We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Extended Binary Gradient Pattern (eBGP): A Micro- and Macrostructure-Based Binary Gradient Pattern for Face Recognition in Video Surveillance Area

Nuzrul Fahmi Nordin, Samsul Setumin, Abduljalil Radman and Shahrel Azmin Suandi

Abstract

An excellent face recognition for a surveillance camera system requires remarkable and robust face descriptor. Binary gradient pattern (BGP) descriptor is one of the ideal descriptors for facial feature extraction. However, exploiting local features merely from smaller region or microstructure does not capture a complete facial feature. In this paper, an extended binary gradient pattern (eBGP) is proposed to capture both micro- and macrostructure information of a local region to boost up the descriptor performance and discriminative power. Two topologies, the patch-based and circular-based topologies, are incorporated with the eBGP to test its robustness against illumination, image quality, and uncontrolled capture conditions using the SCface database. Experimental results show that the fusion between micro- and macrostructure information significantly boosts up the descriptor performance. It also illustrates that the proposed eBGP descriptor outperforms the conventional BGP on both the patch-based topology and the circular-based topology. Furthermore, a fusion of information from two different image types, orientational image gradient magnitude (OIGM) and grayscale image, attained better performance than using OIGM image only. The overall results indicate that the proposed eBGP descriptor improves the recognition performance with respect to the baseline BGP descriptor.

Keywords: surveillance system, face recognition, binary gradient pattern (BGP), facial feature extraction, patch-based topology, circular-based topology

1. Introduction

Face recognition is one of the biometric verification methods that offers a wide range of applications such as law enforcement, forensics, biometric authentication, surveillance, and health monitoring [1]. Face recognition has also been used

to authenticate payment using mobile wallet, and the social media company like Facebook uses face recognition algorithm for the purpose of image tagging [2]. One of the advantages of face recognition is being contactless between the subject and camera. Given the advantages offered by face recognition and with the advancement in computing power, significant research and methods have been proposed over the years in face recognition domain. In fact, a robust facial recognition system must be able to work with various real-life situations or unconstrained conditions, such as but not limited to pose, lighting, image or camera quality, occlusion, rotation, and translation. The system must also be able to perform extremely well in a domain where limited sample is available. In surveillance monitoring applications, a typical approach is to sample face appearing in videos and then match them with facial models generated from high-quality target face image [3, 4].

Feature extraction is the process of capturing feature of interest from the face and represents it in the form of feature vector. The extraction process is usually done by a face descriptor. This descriptor must be able to work with multiple variations such as illumination, occlusion, face expression, and image quality [4]. Indeed, there is a collection of face descriptors proposed over the years such as scale-invariant feature transform (SIFT) [5], speeded up robust feature (SURF) [6], local binary pattern (LBP) [7], and histogram of oriented gradient (HOG) [8]. In terms of facial feature representation, there are two types of representations that many descriptors have evolved around over the years. They are global and local feature representations. Global-based feature extraction like principal component analysis [9], linear discriminant analysis [10], and independent component analysis [11] preserves the statistical information of the face by turning each face image into a high-dimensional feature vector. Meanwhile, local-based feature splits input image into smaller patches and extracts the micro textural details from each patch before fusing these features back to form the global shape information. Local-based feature extraction has shown to be resilient to multiple variations by enforcing spatial locality in both pixel and patch levels. For instance, local feature descriptor is robust to local deformation in expression and occlusion. LBP [7] is an example of feature extraction method that works on this principle which achieved reasonably good performance but heuristic in nature. Recently, LBP has drawn great intention as a face descriptor due its reputation as a powerful texture descriptor [9]. LBP extracts local-based spatial structure of an image by thresholding intensity of center pixel with its neighborhood. The product of this operation is characterized as local binary pattern, which then the distribution of binary pattern over the whole image is used to form the LBP histogram vector or feature vector. Neighborhood pixels are sampled on a circle, and any neighbor which does not fall exactly on the center of the pixel has an intensity computed from interpolation [7]. Due to some shortcomings of LBP, for instance, LBP produces long histogram, and therefore it is memory-consuming [12], LBP is very sensitive for image rotation and noise [13], and it only captures microstructure and ignores macrostructure of the texture resulting in missing extra discriminative power [14]. Several variants of LBP have been proposed in the literature, for example, rotation-invariant LBP [13], median robust extended local binary pattern (MRELBP) [15], and binary gradient pattern (BGP) [14]. This paper touches on a number of relevant existing LBP-based descriptors. The rest of this paper is organized as follows. In Section 2, two state-of-the-art descriptors (the LBP [7] and its variant, the BGP [14]) would be briefly reviewed since we would embed the proposed extended BGP (eBGP) into these two descriptors. Section 3 describes the proposed eBGP descriptor. The evaluating results are analyzed and discussed in Section 4. Finally, conclusions are drawn in Section 5.

2. From local binary pattern (LBP) to binary gradient pattern (BGP)

LBP [7] is one of various texture descriptors and is known for being computationally efficient [16]. It extracts local-based spatial structure of an image by thresholding intensity of center pixel with its neighborhood pixel P within a radius R. The product of this operation is characterized as local binary pattern, which then the distribution of binary pattern over the whole image is used to form the LBP histogram vector or feature vector. The original LBP works on 3×3 square neighborhood and only considers the sign information to form the LBP pattern. Neighborhood pixels are sampled on a circle, and any neighbor which does not fall exactly on the center of the pixel has an intensity computed from interpolation [7]. **Figure 1(a)** illustrates LBP neighborhoods around the center pixel with $R = 1$. Assuming all the pixels hold values as in **Figure 1(b)**, thresholding all eight neighborhood pixels with the center pixel using Eq. (1) will produce the result as in **Figure 1(c)**. This binary string is then multiplied with weights, and the sum of these values corresponds to the LBP label for that particular pixel. The distribution of LBP labels across the entire image is represented in a histogram as a feature vector:

$$LBP_{R,P}(c) = \sum_{i=0}^{P-1} s(g_i - g_c) \, 2^i, \, s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \quad (1)$$

where g_i and g_c are the gray values of the center pixel and its neighbors, respectively, P is the number of neighbors, and R is the radius of the neighborhood. LBP offers few advantages in terms of low computational complexity, illumination invariant, and ease of implementation, but it has significant disadvantages. In LBP implementation, the individual operator of particular (P,R) produces different histogram length. For instance, in (8,1) neighborhood, LBP generates $2^P = 256$ ($P = 8$) histogram bins, while for (16,2) neighborhood, 2^{16} histogram bins are produced. This is a significant drawback as LBP produces long histogram and therefore memory-consuming. The LBP is also intolerant to image rotation and highly sensitive to noise where noise on the center pixel will dominate local characteristic [12]. Furthermore, the LBP only captures microstructure and ignores macrostructure of the texture resulting in missing extra discriminative power.

The success of LBP has continued since then. A variety of LBP-based descriptors have been proposed recently to overcome all shortcomings toward noise, illumination, color, and temporal information. Huang and Yin [14] proposed an improved version of LBP, called binary gradient pattern (BGP), by introducing structural pattern and image gradient orientation (IGO) implementation in multiple directions rather than on X and Y directions only, as in the conventional manner. The implementation of IGO in multiple directions helps to improve discriminative power of the proposed descriptor. **Figure 2** shows how BGP encodes binary string
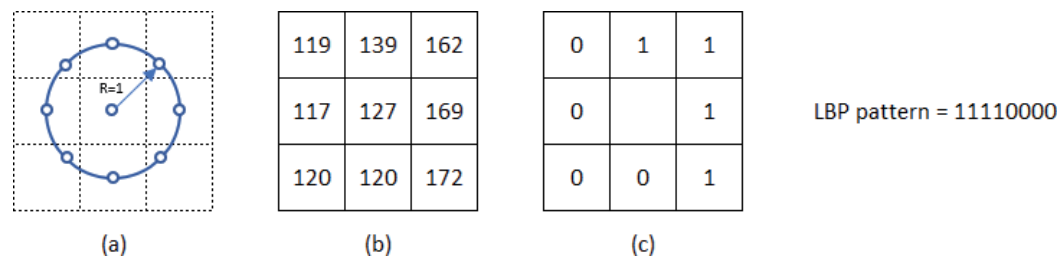

119	139	162		0	1	1
117	127	169		0		1
120	120	172		0	0	1

LBP pattern = 11110000

(a)　　　　　　(b)　　　　　　(c)

Figure 1.
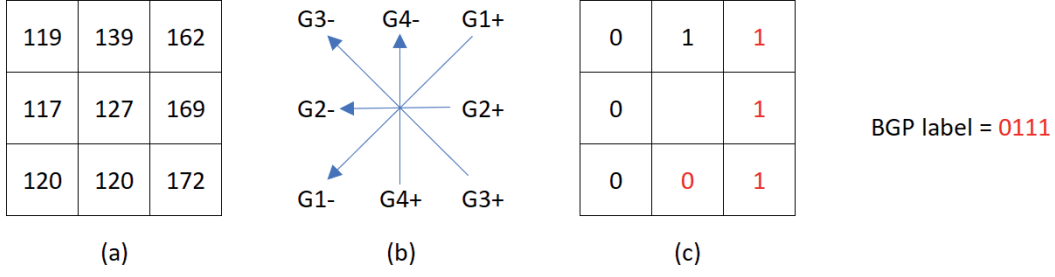*LBP neighborhood and thresholding.*

