink), head poses, and occlusion (sun glasses and bangs). Figure 11 shows several examples in which original color images were converted to gray-scale images and decomposed into mean and variance moment images. For the dataset, we conducted two experiments for comprehensive evaluation. Because of color images, the first experiment was examined with gray-scale images following the same protocol as other datasets, while the other was carried on other color channels to give several our perspectives.

**The FERET dataset**<sup>6</sup> [41], collected in 15 sessions during four years, was a large benchmark used extensively for comparison. The dataset comprised a total of 14,126 images from 1199 individuals. A subset adopted for evaluations had 1400 images of 200 subjects (7 images per person), including variations in poses, expression, and illumination. Figure 12 shows images under 7 states of each person.

## 5.2 Results with the ORL and YALE datasets

For the ORL dataset,  $N_{train}$  training images per subject were randomly selected ( $N_{train} = 2, 4, 5, 8$ ) while the remaining ( $10 - N_{train}$ ) images were used for testing. For

the YALE dataset,  $N_{train}$  training images per subject were randomly selected ( $N_{train} = 2, 4, 6, 8$ ) and the remaining ( $11 - N_{train}$ ) images were carried out for testing. The measurements were repeated 100 times by shuffling data process. Results were shown in Tables 1 and 2 for the average classification rate.

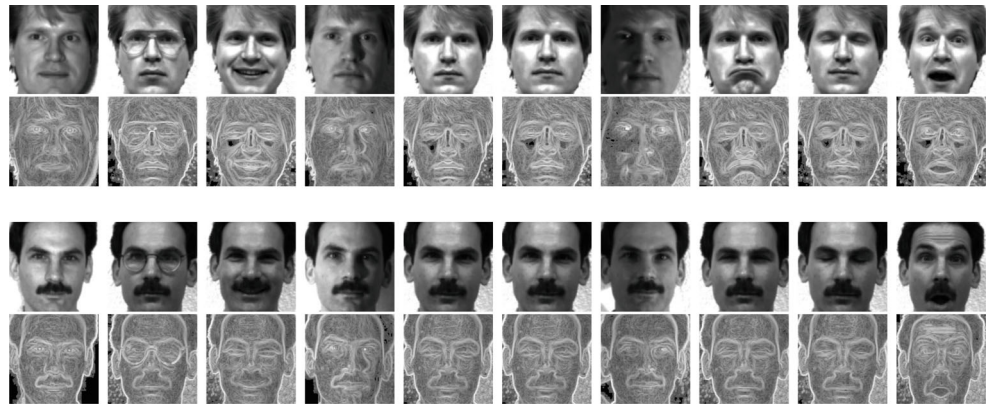
Tables 1 and 2 summarised our experimental results under the various configurations. First, the CLBP\_S( $m_1$ ) and CLBP\_M( $m_1$ ), the WSBP\_S and WSBP\_M produced similar recognition results. The CLBP\_S( $m_1$ ) and WSBP\_S played a major role in achieving good results, suggesting that the S component encoded more valuable information from each face image. Second, when utilizing both mean ( $m_1$ ) and variance ( $\mu'_2$ ) moments as the input data for CLBP, the CLBP\_S( $m_1, \mu'_2$ ) and CLBP\_M( $m_1, \mu'_2$ ) dramatically increased the classification rate than considering only the first-order ( $m_1$ ) moment, suggesting that fusion of S and M components improved the performance. Indeed, the CLBP\_S\_M( $m_1$ ) produced better results. For instance, when  $(P, R) = (4, 2)$ , the best results obtained by CLBP\_S\_M( $m_1$ ) with the number of training images ( $N = 2, 4, 5$ , and 8) for the ORL dataset were 86.22%, 96.2%, 98.25%, and 99.45%, respectively. Similarly, the WSBP descriptors reached 87.91%, 96.77%, 98.51%, or 99.41%, respectively. On the other hand, WSBP and CLBP\_S\_M( $m_1, \mu'_2$ ) were similar. For the YALE dataset (Table 2), our WSBP outperformed the other methods: 91.95%, 97.14%, 98.72%, and 99.44% at the scale of  $(P, R) = (4, 2)$ .

Also, our proposed method was compared with other state-of-the-art methods as shown in Table 3, that summarized techniques and recognition rate corresponding to each method. Ten methods on the top, including ours, were based on several hand-crafted features, and three remaining ones were based on deep features. Our method achieved 98.51% and 98.72% recognition rates for ORL and YALE datasets, which were greater than other methods, suggesting that our descriptor was robust against visual challenges, such

<sup>5</sup><https://www2.ece.ohio-state.edu/~aleix/ARdatabase.html>

<sup>6</sup>[http://www.itl.nist.gov/iad/humanid/feret/feret\\_master.html](http://www.itl.nist.gov/iad/humanid/feret/feret_master.html)

**Fig. 8** The YALE dataset samples which are decomposed into pairs of moment images (mean  $m_1$ , variance  $\mu'_2$ ) for each two-row pair



as illumination variation, facial expressions, head poses (multi-orientation), and occlusion.

### 5.3 Results with Caltech 1999 and KDEF datasets

Since the number of images for each class in Caltech 1999 dataset varied, we did not change the number of training images for each class like the previous experiments. Here, we randomly chose half of the images in each class as a training set, and the remaining ones were used as a testing set. Table 4 showed our results for three configurations of (P, R), respectively, suggesting that our WSBP and CLBP\_S\_M( $m_1, \mu'_2$ ) achieved the highest recognition rate. Table 5 compared our results with a deep learning approach based on Deep Stack Denoising Sparse Autoencoders (DSDSA) [13]. As can be seen from these tables, even using one single small scale (P, R) = (4, 2), our descriptor WSBP (Ours 1), CLBP\_S\_M( $m_1, \mu'_2$ ) reached the recognition rate of 98.83%, 98.96%, respectively, which were greater than the performance of DSDSA. When we considered CLBP\_S\_M( $m_1, \mu'_2$ ) at the scale of (P, R) = (6, 3), the performance was 99.03% (Ours 2).

For the KDEF dataset, we conducted the face recognition task by changing the number of training images for each person to verify the accuracy rates of each train/test

portion. Indeed, several training images  $N_{train}$  ( $N_{train} = 2, 3, 4, 5$ ) were randomly chosen while the remaining images ( $7 - N_{train}$ ) were used for test. Here, the evaluation was repeated 100 times by shuffling data to get the average accuracy. Table 6 showed our results for three configurations of (P, R), respectively. Specifically, at the scale of (P, R) = (4, 2), our descriptor WSBP substantially increased the accuracy up to 94.11%, 97.87%, 99.07%, and 99.33% with the number of training images  $N_{train} = 2, 3, 4$ , and 5, respectively. Such high performance suggests that our descriptor could effectively deal with visual challenges such as diverse facial expressions, illumination or occlusions.

