A ZigZag Pattern of Local Extremum Logarithm Difference for Illumination-Invariant and Heterogeneous Face Recognition

Hiranmoy Roy[1(✉)] [iD] and Debotosh Bhattacharjee[2]

[1] Department of Information Technology, RCC Institute of Information Technology, Canal South Road, Beliaghata, Kolkata 700015, India
hiranmoy.roy@rcciit.org

[2] Department of Computer Science and Engineering, Jadavpur University, Kolkata 700032, India
debotoshb@hotmail.com

Abstract. A novel methodology for matching of illumination-invariant and heterogeneous faces is proposed here. We present a novel image representation called local extremum logarithm difference (*LELD*). Theoretical analysis proves that *LELD* is an illumination-invariant edge feature in coarse level. Since edges are invariant in different modalities, more importance is given on edges. Finally, a novel local zigzag binary pattern *LZZBP* is presented to capture the local variation of *LELD*, and we call it a zigzag pattern of local extremum logarithm difference (*ZZPLELD*). For refinement of *ZZPLELD*, a model based weight value learning is suggested. We tested the proposed methodology on different illumination variations, sketch-photo and NIR-VIS benchmark databases. Rank-1 recognition of 96.93% on CMU-PIE database and 95.81% on Extended Yale B database under varying illumination, show that *ZZPLELD* is an efficient method for illumination invariant face recognition. In the case of viewed sketches, the rank-1 recognition accuracy of 98.05% is achieved on CUFSF database. In the case of NIR-VIS matching, the rank-1 accuracy of 99.69% is achieved and which is superior to other state-of-the-art methods.

Keywords: Illumination-reflectance model
Modality-invariant feature · Illumination-invariant feature
Local extremum logarithm difference · Local zigzag binary pattern
Heterogeneous face recognition

1 Introduction

Biometric authentication is becoming a ubiquitous component of any modern systems, such as mobile phone, smart TV, computer, etc. There is a set of unique biometric features, such as fingerprints, faces, retinas, DNA samples, ears, etc. Among these features, faces are the most easily available and easily recognizable feature. The valuable part of any face biometrics is that the recognition

or authenticate can be done without any expertise. However, by the naked eye without an expert, it is quite impossible to authenticate depending on fingerprints, DNA samples, retinas and ears. Another advantage of face biometric is that a single face consists of a bundle of unique biometric features like eyes, nose, etc. Therefore, in recent years, a lot of work has been done on face biometric authentication based on real life applications. Depending on the broad range of applications, different face recognition algorithms have already been proposed by many researchers [43–45]. The performances of these algorithms are primarily tested on face images collected under well-controlled studio conditions. In effect, most of them have found difficulty in the case of handling natural images, which are captured under different illumination variations. With the variations of illumination, the result degrades drastically. On the other hand, different real life applications need faces captured in different situations. Near-infrared cameras are used to capture faces at night, and for illumination-invariant face recognition [1]. To detect liveness by capturing body heat, thermal-infrared (TIR) cameras are used. Often, it may happen that there are no available fingerprints, no available DNA samples and devices, which are present, have captured poor quality images. In those situations, the only solution is the use of face sketches generated by interviewing the eye witness. Hence we can say that various scenarios and necessities create different modalities of faces. In situations like this, it becomes difficult to use conventional face recognition systems. Therefore, an interesting and challenging field of face biometric recognition has emerged for forensics, called heterogeneous face recognition (HFR) [2].

The problem of heterogeneous face recognition has received increasing attention in recent years. Up to now, many different techniques has been proposed in the literature to solve the problem. We can easily classify these solutions into three broad categories: image synthesis based methods, common subspace learning based methods and modality-invariant feature representation based methods.

– **Image synthesis:** In this category, a pseudo-face or pseudo-sketch is generated using synthesis techniques to transform one modality image into another modality and then some classification technique is used. The pioneering work of Tang and Wang [3], where they introduced an eigen transformation based sketch-photo synthesis method. The same mechanism was also used by Chen *et al.* [4] for NIR-VIS synthesis. Gao *et al.* [5] proposed an embedded hidden markov model and a selective ensemble strategy to synthesise sketches from photos. Wang and Tang [6] again proposed a patch based Markov Random Field (MRF) model for the sketch-photo synthesis. Li *et al.* [7] used the same MRF model for TIR-VIS synthesis. Gao *et al.* [8] proposed a sparse representation based pseudo-sketch or pseudo-photo synthesis. Another sparse feature selection (SFS) and support vector regression for synthesis was proposed by Wang *et al.* [9]. Wang *et al.* [10] proposed a transductive learning based face sketch-photo synthesis (TFSPS) framework. Recently, Peng *et al.* [11] proposed a multiple representation based face sketch-photo synthesis.
– **Common Subspace learning:** In this category, different modality face images are projected into a subspace for learning. Lin and Tang [12]

introduced a common discriminant feature extraction (CDFE) for face sketch-photo recognition. Yi *et al.* [13] proposed a canonical correlation analysis based regression method for NIR-VIS face images. A coupled spectral regression (CSR) based learning for NIR-VIS face images was proposed by Lei and Li [14]. A partial least square (PLS) based subspace learning method was proposed by Sharma and Jacobs [15]. Mignon and Jurie [16] proposed a cross modal metric learning (CMML) for heterogeneous face matching. Lei *et al.* [17], proposed a coupled discriminant analysis for HFR. A multi-view discriminant analysis (MvDA) technique for single discriminant common space generation was proposed by Kan *et al.* [18].