Figure 2.
Basic BGP operator with eight neighbors.

from a region of interest (ROI). Given a set of grayscale intensity value of 9 pixels as in **Figure 2(a)**, BGP computes binary correlations between symmetric neighbors of central pixel from multiple k directions. With the number of neighbors always twice than the number of directions k, in (8,1) spatial resolution, there are four different thresholding directions denoted as G1, G2, G3, and G4 as shown in **Figure 2(b)**. Principal binary, B_i^+, is computed from all directions using Eq. (2), and its associated binary B_i^- from Eq. (3), where G_i^+ and G_i^- are intensity values of the pixels. The resulting principal binary numbers and its associated are shown in **Figure 2(c)**:

$$B_i^+ = \begin{cases} 1, if\, G_i^+ - G_i^- \geq 0 \\ 0, if\, G_i^+ - G_i^- < 0 \end{cases} \tag{2}$$

$$B_i^- = 1 - B_i^+, i = 1, 2, ..., k \tag{3}$$

$$L = \sum_{i=1}^{k} 2^{i-1} B_i^+ \tag{4}$$

Binary string for the ROI is constructed from four principal binary numbers which is equivalent to 0111, and the label L is computed from Eq. (4). Because the principal and associated binary numbers are always complementary, only a single bit is required to describe the direction, this allowing for more compact representation of BGP label by only considering principal binary numbers. The total number of BGP label N_L is determined by the numbers of principal binary only, which is also equivalent to the number of directions k. At any spatial resolution, N_L equals to $2k$. Using **Figure 2(b)** as an example, features extracted from four directions in (8,1), spatial resolution will produce 2^4 or 16 different labels (i.e., from 0000 to 111/from 0 to 15). Structural pattern is a binary string which has continuous "1's indicating a stable local change in texture and essentially describes the orientation of local edge texture. On the other hand, a nonstructural pattern is a binary string with a discontinuous "1's, which contains arbitrary changes of local texture which is likely to indicate noise or outliers. From statistical experiment conducted by Huang and Yin [14] on 2600 face images, 95% of the patterns in typical BGP face having continuous "1's.

The number of structural labels N_{sp} at any spatial resolution equals to the number of neighbors P. With eight neighbors, there will be 16 different labels where eight of it made up a structural label and the remaining belong to nonstructural label. For example, 0000, 0001, 0011, 0111, 1000, 1100, 1110, and 1111 are structural patterns in $BGP_{8,1}$, and each structural pattern location map is illustrated in **Figure 3**. In BGP implementation, nonstructural patterns are discarded and not given a label in contrast to nonuniform pattern in LBP implementation. Location map of nonstructural patterns in **Figure 3** shows that nonstructural patterns contain less meaningful information and are often caused by noise and outliers.
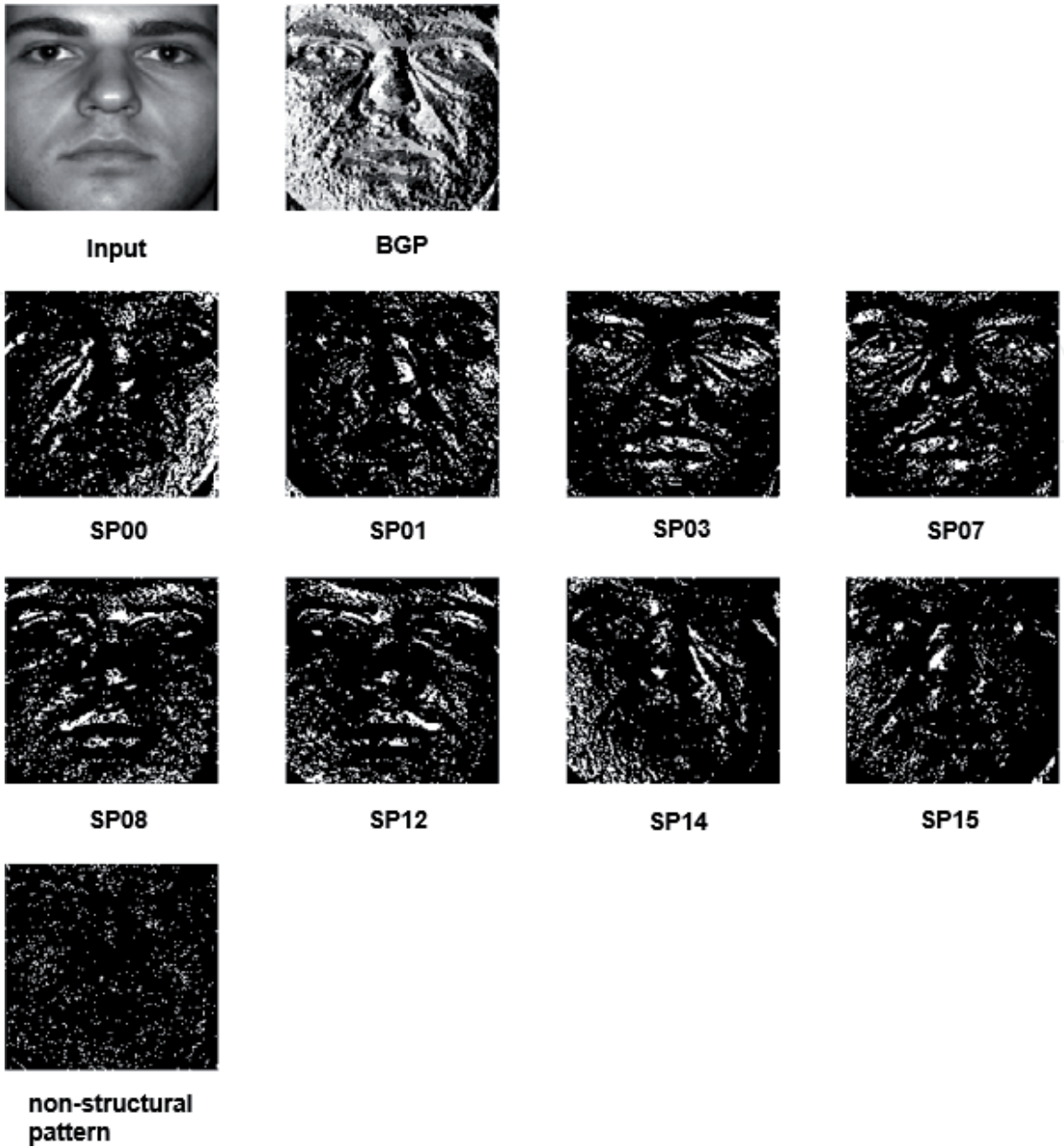
Figure 3.
Face and location maps of eight structural patterns (SP00-SP15) and nonstructural pattern.

To further enhance discriminative power and robustness of BGP, Huang and Yin [14] introduced another descriptor by applying BGP on orientational image gradient magnitude (OIGM) which is abbreviated as BGPM. The use of image gradient magnitude (IGM) enhances the strength of edge information which effectively allows BGPM to gain greater discriminant ability with only small increment in complexity. The overall process of BGPM descriptor is depicted in **Figure 4**.