### 5.4 Results with the AR dataset

#### 5.4.1 Evaluation with gray-scale images

We carried out the cross-validation for training and testing. The different number of training images ( $N_{train} = 10, 13, 15, 20$ ) and the remaining ( $26 - N_{train}$ ) were used for training and testing sets to guarantee they had unseen images. Results using 100 shuffle splits were summarized in Table 7 for the recognition rate and Table 8 for comparing with other methods, respectively.

**Fig. 9** The Caltech 1999 dataset samples which are decomposed into pairs of moment images (mean  $m_1$ , variance  $\mu'_2$ ) for each two-row pair



**Fig. 10** The KDEF dataset samples which are decomposed into pairs of moment images (mean  $m_1$ , variance  $\mu'_2$ ) for each two-row pair



Table 7 showed several LBPs results obtained with various parameters. As it can be seen, the CLBP\_S( $m_1$ ) and CLBP\_M( $m_1$ ) obtained the base results. With a specific parameter of (P, R) = (6, 2), the best results obtained by CLBP\_S( $m_1$ ) with  $N_{train} = 10, 13, 15, 20$  were respectively 82.87%, 88.83%, 90.79%, 95.90%. Similarly, the results of CLBP\_M( $m_1$ ) were respectively 80.09%, 86.22%, 89.83%, 95.67%. However, the recognition rates were improved significantly when bilaterally complementing S and M components with CLBP\_S\_M( $m_1$ ), the results were 98.46%, 98.68%, 99.04%, 99.93%, respectively. In this case, the CLBP\_S\_M( $m_1$ ) had a significant improvement which could increase 12.46% at  $N_{train} = 13$  (compared to CLBP\_M( $m_1$ ) and CLBP\_S( $m_1$ )). Moreover, CLBP\_S\_M( $m_1, \mu'_2$ ) and WSBP descriptors also significantly increased the performance when reaching 98.79% and 99.37%. With this parameter, our proposed framework WSBP outperformed the CLBP\_S\_M( $m_1$ ) (0.69%) and CLBP\_S\_M( $m_1, \mu'_2$ ) (0.58%). Notice that the improve-

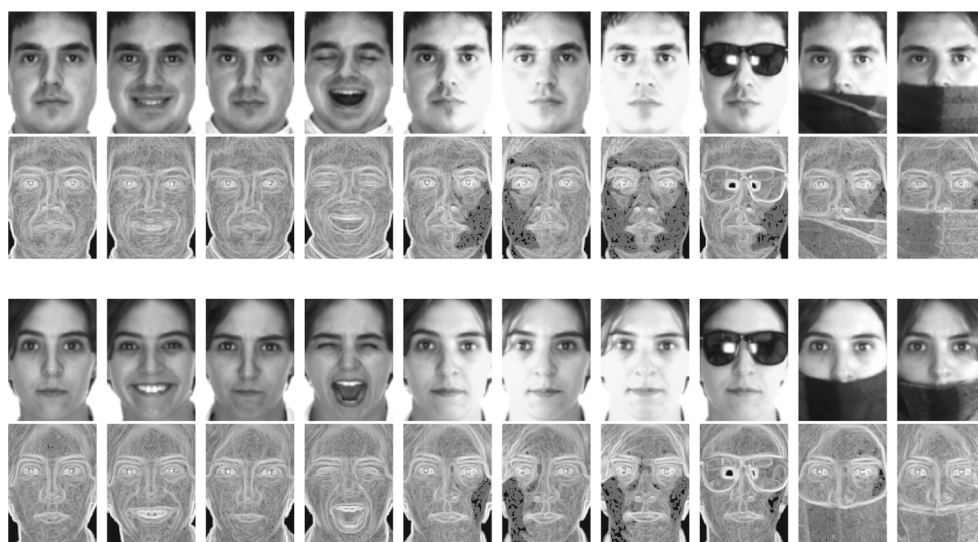
ment of WSBP could reach 13% compared to the original LBPs.

Table 8 compared our method with others. In terms of the recognition rates, Ours outperformed the state-of-the-art methods, including hand-crafted features and deep features techniques. Also, our WSBP was better than Multi-resolution dictionary [30] (82.19%), MNTCDP [21] (96.18%), Local Multiple Patterns [49] (98.00%), or even deep facial features CS [2] (93.99%) by a substantial margin. The remaining algorithms, including EL-LBP [44] (98.27%) and deep feature FDDL + CNN [39] (98%) were comparable with our descriptors, and yet ours prevailed.

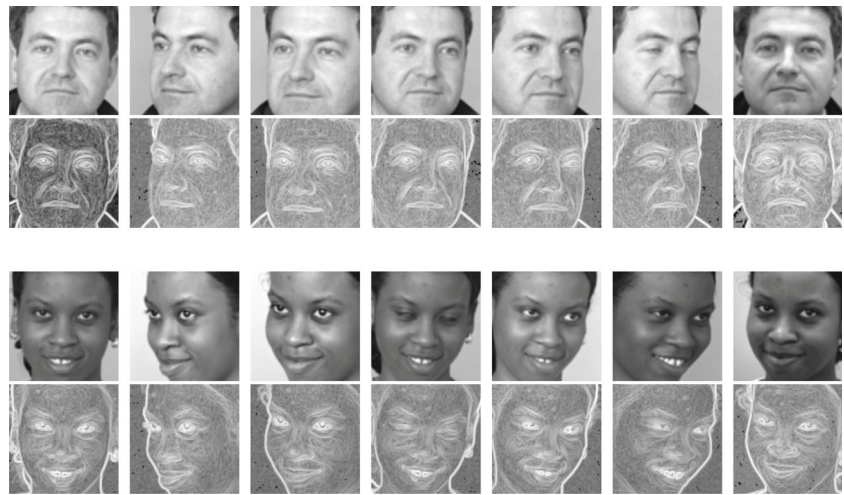
#### 5.4.2 Evaluation with color channel

The motivation of this experiment was to check the behaviors of facial descriptors for the color channel. The experiment was conducted with the HSV color images by keeping the other experimental setting was similar to

**Fig. 11** The AR dataset samples which are decomposed into pairs of moment images (mean  $m_1$ , variance  $\mu'_2$ ) for each two-row pair