Based on a series of results obtained from multiple databases such as Extended Yale B [17], AR [18], CMU Multi-PIE [19], FERET [20], and LFW [21] against a wide range of descriptors, BGPM is proven to be the best descriptor for each database. The BGPM descriptor has achieved invariance against illumination changes and local distortions while reducing the vector dimensionality. BGP compact representation makes BGP extremely fast and uses much fewer pattern labels than LBP at any spatial resolution. For instance, in a system with spatial resolution of (8,1), BGP histogram only needs 9 bins, with 8 bins for structural patterns, and 1 bin for nonstructural patterns, in contrast to the LBP which requires 59 bins. BGP and BGPM have been demonstrated to possess strong spatial locality and orientation properties which lead to effective discrimination.
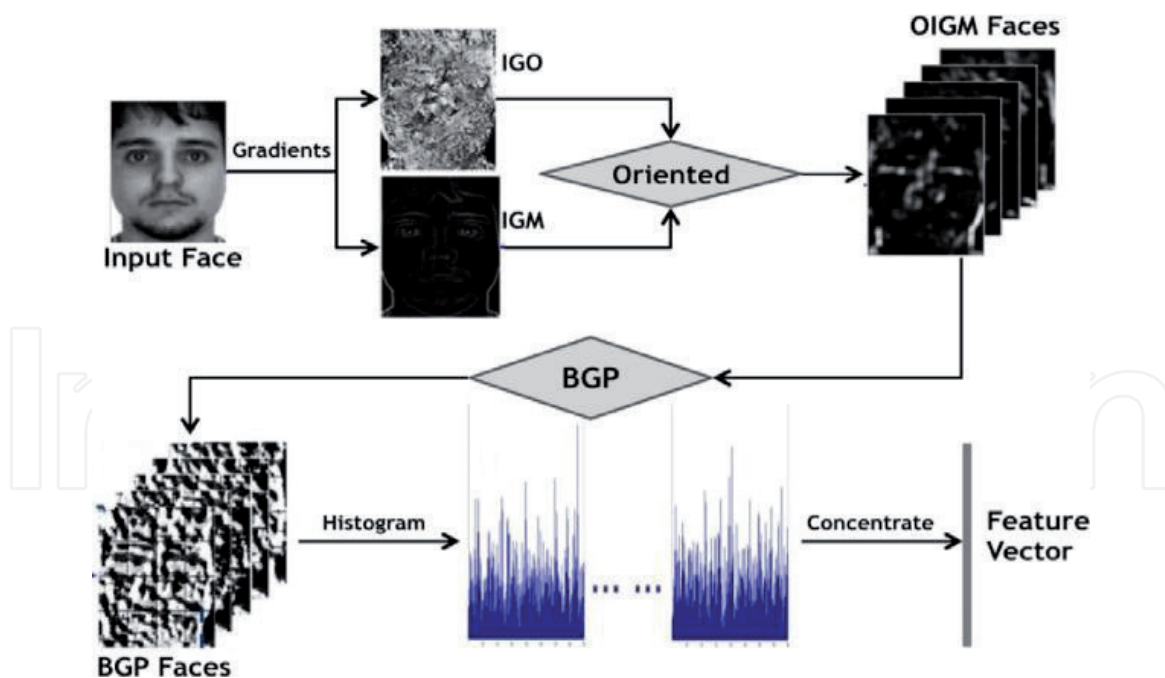
Figure 4.
Framework of BGPM descriptor [14].

Although BGP has shown to be efficient in processing time and achieving outstanding results in several databases, BGP was never being tested with a proper surveillance database like [22], which consists of low-resolution non-frontal face images taken by different camera quality. Like most of other local-based descriptors, BGP exploits information from microstructure only, however exploiting facial feature from macrostructure to complement the microstructure feature resulting in a more complete image representation [23–24], especially for surveillance applications where noise, occlusion, and head position might impact the descriptor performance. In this paper, information from both micro- and macrostructures are captured and integrated into the BGP descriptor to boost up its performance for video surveillance applications. The new proposed descriptor is termed as an extended BGP (eBGP).

3. Extended binary gradient pattern (eBGP)

An eBGP extends the BGP descriptor by exploiting macrostructure information from topology with larger spatial resolution. There are many different types of macrostructure topologies that have been proposed for other LBP variants [25]. In this paper, the patch-based topology with eight neighborhood patches and the circular topology are evolved with the proposed eBGP descriptor. Both topologies have been implemented by [24, 26], where each topology has its pros and cons with the implementation. Regardless of the topology, the microstructure information is always extracted using the same approach as in BGP. Herein, the eBGP is explained with the focus on extracting features from macrostructure based on the patch-based topology with eight neighborhood patches and the circular-based topology.

3.1 Patch-based topology

Patch-based topology is inspired by multi-scale block local binary pattern (MBLBP) [24]. In this topology, macrostructure is made up of nine patches of pixels as in **Figure 5**. All these patches have the same size and width, while the center patch represents the ROI microstructure. Thereby, a default BGP operator is applied
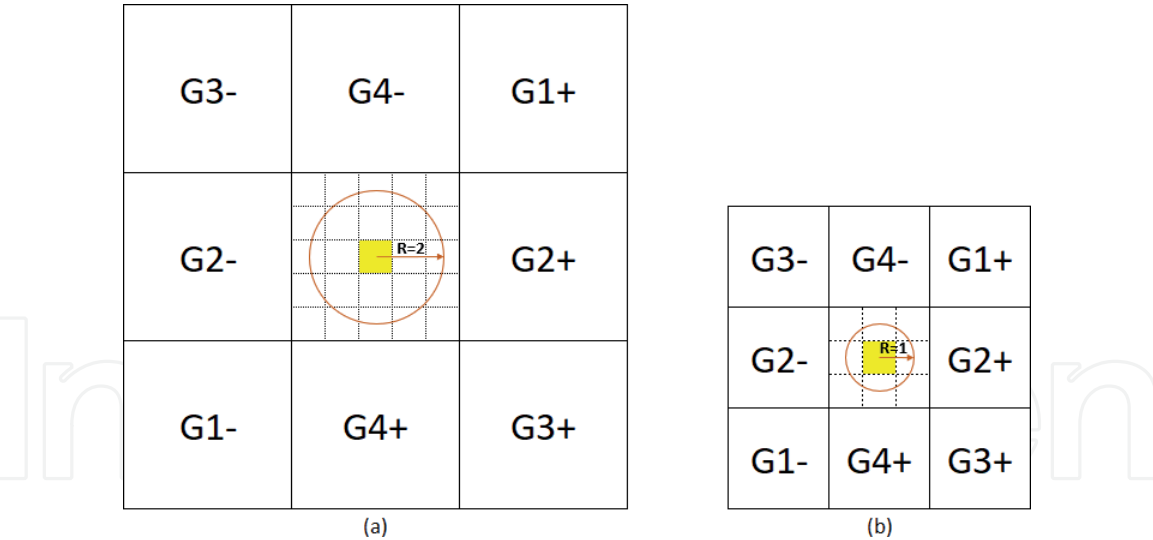
Figure 5.
Topology for macrostructure information extraction. (a) Patch of 5 × 5 pixels for R = 2. (b) Patch of 3 × 3 pixels for R = 1.

to the center patch in order to extract the microstructure information, whereas the macrostructure information is extracted from the eight neighborhood patches. Accordingly, multiple sizes of patches could be selected from this topology, and the size of the structure is determined by the spatial resolution of the center patch.

For instance, when exploiting microstructure information from (8,1) spatial resolution, the size of the center patch will be 3 × 3 pixels as illustrated in **Figure 5(b)**. In this implementation, all patches have the same size and do not overlap each other; therefore the macrostructure is formed from nine patches of 3 × 3 pixels. **Figure 5(a)** depicts the macrostructure topology formed from 9 patches of 5 × 5 pixels when microstructure information is exploited from (16,2) spatial resolution. For comparison purposes, this research will evaluate two structures as illustrated in **Figure 5(a)** and **(b)**, to match BGP results exploited from (8,1) and (16,2) spatial resolution. Using **Figure 5(a)** as an example, each neighborhood patch contains 25 pixels with each pixel having its own grayscale value. Unlike the center patch, no feature is extracted from the individual neighborhood patch. Instead, each neighborhood patch is represented by a single intensity value which will be used for thresholding. In this topology, the patch's mean and median will be applied to represent the patch intensity. The patch's mean (G) of a neighborhood patch (P), accounted from 25 pixels in a single 5 × 5 patch, is computed as follows:

$$G_P = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad (5)$$

where x is the intensity value of each pixel and n is the number of pixels in the patch P.

On the other hand, the patch median is computed by finding the middle value of ordered pixel values. Additional experiments are conducted in this research to find the best representation for the patch-based topology. As an example, feature extraction from macrostructure is illustrated in **Figure 6**. **Figure 6(a)** shows the patch-based topology with the size of 3 × 3 pixels and its intensity value. In each patch, a median is calculated from all pixels within the patch, and the median now represents the image intensity of the patch as shown in **Figure 6(b)**. The following steps are similar to what has been explained in BGP. By thresholding each patch with symmetric neighbors in four directions using Eqs. (2) and (3), four pairs of binary numbers are generated as shown in **Figure 6(c)**. Once all the principal bits are computed, the
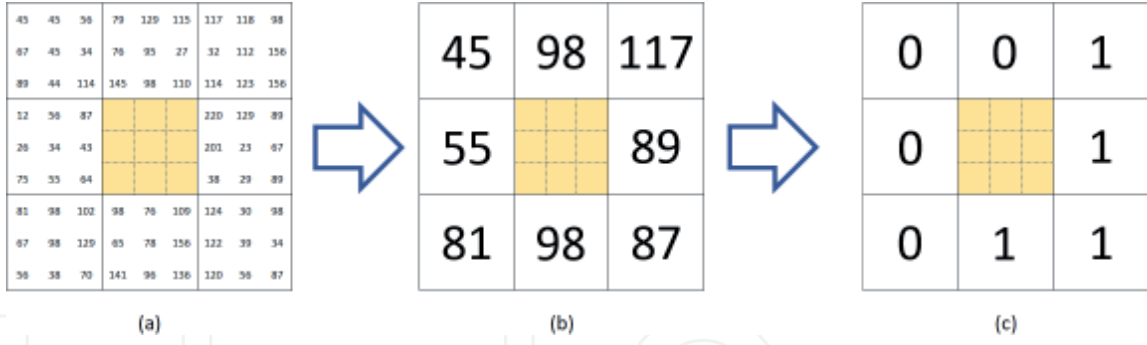
Figure 6.
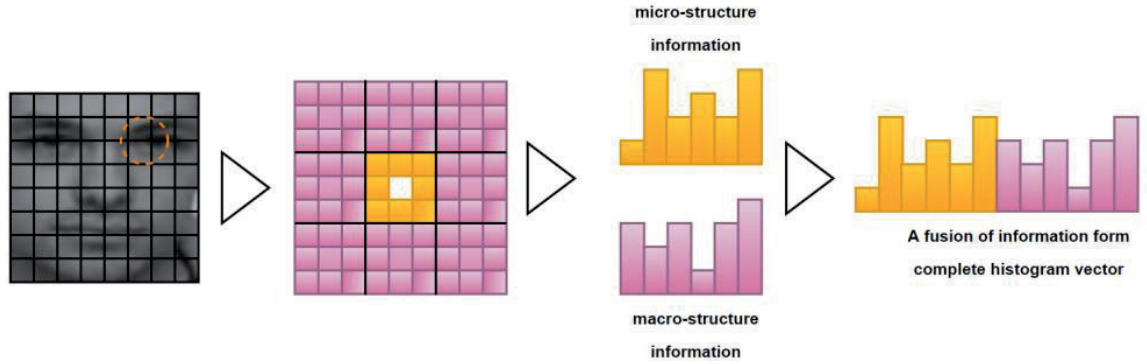*Feature extraction from the macrostructure using median as the patch intensity.*