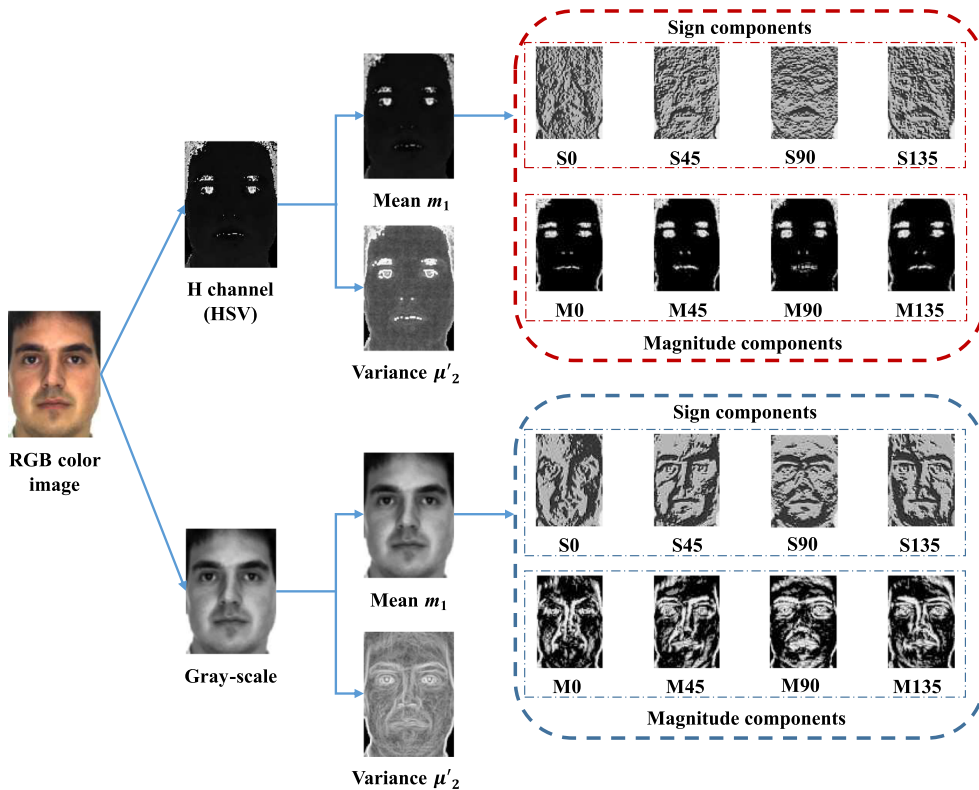
**Fig. 12** The FERET dataset samples which are decomposed into pairs of moment images (mean  $m_1$ , variance  $\mu'_2$ ) for each two-row pair



that of the gray-scale image. First, an RGB color image was converted into an HSV color image. Second, the Hue channel was extracted from the HSV space, called it H image, and fed it as an input for our experiment. Our descriptors were able to extract the eyes, eye-brows and mouth from H image probably because these areas had the distinctive colors. And yet the Sign (S) and Magnitude (M) components, computed by  $CLBP_{\alpha}$  on  $m_1$ , could not

discriminate the subtle color change occurring within the facial skin area, as shown in Fig. 13.

An evaluation with color channel was conducted with the same protocol settings of the gray-scale case. Result with 100 shuffle splits was summarized in Table 9. Note that  $(P, R) = (4, 2)$ , and  $N_{train} = 13$  were specific parameters chosen from Table 9, suggesting that S worked better than M for three cases. For instance, the accu-



**Fig. 13** Illustration of resulting images for H (Hue) and gray-scale images. The upper part contains the mean ( $m_1$ ) and new variance ( $\mu'_2 = \sqrt[3]{\mu_2}$ ) images of H image by using the structuring element  $\mathcal{B} = \{(1, 5); (2, 8)\}$  and its Sign-Magnitude components calculated by

$CLBP_{\alpha_i}$  operators. The lower part is similar to the upper one but using the gray-scale image with the structuring element  $\mathcal{B} = \{(1, 8)\}$ . All resulting images are put together for a comprehensive viewing

**Table 1** Recognition rates for the ORL dataset

$N_{train}$	(P, R) = (4, 2)				(P, R) = (6, 2)				(P, 2) = (6, 3)			
	2	4	5	8	2	4	5	8	2	4	5	8
CLBP_S( $m_1$ )	84.44	94.92	96.97	98.56	84.18	94.82	96.93	98.63	82.34	93.63	96.13	98.62
CLBP_S( $m_1, \mu'_2$ )	85.20	95.35	97.32	98.83	85.02	95.09	97.19	98.75	83.22	94.31	96.75	98.88
WSBP_S	86.32	95.70	97.72	98.98	84.33	94.85	96.94	98.66	83.79	94.63	96.92	98.94
CLBP_M( $m_1$ )	75.62	82.34	84.52	86.30	76.05	82.6	84.78	86.68	74.19	80.98	83.21	86.86
CLBP_M( $m_1, \mu'_2$ )	82.24	93.99	96.70	99.21	82.41	94.23	96.91	99.30	80.85	93.39	96.15	98.33
WSBP_M	75.95	82.54	84.89	87.56	76.56	83.12	85.38	87.86	74.93	82.01	84.17	88.46
CLBP_S_M( $m_1$ )	86.22	96.20	98.25	99.45	86.34	96.25	98.26	99.53	84.45	95.23	97.63	99.29
CLBP_S_M( $m_1, \mu'_2$ )	86.09	96.28	98.33	99.59	86.07	96.25	98.36	99.59	84.52	95.44	97.77	99.35
WSBP	87.91	96.77	98.51	99.41	85.43	96.13	98.02	99.44	84.99	95.78	97.74	99.21

racy for CLBP\_S( $m_1$ ), CLBP\_S( $m_1, \mu'_2$ ), and WSBP\_S was 49.41%, 93.56%, and 90.28%, respectively, while that of CLBP\_M( $m_1$ ), CLBP\_M( $m_1, \mu'_2$ ), and WSBP\_M was low as 24.36%, 55.32%, and 57.06%, respectively, indicating that the combination of S and M somehow impaired the overall accuracy compared to the S case. In addition, the accuracy for CLBP\_S\_M( $m_1$ ), CLBP\_S\_M( $m_1, \mu'_2$ ), and WSBP reached 33.73%, 82.52%, and 83.05%, respectively.

CLBP $_{\alpha}$ , which was inspired from CLBP [14], was designed for the gray-scale case to complement the crucial component M. It was not very effective in discriminating the color change within facial skin (see Magnitude components of H and the gray-scale images in Fig. 13). On the other hand, S component worked very well on H image by utilizing statistical moments ( $m_1, \mu'_2$ ), since its accuracy was comparable with the state-of-the-art methods. For instance, the accuracy of CLBP\_S( $m_1, \mu'_2$ ) and WSBP\_S reached 93.56% and 90.28%, respectively, while CLBP\_S( $m_1$ ) reached 49.41%. Notice that the

accuracies of CLBP\_S( $m_1, \mu'_2$ ) and WSBP\_S were better than that of CLBP\_S( $m_1$ ) since margins were 44.15% and 40.87%, respectively. These results suggest that our descriptor was designed to extract the spatial relationship of the neighboring pixels, not to simply discriminate the magnitude between pixels.

### 5.5 Results with the FERET dataset

Previous works [10, 29] in the literature performed their experiments with a protocol by using only frontal sets (Fa, Fb, Fc, Duplicate I, and Duplicate II) where Fa with 1196 images known as the gallery and others known as probes. Unlike the previous protocol, we did with a subset created from 1400 images (ba, bd, be, bf, bg, bj, bk) in which each person had two facial expression images, two left pose images, two right pose images, and one illumination image. This subset was more challenging than the previous one since it comprised not only frontal faces but also multiple orientations, expressions. Moreover, experimental results

**Table 2** Recognition rate for the YALE dataset

$N_{train}$	(P, R) = (4,2)				(P, R) = (6,2)				(P, R) = (8,4)			
	2	4	6	8	2	4	6	8	2	4	6	8
CLBP_S( $m_1$ )	89.77	93.65	94.8	95.67	89.70	93.84	94.81	95.40	91.04	94.53	94.96	95.76
CLBP_S( $m_1, \mu'_2$ )	89.53	94.34	95.24	96.18	89.54	94.40	95.47	96.04	90.97	94.69	95.57	96.27
WSBP_S	87.23	91.93	93.29	93.91	87.50	91.90	93.31	93.91	89.54	93.50	94.53	95.04
CLBP_M( $m_1$ )	80.87	86.64	87.85	88.42	80.30	86.19	87.60	88.04	81.00	87.41	88.93	89.47
CLBP_M( $m_1, \mu'_2$ )	83.17	89.80	92.21	94.09	83.13	89.77	92.33	94.44	83.96	89.94	91.79	93.42
WSBP_M	81.7	88.53	90.07	91.33	80.68	88.11	89.71	90.49	82.65	88.53	90.03	90.71
CLBP_S_M( $m_1$ )	90.36	96.55	98.4	99.27	86.05	95.90	98.12	98.89	73.39	94.5	96.48	98.51
CLBP_S_M( $m_1, \mu'_2$ )	88.61	95.91	97.69	98.62	81.47	95.10	97.39	98.29	70.82	94.08	97.24	98.20
WSBP	91.95	97.14	98.72	99.44	87.13	96.44	98.19	98.91	72.52	94.87	97.37	98.04

**Table 3** Performance comparison with the ORL and YALE datasets

Method	Techniques	ORL	YALE
Ours	WSBP	98.51	98.72
Truong et al. [44]	EL-LBP	97.12	95.27
Liang et al. [25]	Wavelet + PC	–	98.60
Karczmarek et al. [20]	CCBLBP(7px,3×3,400)	98.45	–
Kas et al. [22]	AECLBP-S (B16)	97.02	84.00
Moussa et al. [34]	DCT + PCA + GA	92.62	95.50
Kas et al. [21]	LQPAT	97.8	90.8
Luo X. et al. [30]	Multi-resolution dictionary	92.15	–
Khanbebin et al. [35]	DR-LBP + LDA	95.95	94.09
Pillai A. et al. [42]	LDENP + Chi-square distance	94.60	–
Biswas et al. [2]	Compressive Sensing (CS)	91.50	–
Chen et al. [8]	Gabor + DBN (GDBN)	94.98	98
Görgel et al. [13]	Deep Autoencoders (DSDSA)	97.50	98.16

under this subset could reflect changing accuracy rates of each train/test portion.

The experiment was carried out with random  $N_{train}$  training images of each class ( $N_{train} = 1, 2, 3, 4, 5, 6$ ), and  $N_{rest}$  testing images ( $N_{rest} = 7 - N_{train}$ ) by 100 splits for average accuracy. Table 10 illustrates the achieved results on  $CLBP_{\alpha_i}$  operators at various scales of (P, R). For most cases, WSBP obtained the best results and reached over 90% accuracy with 2 training images only. Table 11 compared a few recent methods and stated that our descriptors achieved the best performance. In detail, WSBP with 3 training images exceeded MNTCDP [21] at 2.57%, which was not easy to deal with the challenging FERET dataset having images under multiple orientations. As mentioned above, FERET had two different protocols for evaluations. It would not be a fair evaluation if we compared such methods under different protocols. And yet, it is interesting to evaluate the previous reports. For this purpose, we performed the average accuracy result based on recent reports: CLBP [10], SLBFLE [29], and WPCBP+FLD (HI) [47] (see Table 11),

**Table 4** Recognition rates for the Caltech dataset

	(P, R) = (4, 2)	(P, R) = (6, 2)	(P, R) = (6, 3)
CLBP_S( $m_1$ )	90.99	90.65	91.48
CLBP_S( $m_1, \mu'_2$ )	97.62	97.47	97.81
CLBP_M( $m_1$ )	91.74	91.84	91.84
CLBP_M( $m_1, \mu'_2$ )	97.21	97.25	97.33
CLBP_S_M( $m_1$ )	96.20	95.6	96.01
CLBP_S_M( $m_1, \mu'_2$ )	98.96	98.91	99.03
WSBP	98.83	97.89	98.32

wherein these methods performed efficiently with the subset of the frontal face cases.

## 5.6 Robustness against degraded images

In practical surveillance scenarios, the degradation of images often happened during the acquisition process and could significantly affect the system performance. Therefore, motivation of this experiment was to examine how our facial descriptors dealt with such problems. In the first scenario, the Gaussian noise was added to the original image. For instance, five different levels of Gaussian noise were added by  $levels = \{10\%, 20\%, 30\%, 40\%, 50\%\}$  using the Matlab function “imnoise”. In the second scenario, occlusion was simulated by adding a white rectangle of random positions within the face region. Each rectangle had various sizes, ranging from [20, 20] to [30, 60] with a Matlab function “insertShape”. Figure 14 showed the both scenarios.

In each scenario, five images were chosen in each class as training samples, whereas the rest as testing samples, by splitting the data into 100. The average recognition rates from the different methods were shown in Table 12. Here, we fine-tuned the structuring element  $B_2 = \{(1, 5); (2, 6)\}$  to obtain the best achievements for WSBP

**Table 5** Performance comparison with the Caltech dataset

Method	Result
Ours 1 (WSBP)	98.83
Ours 2 (CLBP_S_M( $m_1, \mu'_2$ ))	99.03
Görgel et al. [13]	97.50