Figure 7.
Patch-based feature extraction flow. The center patch represented by the orange box and the neighborhood patches by the purple boxes.

label can be calculated using Eq. (4). In general, the flow for macrostructure extraction is like microstructure except for its representative value used during thresholding. Indeed, the microstructure information is extracted from neighborhood pixels, while the macrostructure information is extracted from neighborhood patches.

Since there are only eight neighbor patches, regardless of the structures' size, the generated histogram vector which represents the macrostructure information is bound to the maximum of 16 bins. Observing only a structural pattern will greatly reduce the dimensionality of macrostructure information to eight bins. The total length of the histogram vector (H_t) is computed as follows:

$$H_t = \sum_{k=1}^{N} (P_R + 8)_k \qquad (6)$$

where N is the number of blocks and P_R is the number of neighborhood pixels used for extracting the microstructure information at the center patch and 8 is the length of the histogram vector extracted from the macrostructure. Using **Figure 6(b)** as an example, at each k_{th} block, the length of histogram vector is 16, where 8 comes from the microstructure and the other 8 from the macrostructure.

Subsequently, information fusion between micro- and macrostructures is conducted through concatenating the feature vectors of both the microstructure and the macrostructure, as illustrated in **Figure 7**. At this point, both feature vectors are contributed by the same weight. **Figure 8** demonstrates an example of face image represented using the patch-based topology. **Figure 8** illustrates that eBGP on the patch-based topology capable to capture the micro textural details and the macrostructure provides complementary information to the small details. Moreover, the macrostructure information contains less detailed information and may reduce the noise or outlier embedded in the image.
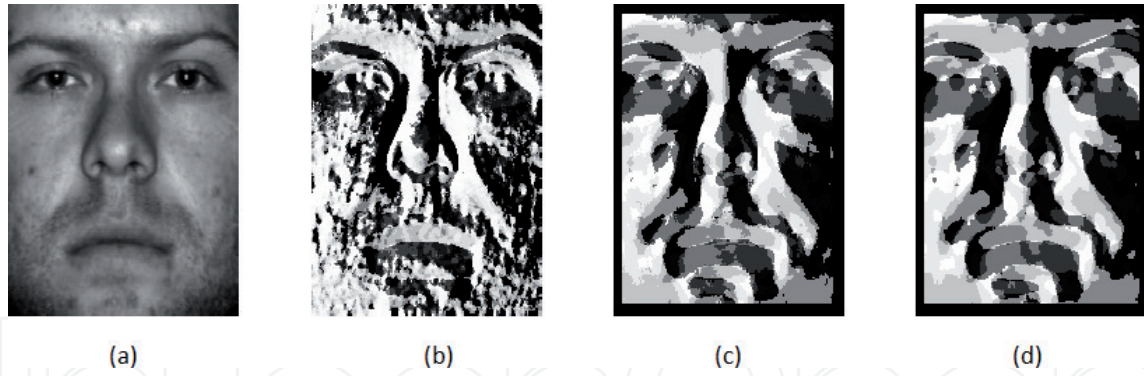
Figure 8.
Sample image with 5 × 5 pixel patch-based structure: (a) the original image, (b) the image extracted using the microstructure, (c) the image extracted using the macrostructure based on the local median, and (d) the image extracted from the macrostructure using the local mean.

3.2 Circular-based topology

Circular-based topology borrows the basic implementation of LBP which identifies a neighborhood as a set of pixels on a circular ring. In this topology, two levels of information are extracted from neighborhood at two different spatial resolutions. The first level of information is the microstructure information, which is extracted from a set of pixels on a circular ring with radius R_1. Meanwhile, the macrostructure information is extracted from neighborhood pixels that lie on a circular ring of radius R_2. The same BGP operator is used to extract information from the two different spatial resolutions with smaller spatial resolution that represents the microstructure and larger spatial resolution that represents the macrostructure. The visual illustration of circular-based topology implementation is presented in **Figure 9**. Circular rings with R_1 and R_2 represent the two different spatial resolutions $(P;R)$. Assuming R_1 is 1, running BGP descriptor on (8,1) neighborhood extracts the microstructure information of ROI. In this implementation, R_2 is always larger than R_1, and thus R_2 must be set to any number >1.

Figure 10(a) shows a sample of image intensity that falls on circular rings R_1 and R_2 with spatial resolution (8,1) and (24,3), respectively. In this example, the microstructure information is extracted from 8 pixels, while the macrostructure information is extracted from 24 pixels as shown in **Figure 10(b)**. Using the same method in BGP, principal and associated bits are calculated using Eqs. (2) and (3) by thresholding symmetric neighbors in multiple directions. The computed binary pairs are shown in **Figure 10(c)** with 4 and 12 principal bits generated from 8 and 24 neighbors, respectively. Finally, label for both micro- and macrostructures is computed using Eq. (4).

In BGP scheme, the length of histogram vector is equal to the number of neighbors at any spatial resolution. Similar to the patch-based topology, the generated histogram vector which embeds micro- and macrostructure information is concatenated to form a final representation of features for each ROI. The total length of histogram vector in this scheme can be computed using:

$$H_t = \sum_{k=1}^{N} (P_1 + P_2)_k \qquad (7)$$

where N is the number of blocks and P_1 and P_2 are the number of neighborhood pixels on the circular rings of radius R_1 and R_2, respectively. For instance, if $R_1 = 2$ and $R_2 = 4$, features are exploited from 16 and 32 neighborhood pixels, respectively. Thus, the combination of the two spatial resolutions will produce a histogram vector with a length of 48 at each k_{th} block. Resulting from this observation, R_2 is set to
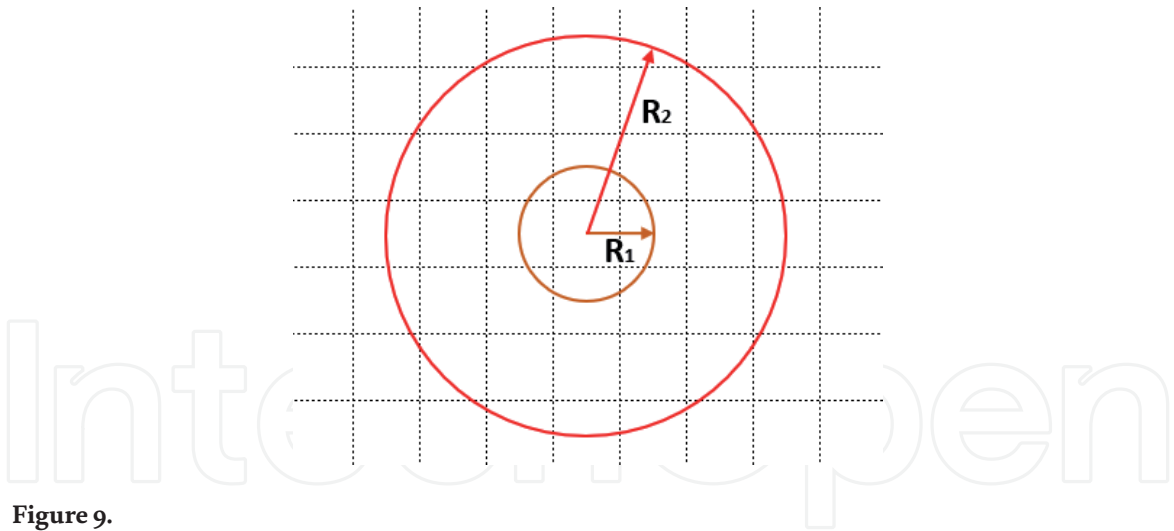
Figure 9.
Circular-based topology.
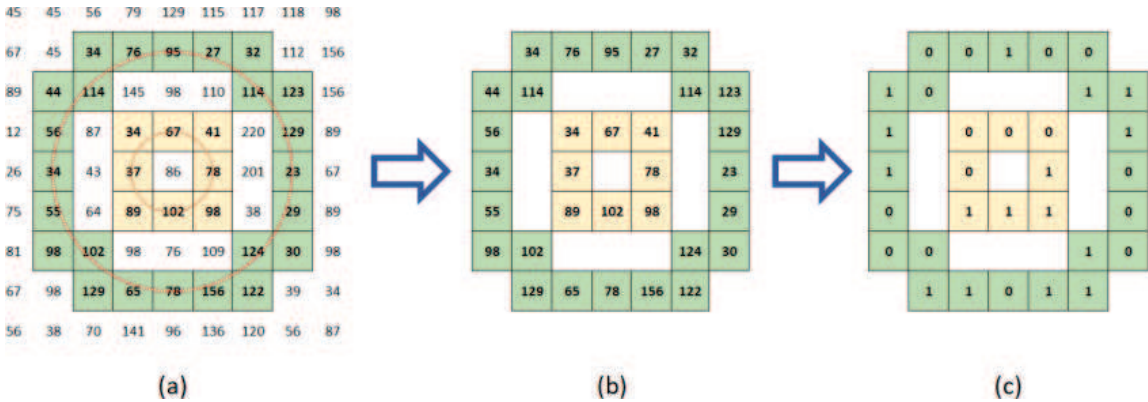
Figure 10.
The microstructure information is devised from 8 pixels on the smaller ring, while the macrostructure information is devised from 24 pixels on the larger ring without any interpolation.
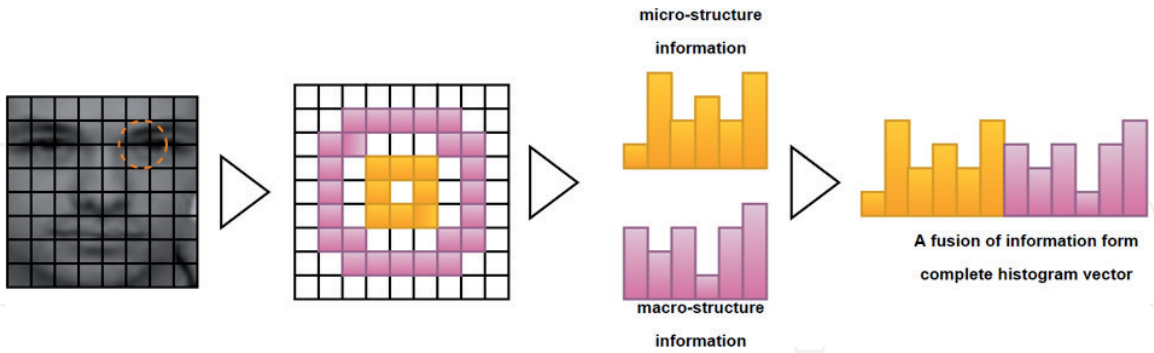
Figure 11.
Circular-based feature extraction flow with $R_1 = 1$ and $R_2 = 3$.

5 to limit the feature dimensionality of macrostructure to 40 because having larger spatial resolution will only increase the feature vector dimensionality. In contrast, R_1 is limited to 4 because larger spatial resolution will prevent BGP operator from capturing micro edge and micro texture features which are mostly exploited from a smaller region.

Figure 11 illustrates the general flow of feature extraction in the circular-based topology. Overall, this topology employs BGP operator on two different spatial resolutions, where the smaller resolution is for the microstructure information and the larger resolution is for the macrostructure information. In this research, no interpolation has been done to neighboring pixels where the circle does not fall

(a)　　　　　　　　(b)　　　　　　　　(c)

Figure 12.
Sample image with R₁ = 2 and R₂ = 5 circular-based topology: (a) the original image, (b) the image extracted from the microstructure (R₁ = 2), and (c) the image extracted from the macrostructure (R₂ = 5).

exactly on the center of pixels. **Figure 12** presents a sample image that is extracted from the two spatial resolutions $R_1 = 2$ and $R_2 = 5$.

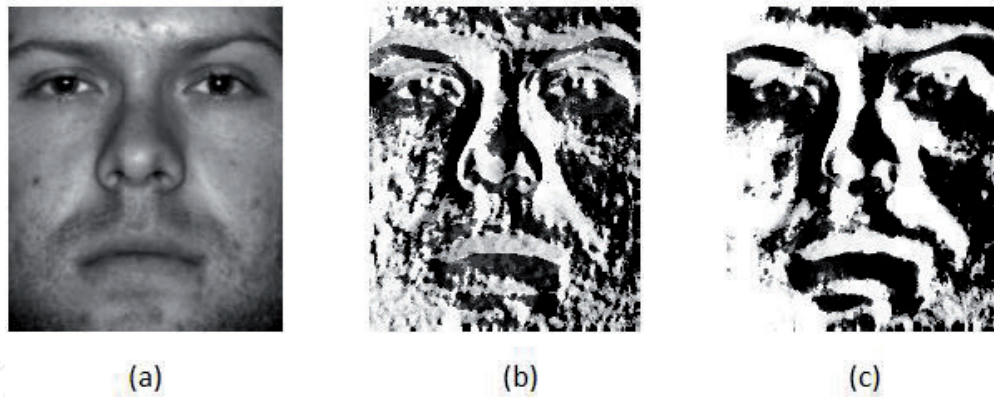Similar to the patch-based topology, BGP captures the micro-oriented edges from the small structure while capturing less details of information at a much larger spatial resolution. But the combination of these two information will complement each other in providing a complete face representation.

4. Results, discussion, and analysis

To illustrate a real-world video surveillance system, the effectiveness of the proposed eBGP descriptor was evaluated using the Surveillance Camera Face (SCface) database [22]. The SCface database consists of low-resolution non-frontal face images taken by different camera quality. A series of experiments were planned to test all proposed topologies and structures on the SCface database. The performance of the proposed eBGP descriptor was evaluated against illumination, image quality, single sample per person, and real-world capture condition.

In fact, the SCface database is the most challenging database for face recognition, where its images were taken in uncontrolled indoor environment. The SCface database consists of 4160 images from 130 subjects. All images were taken at three distinct distances from the camera, where the cameras are installed at 2.25 m above the floor. Images were captured at distance 1 while the subject position is 4.20 m away from the camera, whereas for distances 2 and 3, the subject positions were at 2.60 and 1.00 m, respectively. The outdoor light was only the source of illumination, which came through a window on one side. The images were captured from five different quality commercial surveillance video cameras and two infrared night-vision cameras, in uncontrolled lighting so as to mimic the real-world conditions. Furthermore, full frontal mug shot image for each subject was captured using a high-quality photo camera with the capture conditions exactly the same as would be expected for any law enforcement. The high-quality photo camera for capturing visible light mug shots was installed the same way as the infrared camera but in a separate room with the standard indoor lighting, and it was equipped with adequate flash. In our experiments, the high-quality mug shot image of each person was used as a training gallery, while the remaining images from the five surveillance cameras and distances were used as test images, as depicted in **Figure 13**. With the focus of this research toward images in visible spectrum and single sample per person, especially for real-world surveillance system, the images taken from IR night-vision cameras and mug shot rotation were not used in this research. As preprocessing

Figure 13.
Sample images from the SCface database of distance 3: (a) the high-quality mug shot. (b–f) The images taken from five different surveillance cameras. (g and h) The images were taken from IR night-vision cameras.

steps, all images in the SCface database were aligned based on the provided eye coordinates, so that the eyes' line lies on a straight line. The images were then scaled and cropped to 64x64 pixel as has been implemented in [22].

The performance of the proposed eBGP descriptor was evaluated using the histogram intersection, where the histogram intersection computes the similarity between two discretized probability distributions or histogram vectors. Given H^T is the histogram vector of a training image reference and H^P is the histogram vector of a probe image, each one containing n bins, the intersection between them is defined as follows:

$$H^T \cap H^P = \sum_{j=1}^{n} \min\left(H_j^T, H_j^P\right) \qquad (8)$$

where H^T and H^P are generated from distribution of labels computed from eBGP operator and the *min* function takes as arguments two values and return the smallest one. Any histogram pair that returns the highest intersection value based on Eq. (8) than any other pairs is considered to be matched and assigned to the label. By comparing this label against ground truth label, the recognition rate is determined by counting the occurrence of the correct label over the number of test images. Recognition rate is computed as follows:

$$Recognition\ rate\ (\%) = \frac{N_L}{N} \times 100\% \qquad (9)$$

where N_L is the total number of test images which are correctly matched and N is the total number of test images.

It is vital to stress that the classifier plays a decisive role in achieving better recognition rate. In this research, the experiments were dictated in such a way to

focus on recognition rate improvement due to macrostructure information fusion. Hence, the recognition rate of the proposed eBGP descriptor and its baseline BGP descriptor were computed and compared to verify the recognition rate improvement. For comparative analysis, results of BGP descriptor on the SCface database are produced by running the BGP code requested from [14]. This is to ensure analysis of the result can be done without any concern on the validity of the results. In fact, Huang and Yin [14] do not use the SCface database in their work; thus BGP code was altered to work with the SCface database.