**Table 6** Recognition rate for the KDEF dataset

$N_{train}$	(P, R) = (4, 2)				(P, R) = (6, 2)				(P, R) = (6, 3)			
	2	3	4	5	2	3	4	5	2	3	4	5
CLBP_S( $m_1$ )	90.69	93.18	95.18	95.90	90.68	93.23	95.07	96.11	91.19	93.63	95.40	96.31
CLBP_S( $m_1, \mu'_2$ )	94.85	97.23	98.47	98.91	94.84	97.18	98.46	98.79	94.98	97.33	98.60	98.99
WSBP_S	89.66	92.11	94.20	95.26	89.58	92.32	94.24	95.26	90.51	92.79	94.93	95.68
CLBP_M( $m_1$ )	78.14	81.68	85.35	87.47	78.45	81.77	85.32	87.41	78.88	82.76	85.78	87.99
CLBP_M( $m_1, \mu'_2$ )	84.51	91.16	94.75	95.82	84.33	91.25	94.73	95.75	84.88	91.19	94.84	96.15
WSBP_M	78.79	82.85	86.80	88.41	79.43	83.30	87.00	88.51	79.92	83.84	87.17	89.11
CLBP_S_M( $m_1$ )	88.07	91.25	93.32	93.91	88.29	91.25	93.41	94.10	88.43	91.53	93.41	94.11
CLBP_S_M( $m_1, \mu'_2$ )	91.30	95.48	97.13	97.81	91.15	95.37	97.04	97.91	91.26	95.46	97.16	97.87
WSBP	94.11	97.87	99.07	99.33	90.12	96.51	98.34	98.79	90.76	97.11	98.84	99.36

and CLBP\_S\_M( $m_1, \mu'_2$ ). This structuring element made our descriptors more robust against noise and occlusion comparing to other methods.

### 5.7 The processing time

This section describes the computational cost of several descriptors based on LBPs. Experiments of the ORL dataset for 400 images with  $92 \times 112$  pixels were carried out with a machine with 3.5GHz CPU, 32GB RAM, and Windows 10 64-bit operating system. Table 13 showed the computational cost from two aspects: firstly, the processing time for the feature descriptor extraction phase and; secondly, the processing time for the matching phase (in seconds) of various descriptors with three different configurations of (P, R). The processing time measured here was based on the

structuring element  $\mathcal{B} = \{(1, 6)\}$ , where the training set had 200 images and the testing sets had 200 images.

Table 13 showed that the WSBP required a longer processing time for the feature extraction and matching phase than CLBP\_S( $m_1$ ) or WSBP\_S. Indeed, it took much more processing time proportionally to a size of (P, R) due to larger dimension. And yet, notice that our WSBP descriptor was effective when it was compared with CLBP\_S\_M( $m_1, \mu'_2$ ) since both recognition rates were approximately the same; see Tables 13 and 1.

## 6 Summary and discussion

Based on our experiments, we summarize and discuss several advantages of our proposed descriptors:

**Table 7** Recognition rate for the AR dataset

$N_{train}$	(P, R) = (4, 2)				(P, R) = (6, 2)				(P, R) = (6, 3)			
	10	13	15	20	10	13	15	20	10	13	15	20
CLBP_S( $m_1$ )	83.06	89.20	91.05	96.1	82.87	88.83	90.79	95.9	84.33	89.92	91.77	96.58
CLBP_S( $m_1, \mu'_2$ )	98.20	98.25	98.52	99.92	98.16	98.21	98.5	99.92	98.27	98.31	98.55	99.93
WSBP_S	81.87	87.81	89.63	95.48	81.64	87.65	89.41	95.20	83.39	88.62	90.64	96.03
CLBP_M( $m_1$ )	81.49	87.35	90.56	95.88	80.09	86.22	89.83	95.67	80.99	87.41	90.80	96.43
CLBP_M( $m_1, \mu'_2$ )	94.67	96.78	97.83	99.77	94.49	96.65	97.63	99.75	94.71	96.66	97.75	99.75
WSBP_M	83.24	89.48	92.08	97.27	82.13	88.33	91.51	96.88	82.79	88.93	91.85	97.62
CLBP_S_M( $m_1$ )	98.57	98.70	99.09	99.93	98.46	98.68	99.04	99.93	98.49	98.64	99.01	99.93
CLBP_S_M( $m_1, \mu'_2$ )	98.47	98.81	99.19	99.93	98.39	98.79	99.16	99.93	98.41	98.7	99.15	99.93
WSBP	98.71	99.42	99.55	99.89	98.62	99.37	99.53	99.86	98.92	99.52	99.66	99.89

**Table 8** Performance comparison with the AR dataset

Method	Techniques	Result
Ours	WSBP	99.52
Truong et al. [44]	EL-LBP	98.27
Kas M. et al. [21]	MNTCDP	96.18
Luo X. et al. [30]	Multi-resolution dictionary	82.19
Yang W. et al. [49]	Local Multiple Patterns	98.00
Ouanan et al. [39]	FDDL + CNN	98.00
Biswas et al. [2]	Compressive Sensing (CS)	93.99

- The WSBP descriptor is designed to extend the LBPs with local difference Sign-Magnitude distributions on statistical moments. As a pre-processing step, statistical moment images obtained by spatial support  $\mathcal{B}$  of local filters can eliminate noise coming from contrast change or illumination variation (mean moment) and yet derive useful information from the salient regions in a face image (variance moment) (see Fig. 6).
- The classical LBPs consider neighborhoods bilaterally in a circle, whereas our WSBP descriptors are to exploit  $CLBP_{\alpha_i}$  operators along with multiple directions, i.e. four directions, independently and combine them in the final descriptors. It is found that they are robust against different lighting conditions, head poses, and facial expressions to achieve high performance (see  $CLBP\_S\_M(m_1)$ ,  $CLBP\_S\_M(m_1, \mu'_2)$ , and WSBP in Tables 1, 2, 4, 7, and 10).
- Since the WSBP is built by fusing CLBPs along four different directions  $\{\alpha_i\} = \{0^0, 45^0, 90^0, 135^0\}$ , it works well with a single scale (P, R) of CLBP operators. It is then unnecessary to exploit a multi-scale approach

**Table 10** Recognition rate for the FERET dataset

(P, R)	$N_{train}$	$CLBP\_S\_M(m_1)$	$CLBP\_S\_M(m_1, \mu'_2)$	WSBP
(4,2)	1	66.27	69.81	75.16
	2	79.25	86.10	91.20
	3	83.22	90.89	94.78
	4	86.43	94.15	96.86
	5	87.81	95.35	97.53
	6	87.99	94.30	97.90
(6,3)	1	65.43	69.12	73.45
		78.34	85.81	90.68
		82.49	91.00	94.48
		85.77	94.33	96.61
		86.93	95.58	97.22
		87.67	94.60	97.27
	2	66.04	70.10	69.42
		78.93	86.44	90.05
		83.12	91.30	94.66
		86.41	94.48	96.93
		87.59	95.53	97.57
		88.11	94.45	97.84

using many parameters (P, R) because it could lead to a high dimensional descriptor.

- Evaluation using six face datasets suggests that our descriptors outperform state-of-the-art methods, such as EL-LBP [44], AECLBP-S (B16) [22], Multi-resolution dictionary [30], DR-LBP + LDA [35], LDENP [42]. Moreover, our WSBP descriptors achieves better results than some deep facial features such as Deep Belief Net (GDBN) [8], Deep Autoencoders (DSDSA) [13], Compressive sensing (CS) [2], or FDDL + CNN [39] (see Tables 3, 5, and 8).

**Table 9** Performance comparison with S, M and S-M components where the structuring element was  $\mathcal{B} = \{(1, 5); (2, 8)\}$

$N_{train}$	(P, R) = (4, 2)				(P, R) = (6, 2)				(P, R) = (6, 3)			
	10	13	15	20	10	13	15	20	10	13	15	20
$CLBP\_S(m_1)$	38.83	49.41	55.98	70.00	37.59	47.68	54.02	68.52	40.48	51.51	58.25	71.72
$CLBP\_S(m_1, \mu'_2)$	90.22	93.56	94.98	98.43	89.64	93.16	94.58	98.27	91.19	94.18	95.48	98.80
WSBP_S	84.35	90.28	92.24	95.48	80.20	87.22	89.88	93.87	83.88	90.10	92.30	95.58
$CLBP\_M(m_1)$	19.99	24.36	26.99	32.48	20.03	24.39	27.16	33.04	19.70	24.33	27.16	32.85
$CLBP\_M(m_1, \mu'_2)$	46.69	55.32	57.77	66.98	47.40	56.57	58.92	68.08	46.93	55.72	58.49	67.62
WSBP_M	48.88	57.06	61.03	66.14	45.18	53.06	57.00	62.12	46.17	54.22	58.04	63.04
$CLBP\_S\_M(m_1)$	31.33	33.73	35.17	40.26	30.79	33.20	34.54	39.36	29.87	32.73	34.57	40.22
$CLBP\_S\_M(m_1, \mu'_2)$	69.93	82.52	84.22	90.43	69.99	82.60	84.40	90.70	69.93	82.28	84.50	90.82
WSBP	74.11	83.05	86.46	91.00	67.16	76.34	80.57	85.97	71.55	80.52	84.38	89.38



**Table 11** Performance comparison with the FERET dataset

Descriptor	1	2	3	4	5	6
Average recognition rate for different $N_{train}$						
WSBP	75.16	91.20	94.78	96.86	97.53	97.90
CLBP_S- $M(m_1, \mu'_2)$	70.10	86.44	91.30	94.48	95.53	94.45
MNTCDP [21]	61.57	83.56	92.21	96.48	98.52	99.70
AECLBP-S-MxC [21]	51.42	73.08	83.59	90.44	93.99	97.19
LQPAT [4, 21]	52.48	74.14	83.94	91.96	94.97	97.24
LMP [49]	49.47	67.79	76.22	85.28	90.78	97.65
FLRDP [16]	–	–	–	74.58	82.85	90.70
ALTP [48]	63.75	77.30	72.75	87.33	92.25	91.00
Average recognition rate with the previous protocol						
CBP [10]	92.00					
SLBFLE (R=4) [29]	96.95					
WPCBP+FLD (HI) [47]	97.5					

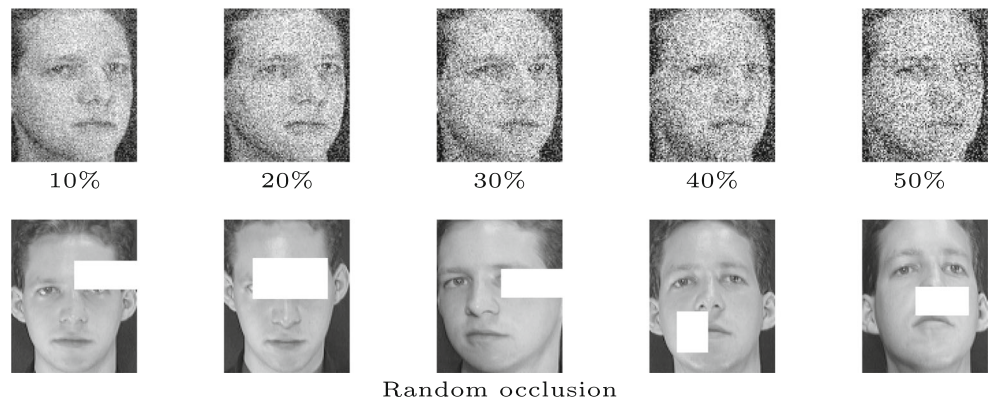
- According to an additional experiment with the color channel, it is found that the Magnitude transform captures the relationship of pixel magnitude on gray-scale image very well, but is not effective with Hue image (see Fig. 13), since a combination of Sign-Magnitude of  $CLBP_\alpha$  in Hue space performs worse than that of the gray-scale case. Also, fusing statistical moments ( $m_1$  and  $\mu_2$ ) in  $CLBP_S$  and  $WSBP_S$  achieves the higher accuracy in Hue space, by ignoring texture pixel intensity. This evaluation suggests a new direction in face recognition problems, such as integrating many color channels to enhance face spoofing detection performance [43].
- Although the YALE dataset contains some facial expression cases, it would be interesting in testing how our descriptor affords systematic variation of facial expressions. In addition, we use the KDEF dataset

- which has seven facial expressions for each subject to study the effect of facial expressions. Result suggests that our descriptor deals with such cases very well.
- Our facial descriptors using both mean ( $m_1$ ) and variance ( $\mu_2$ ) have shown their robustness against degraded images by evaluating the ORL dataset that contains artificial noise. Given that the Gaussian noise level of 50% makes the degraded face more challenging to recognize by human eyes, our  $WSBP(\mathcal{B}_2)$  descriptor still reaches the acceptable accuracy of 93.05% for noise and 85.09% for occlusion, which are much higher than those of other LBPs.

### 7 Conclusions and future work

We present a set of descriptors wherein the local difference distributions in local binary patterns are exploited by

**Fig. 14** The modified ORL images: (1) the first row has five levels of Gaussian noise (10%-50%), (2) the second one has occlusion by white random rectangles



**Table 12** Performance comparison for different descriptors with the ORL dataset added with Gaussian noise and occlusion