4.1 Experiment settings and preprocessing

As a preprocessing step, each image is first transformed into OIGM images using the same method used by the BGP descriptor. OIGM images are then divided into N numbers of non-overlapped blocks before applying eBGP descriptor, where N is set to 16 in this research.

4.2 Results of patch-based topology

For better presentation, several notations are used to describe the experiment setup and implementation. $BGPM_{(P;R)}$ is the implementation used in the BGP descriptor of spatial resolution (P,R), while $eBGPM_{(P;R)}$ is the implementation of the proposed eBGP descriptor with macrostructure information based on the patch-based topology. In this experiment, the patch-based topology uses the patch's median as a default scheme for the thresholding between patches.

Table 1 shows the performance of the proposed descriptor on the SCface database, where $eBGPM_{(16;2)}$ and $eBGPM_{(8;1)}$ represent the extended BGPM (eBGPM) with structures of **Figure 5(a)** and **Figure 5(b)**, respectively. Results of $BGPM_{(16;2)}$ and $BGPM_{(8;1)}$ represent the baseline descriptor. As mentioned before in this section, the images of SCface database were captured by five cameras with three different distances. **Table 1** shows the recognition rate results for each set and the average recognition rate over all cameras. The recognition rate for each set was calculated based on Eqs. (8) and (9).

Distance	Descriptor	Camera					
		1	2	3	4	5	Average
1	$BGPM_{(8;1)}$	3.08	0.77	3.08	3.08	5.38	3.08
	$BGPM_{(16;1)}$	6.15	4.62	4.62	3.85	5.38	4.92
	$eBGPM_{(8;1)}$	4.62	1.54	4.62	3.85	6.15	4.16
	$eBGPM_{(16;1)}$	3.85	7.69	5.38	5.38	8.46	6.15
2	$BGPM_{(8;1)}$	16.15	12.31	6.92	11.54	13.85	12.15
	$BGPM_{(16;1)}$	23.85	13.85	7.69	12.31	13.08	14.16
	$eBGPM_{(8;1)}$	20.77	13.85	10.77	16.92	16.15	15.69
	$eBGPM_{(16;1)}$	23.08	17.69	13.85	16.92	16.15	17.54
3	$BGPM_{(8;1)}$	15.38	19.23	10.00	16.92	11.54	14.61
	$BGPM_{(16;1)}$	18.46	20.00	16.15	14.62	11.54	16.15
	$eBGPM_{(8;1)}$	19.23	17.69	11.54	17.69	13.08	15.85
	$eBGPM_{(16;1)}$	16.15	16.15	15.38	16.15	17.69	16.30

Table 1.
Recognition rate (%) of the proposed eBGP descriptor on the SCface dataset using the patch-based topology.

From **Table 1**, it can be seen that none of the descriptors achieved recognition rate higher than 35% over all cameras and distances. Particularly, the images of distance 1 recorded the lowest recognition rate with an average of 4.58%, while the images of distances 2 and 3 achieved better recognition rates with an average of 14.89 and 15.73%, respectively. **Table 1** also shows that eBGPM$_{(8;1)}$ slightly boosted up the performance comparable with BGPM$_{(8;1)}$ for all distances, where it attained the highest recognition rate over BGPM$_{(8;1)}$ on the distance 2 with an average recognition rate which equals to 3.54%. On the contrary, eBGPM$_{(16;2)}$ has a mix result with respect to its baseline BGPM$_{(16;2)}$; the performance drop can be observed from camera 1 gallery results, where distance 1, distance 2, and distance 3 show lower recognition rate comparable with the baseline descriptor. Similar to eBGPM$_{(8;1)}$, eBGPM$_{(16;2)}$ presented the highest recognition rate on distance 2 gallery images compared to those from distance 1 and distance 3. This is because the gallery images of distance 1, which have been acquired at 4.20 m distance, are low in resolution and small in size. Moreover, the process of scaling and cropping the images into 64 × 64 size leads to loss of the quality and some dominant features. On the other hand, the images of distance 3 are higher in quality and details. However, as the subjects are closer to the camera, which is installed at 2.25 m from the floor, in most natural head position, the upper half of the subject face is more dominant in the captured images as depicted by **Figure 14**. Figure 14 demonstrates that the images of distance 2 are slightly better in quality than the other two distances, but they still suffer from head position. This interprets the superiority of descriptors on this distance.

Due to these discouraging results by both the proposed eBGP descriptor and its baseline BGP, extra experiments were conducted on the SCface database. Since **Table 1** illustrated that the recognition rate is improved with increase of the spatial resolution, consequently the BGPM descriptor is first extended to larger spatial resolution of (24,3). Even though recognition rate increased by including the macrostructure in eBGP, the overall recognition rate is still too low for realistic applications. It might be because the structural pattern and OIGM image were
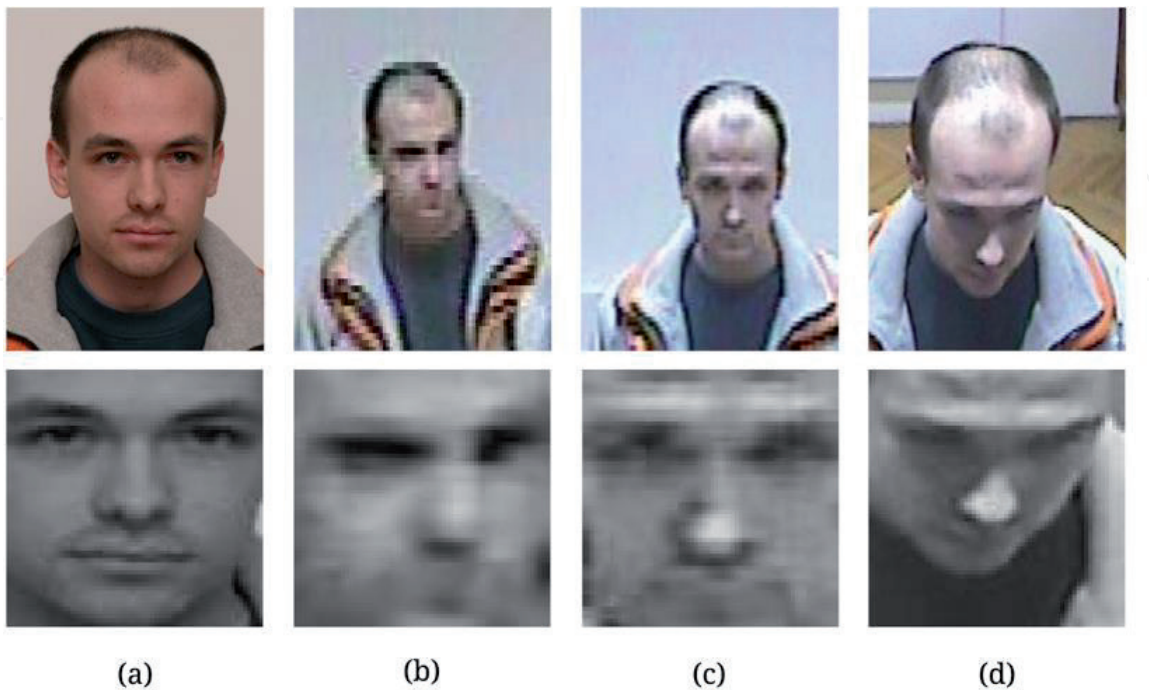

| (a) | (b) | (c) | (d) |

Figure 14.
Samples of the SCface database: (a) training image mug shot and (b–d) test images captured by camera 2 at distances 1, 2, and 3, respectively. The upper row shows the original images, while the lower row shows the images after alignment, scaling, and cropping to 64 × 64.

Distance	Descriptor	Camera					
		1	2	3	4	5	Average
1	$BGPM_{(24;3)}$	5.38	2.31	4.62	4.62	5.38	4.62
	Type I_P	3.85	6.92	4.62	6.92	3.85	6.00
	Type II_P	10.77	6.92	6.92	5.38	10.77	8.31
2	$BGPM_{(24;3)}$	21.54	16.15	13.08	16.15	15.38	16.46
	Type I_P	23.85	20.00	13.85	19.23	15.38	18.46
	Type II_P	34.62	25.38	20.00	25.38	21.54	25.38
3	$BGPM_{(24;3)}$	20.00	18.46	14.62	16.15	11.54	16.15
	Type I_P	16.92	16.92	14.62	16.92	16.15	16.31
	Type II_P	22.31	23.08	15.38	23.85	16.92	20.31

Table 2.
Recognition rate (%) of $BGPM_{(24;3)}$, Type I_P and Type II_P descriptors on the SCface database.

extracted from low-resolution and deformed images (after scaling and cropping have been done). Hence, two additional descriptors were then designed to investigate the effectiveness of structural patterns and OIGM image when exploiting the macrostructure information from low-resolution images. These descriptors still use BGPM in exploiting information from the microstructure, but they extract the macrostructure information in a different way.

The first additional descriptor, denoted as Type I_P in **Table 2**, is equivalent to the $eBGPM_{(16;2)}$ descriptor with one exception. The structural pattern concepts are ignored, and all labels which are produced by (16,2) spatial resolution are assumed to hold some unique features. In this setup, all information from 16 labels are used to populate the histogram vector. This descriptor is designed to investigate if there is any other feature that may be discarded by the structural patterns when dealing with low-quality images. The second descriptor, denoted as Type II_P in **Table 2**, is designed to extract information from both OIGM and grayscale intensity images. This descriptor extracts the microstructure information from the OIGM image and the macrostructure information from the grayscale image. Type II_P descriptor is similar to the other proposed descriptor, where the local microstructure information is extracted from the central patch of ROI using $BGPM_{(16;2)}$. However, instead of using BGP operator to assemble histogram vector from the macrostructure, a standard $LBP_{8,1}^{u2}$ operator is employed to extract the macrostructure information. The patch median of eight neighborhood patches is thresholded with the patch's median of the center patch, so as to produce a string of eight binaries or label. $LBP_{8,1}^{u2}$ descriptor generates over 256 labels, but only 58 uniform patterns are kept for histogram fusion and the remaining are discarded. Histograms from both domains are concatenated and given equal weights.

Results in **Table 2** expose that the Type II_P descriptor achieved better recognition rate than the rest of descriptors. The results also illustrate that Type II_P achieved better performance on images of distance 2 than those from distances 1 and 3. Furthermore, it is notable to mention that employing $BGPM_{(24;3)}$ at larger spatial resolution did not help much in improving the recognition rate as much as Type II_P has achieved.

4.3 Results of circular-based topology

As described in Section 3.2, the macrostructure information are exploited from the outer circle which always has larger spatial resolution $(P;R_2)$ than

Circular eBGP			Camera					
$S^i_{(P,R)}$	$S^o_{(P,R)}$	Type	1	2	3	4	5	Average
(8,1)	(16,2)	I_c	5.38	3.85	3.85	3.08	4.62	4.12
		II_c	6.92	6.92	6.92	6.15	7.69	6.92
	(24,3)	I_c	5.38	4.62	4.62	3.08	5.38	4.62
		II_c	7.69	5.38	6.92	7.69	6.92	6.92
	(32,4)	I_c	5.38	6.15	5.38	6.15	6.15	5.84
		II_c	6.92	6.15	6.92	8.46	6.15	6.92
	(40,5)	I_c	5.38	7.69	4.62	6.15	6.15	6.00
		II_c	9.23	7.69	6.92	7.69	6.15	7.54
(16,2)	(24,3)	I_c	5.38	4.62	6.15	3.85	6.15	5.23
		II_c	6.92	6.92	5.38	6.15	7.69	6.61
	(32,4)	I_c	6.15	6.92	7.69	4.62	6.15	6.31
		II_c	10.00	6.92	3.85	7.69	7.69	7.23
	(40,5)	I_c	5.38	6.92	3.85	6.15	6.15	5.69
		II_c	8.46	7.69	6.15	8.46	6.92	7.54
(24,3)	(32,4)	I_c	5.38	3.85	6.92	3.85	6.15	5.23
		II_c	12.31	7.69	7.69	9.23	6.92	8.77
	(40,5)	I_c	6.15	6.15	5.38	2.31	6.92	5.38
		II_c	10.77	8.46	9.23	9.23	4.62	8.46
(32,4)	(40,5)	I_c	4.62	5.38	6.15	2.31	7.69	5.23
		II_c	9.23	7.69	10.00	10.77	5.38	8.61
Baseline	$BGPM_{(8;1)}$		3.08	0.77	3.08	3.08	5.38	3.08
	$BGPM_{(16;2)}$		6.15	4.62	4.62	3.85	5.38	4.92

Table 3.
Circular-based topology on the SCface dataset at distance 1.

$(P;R_1)$. In other words, more points are used for thresholding when extracting the macrostructure information. For the presentation purpose, $s^i_{(P,R)}$ and $s^o_{(P,R)}$ notations are used to represent the spatial resolution of inner circle and outer circle, respectively. In the circular-based topology, two types of descriptors are designed to evaluate the performance of this topology. Type I_c descriptor is similar to what has been discussed in Section 3.2. Learning from the results obtained based on the patch-based topology, Type II_c descriptor is designed to explore a fusion of texture extracted from grayscale image and OIGM image. This descriptor extracts the local microstructure information from the OIGM image and the macrostructure information from the grayscale image. The histograms generated from these two types of images are concatenated and given equal weights. In this topology, multiple combinations of spatial resolution of inner and outer circles are tested. By limiting R_2 to 5, there are 10 combinations of descriptors at different spatial resolutions. Overall, there are 20 different combinations of descriptors that were put to the test.

Performance of Type I_c and Type II_c descriptors on the SCface dataset at distance 1, distance 2, and distance 3 is presented in **Tables 3, 4**, and **5**, respectively. Similar to the results obtained by the patch-based topology, the average recognition rate of the images that belong to distance 1 from all cameras is the lowest compared to

Circular eBGP			Camera					
$S^i_{(P,R)}$	$S^o_{(P,R)}$	Type	1	2	3	4	5	Average
(8,1)	(16,2)	I_c	20.77	12.31	10.00	11.54	14.62	13.85
		II_c	25.38	19.23	15.38	17.69	14.62	18.46
	(24,3)	I_c	24.62	15.38	11.54	15.38	16.92	16.77
		II_c	25.38	21.54	16.15	19.23	14.62	19.38
	(32,4)	I_c	26.92	17.69	15.38	17.69	13.85	18.31
		II_c	23.85	19.23	16.92	18.46	15.38	18.77
	(40,5)	I_c	29.23	19.23	13.08	19.23	13.85	18.92
		II_c	23.08	19.23	15.38	17.69	16.92	18.46
(16,2)	(24,3)	I_c	26.15	16.15	11.54	13.08	15.38	16.46
		II_c	25.38	22.31	16.15	21.54	19.23	20.92
	(32,4)	I_c	25.38	18.46	13.85	13.85	13.85	17.08
		II_c	24.62	21.54	17.69	21.54	20.00	21.08
	(40,5)	I_c	25.38	20.00	13.08	20.77	15.38	18.92
		II_c	24.62	20.77	16.92	20.00	17.69	20.00
(24,3)	(32,4)	I_c	20.77	18.46	12.31	13.85	14.62	16.00
		II_c	28.46	24.62	16.92	20.77	20.77	22.31
	(40,5)	I_c	22.31	17.69	14.62	16.15	14.62	17.08
		II_c	28.46	23.85	15.38	16.15	16.92	20.15
(32,4)	(40,5)	I_c	22.31	16.92	13.85	17.69	13.85	16.92
		II_c	25.38	25.38	16.92	20.00	16.15	20.77
Baseline	$BGPM_{(8;1)}$		16.15	12.31	6.92	11.54	13.85	12.15
	$BGPM_{(16;2)}$		23.85	13.85	7.69	12.31	13.08	14.16

Table 4.
Circular-based topology on the SCface dataset at distance 2.

those from distance 2 and distance 3 as shown in **Table 3**. One noteworthy observation is that most of Type II_c descriptors at any spatial resolution achieved better recognition rate than Type I_c descriptors. Taking a closer look at the descriptor's performance in **Table 5**, Type II_c descriptor with spatial resolution of $s^i_{(16,2)}$ and $s^o_{(24,3)}$ recorded the best results for all cameras on the test gallery of distance 3. On the other hand, for distance 2 test gallery, Type II_c descriptor with spatial resolution of $s^i_{(24,3)}$ and $s^o_{(32,4)}$ achieved the best result against other combinations.

For further evaluation, **Table 6** demonstrates results of the proposed eBGP descriptor compared with state-of-the-art descriptors such as PCA [27], SIFT and sparse representation-based classification (SRC) [28], and edge-preserving super-resolution (SR) [29], on the SCface database at distance 2. All descriptors applied the same test conditions, where only one mug shot image per subject is used for training, while the remaining low-resolution images from all cameras are used as probe images. The results show that the proposed descriptors based on eBGP achieved the highest recognition rates compared to all other descriptors, especially $eBGPM_{(16;2)}$ (Type II_P) which has the best recognition rate over all camera images. Exploiting information from the macrostructure raised the BGPM results from the fifth highest to first. This indicates the importance of the macrostructure information in shaping a complete face representation in single-reference face recognition problem.