Descriptor	Gaussian noise					Occlusion
	10%	20%	30%	40%	50%	random
WSBP	96.69	95.18	91.88	91.21	88.04	82.59
WSBP ( $\mathcal{B}_2$ )	97.69	96.80	96.10	94.32	93.05	85.09
CLBP_S_M( $m_1, \mu'_2$ ) ( $\mathcal{B}_2$ )	97.14	96.84	95.79	94.39	93.00	87.32
CCBLD(1px,4×4,100) [20]	83.88	61.50	36.43	19.75	13.95	36.35
CCBLD(7px,3×3,400) [20]	97.43	95.40	93.58	90.90	87.48	63.30
LBP	83.68	61.20	33.58	16.08	6.88	32.23
FR(D,5×5,300) [7, 20]	33.63	19.58	12.68	5.33	4.38	23.53
LMP [49]	96.85	95.41	94.16	92.30	88.09	39.55

directions, and then a weighting approach for binary patterns is applied to statistical moment images for an efficient and robust facial feature representation. A comprehensive evaluation with several standard face datasets is carried out to validate our proposal. We have analyzed the behaviors of several descriptors with gray-scale images and found that our method mostly outperforms state-of-the-arts. Also, an analysis with a set of color images has also been examined using the Hue channel for AR dataset. We have also simulated a few practical scenarios, that can be occurred during the data acquisition stage, by adding various Gaussian noise and random occlusion to the ORL dataset. One may understand that the spatial support strategy is a special preprocessing technique to eliminate the noise issues, and selecting the structure element  $\mathcal{B}$  depends on the levels and types of noise. For the scenarios examined in this study, it is found that the structuring element of two circles eliminates noise very efficiently. Although this issue could downgrade the recognition performance, our experimental result is still higher than

others. It shows that our proposed descriptor is robust against the degradation of the given image. Overall, our experimental results suggest that the proposed descriptor is robust against noise, contrast change, illumination variation, and facial expressions by exploiting different directions of binary pattern operators on the mean moment and considering the contribution of binary pattern to the variance moment.

We expect that these descriptors find more applications in the face recognition area and other areas such as facial paralysis analysis and face spoofing detection. Although our proposed framework is novel and high-performing, it has a few issues to be addressed: (1) the computational cost for matching increases when the descriptor dimension becomes larger; (2) it is necessary to fine-tune the optimal  $k$ -parameter for the root extraction variance moment. We plan to focus on how to deal with them. Also, it would be interesting to combine the WSBP descriptors with deep neural network for building powerful descriptors.

**Table 13** The processing time of the descriptors used for the present study with different parameters (FT: feature extraction time, FTS: feature size of LBPs descriptors without using any dimension reduction techniques, and MT: matching time)

Descriptor	(P, R) = (4, 2)			(P, R) = (6, 2)			(P, R) = (8, 4)		
	FT	FTS	MT	FT	FTS	MT	FT	FTS	MT
CLBP_S( $m_1$ )	4.38	2940	0.59	5.27	6468	2.25	6.47	11564	3.71
CLBP_S( $m_1, \mu'_2$ )	7.22	5880	1.4	9.23	12936	4.14	11.56	23128	8.77
WSBP_S	3.61	2940	0.81	4.63	6468	2.16	5.73	11564	3.83
CLBP_S_M( $m_1$ )	6.20	5880	1.64	7.91	12936	3.93	9.97	23128	7.07
CLBP_S_M( $m_1, \mu'_2$ )	11.17	11760	2.94	14.42	25872	8.02	18.74	46256	15.15
<b>WSBP</b>	4.39	5880	1.62	5.88	12936	3.92	7.78	23128	6.79

**Acknowledgements** This work was supported by the Institute of Information & communications Technology Planning & Evaluation (IITP), grant funded by the Korea government (MSIT) (No.2019-0-00231) as well as by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1A6A1A03038540).

We would like to express our deep gratitude to the reviewers and editors, who pointed out the valuable and insightful remarks allowing us to clarify the presentation of this work.

## Declarations

**Conflict of Interests** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ahonen T, Hadid A, Pietikainen M (2006) Face description with local binary patterns: application to face recognition. *IEEE Trans Pattern Anal Mach Intell* 28(12):2037–2041
- Biswas S, Sil J, Maity S (2018) On prediction error compressive sensing image reconstruction for face recognition. *Comput Electr Eng* 70:722–735
- Calvo MG, Lundqvist D (2008) Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behav Res Methods* 40(1):109–115
- Chakraborty S, Singh SK, Chakraborty P (2017) Local quadruple pattern: A novel descriptor for facial image recognition and retrieval. *Comput Electric Engineer* 62:92–104
- Chakraborty S, Singh SK, Chakraborty P (2018) Centre symmetric quadruple pattern: A novel descriptor for facial image recognition and retrieval. *Pattern Recogn Lett* 115:50–58
- Chakraborty S, Singh SK, Chakraborty P (2018) Local gradient hexa pattern: a descriptor for face recognition and retrieval. *IEEE Trans Circ Syst Vid Technol* 28(1):171–180
- Chan CH, Yan F, Kittler J, Mikolajczyk K (2015) Full ranking as local descriptor for visual recognition: A comparison of distance metrics on sn. *Pattern Recogn* 48(4):1328–1336
- Chen Y, Huang T, Liu H, Zhan D (2016) Multi-pose face ensemble classification aided by Gabor features and deep belief nets. *Optik* 127(2):946–954
- Chihaoui M, Elkefi A, Bellil W, Ben Amar C (2016) A survey of 2D face recognition techniques. *Computers* 5
- Deng W, Hu J, Guo J (2019) Compressive binary patterns: designing a robust binary face descriptor with random-field Eigenfilters. *IEEE Trans Pattern Anal Mach Intell* 41(3):758–767
- Ding C, Choi J, Tao D, Davis LS (2016) Multi-directional multi-level dual-cross patterns for robust face recognition. *IEEE Trans Pattern Anal Mach Intell* 38(3):518–531
- Ding C, Tao D (2015) Robust face recognition via multimodal deep face representation. *IEEE Trans Multimed* 17(11):2049–2058
- Görgel P, Simsek A (2019) Face recognition via Deep Stacked Denoising Sparse Autoencoders (DSDSA). *Appl Math Comput* 355:325–342
- Guo Z, Zhang L, Zhang D (2010) A completed modeling of local binary pattern operator for texture classification. *IEEE Trans Image Process* 19(6):1657–1663
- Hernandez-Matamoros A, Bonarini A, Escamilla-Hernandez E, Nakano-Miyatake M, Perez-Meana H (2016) Facial expression recognition with automatic segmentation of face regions using a fuzzy based classification approach. *Knowl-Based Syst* 110:1–14
- Huang P, Gao G, Qian C, Yang G, Yang Z (2017) Fuzzy linear regression discriminant projection for face recognition. *IEEE Access* 5:4340–4349
- Işık S, Özkan K (2015) A comparative evaluation of well-known feature detectors and descriptors. *Int J Appl Math Electron Comput* 3:1–6
- Jridi M, Napoléon T, Alfalou A (2018) One lens optical correlation: application to face recognition. *Appl Opt* 57(9):2087–2095
- Karanwal S, Diwakar M (2020) Two novel color local descriptors for face recognition. *Optik* :166007
- Karczmarek P, Kiersztyn A, Pedrycz W, Dolecki M (2017) An application of chain code-based local descriptor and its extension to face recognition. *Pattern Recogn* 65:26–34
- Kas M, El merabet Y, Ruichek Y, Messoussi R (2018) Mixed neighborhood topology cross decoded patterns for image-based face recognition. *Expert Syst Appl* 114:119–142
- Kas M, El-merabet Y, Ruichek Y, Messoussi R (2020) A comprehensive comparative study of handcrafted methods for face recognition LBP-like and non LBP operators. *Multimed Tools Appl* 79(1):375–413
- Kortli Y, Jridi M, Al Falou A, Atri M (2020) Face recognition systems: a survey. *Sensors* 20(2)
- Le T, Vo MT, Kieu T, Hwang E, Rho S, Baik SW (2020) Multiple electric energy consumption forecasting using a cluster-based strategy for transfer learning in smart building. *Sensors* 20(9)
- Liang H, Gao J, Qiang N (2020) A novel framework based on wavelet transform and principal component for face recognition under varying illumination. *Appl Intell*. <https://doi.org/10.1007/s10489-020-01924-9>
- Liao S, Jain AK, Li SZ (2016) A fast and accurate unconstrained face detector. *IEEE Trans Pattern Anal Mach Intell* 38(2):211–223
- Lin J, Chiu CT (2017) Low-complexity face recognition using contour-based binary descriptor. *IET Image Process* 11(12):1179–1187
- Liouane Z, Lemlouma T, Roose P, Weis F, Messaoud H (2018) An improved extreme learning machine model for the prediction of human scenarios in smart homes. *Appl Intell* 48:2017–2030. [10.1007/s10489-017-1062-5](https://doi.org/10.1007/s10489-017-1062-5)
- Lu J, Liong VE, Zhou J (2018) Simultaneous local binary feature learning and encoding for homogeneous and heterogeneous face recognition. *IEEE Trans Pattern Anal Mach Intell* 40(8):1979–1993
- Luo X, Xu Y, Yang J (2019) Multi-resolution dictionary learning for face recognition. *Pattern Recogn* 93:283–292
- Martinez A, Benavente R (1998) The ar face database. CVC Technical Report 24
- Mehta R, Egiiazarian K (2016) Dominant Rotated Local Binary Patterns (DRLBP) for texture classification. *Pattern Recogn Lett* 71:16–22
- Mi J, Liu T (2016) Multi-step linear representation-based classification for face recognition. *IET Comput Vis* 10(8):836–841
- Moussa M, HMILA M, Douik A (2018) A novel face recognition approach based on genetic algorithm optimization. *Stud Inf Control* 27(1):127–134