Circular eBGP			Camera					
$S^i_{(P,R)}$	$S^o_{(P,R)}$	Type	1	2	3	4	5	Average
(8,1)	(16,2)	I_c	20.77	21.54	13.85	15.38	13.85	17.08
		II_c	25.38	26.15	20.00	23.85	13.85	21.85
	(24,3)	I_c	23.08	20.77	13.08	20.00	11.54	17.69
		II_c	23.08	24.62	20.00	23.85	16.92	21.69
	(32,4)	I_c	20.00	21.54	14.62	17.69	11.54	17.08
		II_c	20.77	24.62	17.69	21.54	14.62	19.85
	(40,5)	I_c	19.23	17.69	15.38	18.46	10.77	16.31
		II_c	23.85	23.85	15.38	20.77	13.85	19.54
(16,2)	(24,3)	I_c	20.77	20.77	13.08	17.69	13.08	17.08
		II_c	26.15	25.38	20.77	24.62	19.23	23.23
	(32,4)	I_c	20.77	18.46	16.15	19.23	10.00	16.92
		II_c	24.62	22.31	16.15	22.31	16.92	20.46
	(40,5)	I_c	19.23	19.23	15.38	18.46	12.31	16.92
		II_c	26.15	21.54	16.15	22.31	11.54	19.54
(24,3)	(32,4)	I_c	17.69	16.15	13.85	17.69	9.23	14.92
		II_c	23.08	20.77	19.23	21.54	15.38	20.00
	(40,5)	I_c	20.00	16.15	13.85	19.23	10.77	16.00
		II_c	23.85	21.54	16.92	18.46	16.15	19.38
(32,4)	(40,5)	I_c	16.15	15.38	13.08	18.46	10.00	14.61
		II_c	20.77	20.77	16.92	21.54	10.77	18.15
Baseline	$BGPM_{(8;1)}$		15.38	19.23	10.00	16.92	11.54	14.61
	$BGPM_{(16;2)}$		18.46	20.00	16.15	14.62	11.54	16.15

Table 5.
Circular-based topology on the SCface dataset at distance 3.

Descriptor	Camera					
	1	2	3	4	5	Average
PCA [27]	7.70	7.70	3.90	3.90	7.70	6.18
SIFT [28]	13.08	12.31	8.46	15.38	9.23	11.69
$BGPM_{(16;2)}$	23.85	13.85	7.69	12.31	13.08	14.16
SRC [28]	29.23	16.15	12.31	25.38	13.08	19.23
Edge-preserving SR [29]	26.92	21.54	15.38	24.61	15.38	20.77
$eBGPM_{(24;3)(32;4)}$ (circular)	28.46	24.62	16.92	20.77	20.77	22.31
$eBGPM_{(16;2)}$ (Type II_P)	34.62	25.38	20.00	25.38	21.54	25.38

Table 6.
Comparison of recognition rate (%) of the proposed eBGP descriptor with state-of-the-art descriptors on the SCface database at distance 2.

5. Conclusion

In this paper, an extended BGP (eBGP) descriptor, which incorporates macrostructure information into BGP descriptor, has been proposed to improve the overall descriptor performance in single-reference face recognition problem.

Results obtained from a series of experiments on the SCface database showed that a fusion of information extracted from micro- and macrostructures is capable of boosting up the performance of BGP descriptor. The proposed eBGP descriptor was tested with the patch-based and circular-based topologies; in overall, the circular-based topology outperformed the patch-based topology in terms of recognition rate. In patch-based topology, 5×5 structure recorded better hike in recognition rate than 3×3 structure, while in circular-based topology, larger spatial resolution showed better hike in the recognition performance. Moreover, a fusion of micro- and macrostructure information extracted from OIGM and grayscale image, respectively, raised the recognition rate higher. In fact, Type II_c setup always illustrated a better performance boost than Type I_c. With regard to thresholding implementation, it is worth to mention that local mean is on par with the local median for the descriptor and does not offer additional boost in the patch-based topology.

Acknowledgements

Author details

Nuzrul Fahmi Nordin[1], Samsul Setumin[1,2], Abduljalil Radman[1,3] and Shahrel Azmin Suandi[1*]

1 Intelligent Biometric Group, School of Electrical and Electronics Engineering, Universiti Sains Malaysia, Nibong Tebal, Malaysia

2 Faculty of Electrical Engineering, Universiti Teknologi MARA Pulau Pinang, Permatang Pauh, Malaysia

3 Faculty of Engineering and Information Technology, Taiz University, Taiz, Yemen

*Address all correspondence to: shahrel@usm.my

IntechOpen

References

[1] Radman A, Suandi SA. Robust face pseudo-sketch synthesis and recognition using morphological-arithmetic operations and HOG-PCA. Multimedia Tools and Applications. 2018;**77**(19):25311-25332

[2] Matta F, Dugelay J-L. Person recognition using facial video information: A state of the art. Journal of Visual Languages and Computing. 2009;**20**(3):180-187

[3] De-la-Torre M, Granger E, Radtke PVW, Sabourin R, Gorodnichy DO. Partially-supervised learning from facial trajectories for face recognition in video surveillance. Information Fusion. 2015;**24**:31-53

[4] Zakaria Z, Suandi SA, Mohamad-Saleh J. Hierarchical skin-AdaBoost-neural network (H-SKANN) for multi-face detection. Applied Soft Computing. 2018;**68**:172-190

[5] Lowe DG. Object recognition from local scale-invariant features. In: Proceedings of the Seventh IEEE International Conference on Computer Vision. 20-27 September; Kerkyra, Greece; 1999. pp. 1150-1157

[6] Bay H, Tuytelaars T, Van Gool L. Surf: Speeded up robust features. In: European Conference on Computer Vision. 7-13 May; Graz, Austria; 2006. pp. 404-417

[7] Ahonen T, Hadid A, Pietikainen M. Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006;**28**(12):2037-2041

[8] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: International Conference on Computer Vision and Pattern Recognition. 20-25 June; San Diego, CA; 2005. pp. 886-893

[9] Wold S, Esbensen K, Geladi P. Principal component analysis. Chemometrics and Intelligent Laboratory Systems. 1987;**2**(1-3):37-52

[10] Belhumeur PN, Hespanha JP, Kriegman DJ. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1997;**19**(7):711-720

[11] Bartlett MS, Movellan JR, Sejnowski TJ. Face recognition by independent component analysis. IEEE Transactions on Neural Networks. 2002;**13**(6):1450-1464

[12] Ren J, Jiang X, Yuan J. Noise-resistant local binary pattern with an embedded error-correction mechanism. IEEE Transactions on Image Processing. 2013;**22**(10):4049-4060

[13] Ojala T, Pietikäinen M, Mäenpää T. Gray scale and rotation invariant texture classification with local binary patterns. In: European Conference on Computer Vision. June 26-July 1; Dublin, Ireland; 2000. pp. 404-420

[14] Huang W, Yin H. Robust face recognition with structural binary gradient patterns. Pattern Recognition. 2017;**68**:126-140

[15] Liu L, Lao S, Fieguth PW, Guo Y, Wang X, Pietikäinen M. Median robust extended local binary pattern for texture classification. IEEE Transactions on Image Processing. 2016;**25**(3):1368-1381

[16] Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions. Pattern Recognition. 1996;**29**(1):51-59

[17] Lee K-C, Ho J, Kriegman DJ. Acquiring linear subspaces for face

recognition under variable lighting. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2005;**27**(5):684-698

[18] Martinez AM. The AR face database. CVC Tech. Report. 1998. 24

[19] Gross R, Matthews I, Cohn J, Kanade T, Baker S. Multi-PIE. Image and Vision Computing. 2010;**28**(5):807-813

[20] Phillips PJ, Rizvi SA, Rauss PJ. The FERET evaluation methodology for face-recognition algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000;**22**(10):1090-1104

[21] Huang GB, Ramesh M, Berg T, Learned-Miller E. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. In: European Conference on Computer Vision Workshop on Faces in Real-life Images. October 200. pp.1-11

[22] Grgic M, Delac K, Grgic S. SCface—surveillance cameras face database. Multimedia Tools Applications. 2011;**51**(3):863-879

[23] Liu L, Fieguth P, Zhao G, Pietikäinen M, Hu D. Extended local binary patterns for face recognition. Information Sciences. 2016;**358**:56-72

[24] Liao S, Zhu X, Lei Z, Zhang L, Li SZ. Learning multi-scale block local binary patterns for face recognition. In: International Conference on Biometrics. 27-29 August 2007; Seoul, Korea;. pp. 828-837

[25] Liu L, Fieguth P, Guo Y, Wang X, Pietikäinen M. Local binary features for texture classification: Taxonomy and experimental study. Pattern Recognition. 2017;**62**:135-160

[26] Liu L, Zhao L, Long Y, Kuang G, Fieguth P. Extended local binary patterns for texture classification.

Image and Vision Computing. 2012;**30**(2):86-99

[27] Martínez AM, Kak AC. Pca versus lda. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001;**23**(2):228-233

[28] Hu X, Peng S, Wang L, Yang Z, Li Z. Surveillance video face recognition with single sample per person based on 3D modeling and blurring. Neurocomputing. 2017;**235**:46-58

[29] Mandal S, Thavalengal S, Sao AK. Explicit and implicit employment of edge-related information in super-resolving distant faces for recognition. Pattern Analysis and Applications. 2016;**19**(3):867-884