35. Najafi Khanbebin S, Mehrdad V (2020) Local improvement approach and linear discriminant analysis-based local binary pattern for face recognition. *Neural Comput Appl.* <https://doi.org/10.1007/s00521-020-05512-3>
36. Napoléon T, Alfalou A (2017) Pose invariant face recognition: 3D model from single photo. *Opt Lasers Eng* 89:150–161
37. Nguyen TP, Vu NS, Manzanera A (2016) Statistical binary patterns for rotational invariant texture classification. *Neurocomputing* 173:1565–1577
38. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal Mach Intell* 24(7):971–987
39. Ouanan H, Ouanan M, Aksasse B (2018) Non-linear dictionary representation of deep features for face recognition from a single sample per person. *Procedia Comput Sci* 127:114–122
40. Pham NT, Lee JW, Park CS (2020) Structural correlation based method for image forgery classification and localization. *Appl Sci* 10(13)
41. Phillips PJ, Moon H, Rizvi SA, Rauss PJ (2000) The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans Pattern Anal Mach Intell* 22(10):1090–1104
42. Pillai A, Soundrapandiyan R, Satapathy S, Satapathy SC, Jung KH, Krishnan R (2018) Local diagonal extrema number pattern: A new feature descriptor for face recognition. *Futur Gener Comput Syst* 81:297–306
43. Shu X, Tang H, Huang S (2020) Face spoofing detection based on chromatic ED-LBP texture feature. *Multimed Syst*
44. Truong HP, Kim YG (2018) Enhanced line local binary patterns (EL-LBP): An efficient image representation for face recognition *Proc of 2018 ACIVS*, pp. 285–296
45. Truong HP, Vo TMD, Le T (2016) Face recognition based on LDA in manifold subspace EAI. *Endorsed Trans Context-aware Syst Appl* 3(9)
46. Vo AH, Hoang Son L, Vo MT, Le T (2019) A novel framework for trash classification using deep transfer learning, vol 7, pp 178631–178639
47. Xu Z, Jiang Y, Wang Y, Zhou Y, Li W, Liao Q (2019) Local polynomial contrast binary patterns for face recognition. *Neurocomputing* 355:1–12
48. Yang W, Wang Z, Zhang B (2016) Face recognition using adaptive local ternary patterns method. *Neurocomputing* 213:183–190
49. Yang W, Zhang X, Li J (2020) A local multiple patterns feature descriptor for face recognition. *Neurocomputing* 373:109–122

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Hung Phuoc Truong** received the B.Sc. degree in information technology and the M.Sc. degree in computer science from the University of Science, Vietnam National University, Ho Chi Minh City, Vietnam, in 2009 and 2012, respectively. He is now a Ph.D. candidate in Computer Science at the Computer Engineering Department, Sejong University, Korea. His research interests include biometrics identification, computer vision, and machine learning.



**Thanh Phuong Nguyen** received his PhD in computer science at Lorraine University, France, in 2010. He has been teaching assistant at Lorraine University from 2009 to 2011, then he has worked as a postdoctoral fellow at CMM, Mines Paristech in 2012, at Robotics and Computer Vision lab, ENSTA Paristech from 2013 to 2015. Since September 2015, he is an Assoc. Prof. at University of Toulon, France and permanent member of SIIM team, LIS laboratory. His research interests include image analysis, visual descriptor, action recognition, texture analysis, shape representation, face analysis, and discrete geometry. He has published about 20 international journals and 30 international conferences.



**Yong-Guk Kim** received the B.S. and M.S. degrees in Electrical Engineering from Korea University, Seoul, Korea, in 1982 and 1984, respectively. His work experience includes LG Electronics and Korea Telecom (KT). He received a Ph.D. focusing on computational vision in 1997 from the University of Cambridge, UK.

From 1995 to 1996, he was a Research Fellow with Helmholtz Robotics Institute, Utrecht, the Netherlands, and from 1998 to 2001, was a Research Associate with Smith-Kettlewell Vision Institute, San Francisco, USA. From 2001, he has been in the Department of Computer Engineering, Sejong University, Seoul, Korea, and is a Full Professor. He was the Director of the start-up incubator and Dean of International Affairs at Sejong University. His research areas are face recognition, facial expression recognition, driver drowsiness detection, fake emotion detection, human head pose estimation, and vision-based autonomous drone.