– **Modality-invariant feature representation:** In this category, images of different modalities are represented using some modality-invariant feature representation. Liao *et al.* [19] used difference of Gaussian (DoG) filter and multi-block local binary pattern (MB-LBP) features for both NIR and VIS face images. Klare *et al.* [20] employed the scale invariant feature transform (SIFT) and multi-scale local binary pattern (MLBP) features for forensic sketch recognition. A coupled information-theoretic encoding (CITE) feature was proposed by Zhang *et al.* [21]. Bhatt *et al.* [22] used a multi-scale circular Weber's local descriptor (MCWLD) for semi-forensic sketch matching. Klare and Jain [23] proposed a kernel prototype random subspace (KP-RS) on MLBP features. Zhu *et al.* [24] used a Log-DoG filter based LBP and a histogram of oriented gradient (HOG) features with transductive learning (THFM) for NIR-VIS face images. Gong *et al.* [25] combined histogram of gradients (HOG) and multi-scale local binary pattern (MLBP) with canonical correlation analysis (MCCA). Roy and Bhattacharjee [26] proposed a geometric edge-texture (GETF) based feature with hybrid multiple fuzzy classifier for HFR. Roy and Bhattacharjee [27] again proposed an illumination invariant local gravity face (LG-face) for HFR. A local gradient checksum (LGCS) feature for face sketch-photo matching was proposed by Roy and Bhattacharjee [28]. Another local gradient fuzzy pattern (LGFP) based on restricted equivalent function for face sketch-photo recognition was again proposed by Roy and Bhattacharjee [29]. A graphical representation based HFR (G-HFR) was proposed by Peng *et al.* [31]. Recently, Roy and Bhattacharjee [30] proposed another edge texture based feature called quaternary pattern of local maximum quotient (QPLMQ) for HFR.

In the synthesis based category, more concentration is applied in the synthesis. The synthesis method itself is a time-consuming technique, where synthesis technique is repeated several times to get a better pseudo sketch or photo. Again, the synthesis mechanism is depending on image quality, and modality i.e. "task-specific". In the common subspace learning category, both modality images are projected into another domain. The projection technique requires huge training data and also generates some loss of information. Due to the loss of information, the accuracy of recognition is also reduced. In the modality-invariant feature representation, local hand-crafted features are directly used, which means no loss of local information and algorithms are more time saving than the other

two categories. One and the only problem in this category is to recognize or search features, which are either common to different modalities or invariant in different modalities.

Modality-invariant feature representation methods are neither time-consuming and task-specific synthesis, nor common subspace based learning, but able to consider the local spatial features. Motivated by the advantages of modality-invariant feature representation, in this paper, we propose a modality-invariant feature representation for different modality images. The existing modality-invariant methods [19–25] tried to generate modality-invariant features using the existing popular hand-crafted features. More emphasis is given on classifier. No one tried to develop new modality-invariant features, except [31]. In [26–29], we developed some new hand-crafted features which are actually modality-invariant. From our visual inspection, we conclude that edges are the most important modality-invariant feature. Psychological studies also says that we can recognize a face from its edges only [32]. It is easy to understand that facial components in a face have maximum edges and they belong to high-frequency components of an image. Another feature i.e. texture information is also important for face matching. Since edges and textures in a face image are sensitive to illumination variations, an illumination-invariant domain with the capability of capturing high-frequency information is necessary. LGCS [28] was only applied on sketch-photo recognition CHUCK database. LGFP [29] was developed only for sketch-photo recognition. In GETF [26] and LG-face [27], we tested the features for both sketch-photo and NIR-VIS recognition. Although the results were impressive, the NIR-VIS existing database (CASIA-HFB) has a very small size and still the result was not 100%. Therefore, we need new methods to handle huge databases, where faces not only have variation in modality, but also in pose, illumination, expression and the obstacle. The goal of the proposed method is to recognize the facial features, which are invariant in different modalities, and illumination.

The artist gives more attention towards edges and texture information at the time of drawing a sketch. In the case of NIR images, the high-frequency information is captured. Therefore, selection of edge and texture features for modality-invariant representation is correct in the sense. We propose an illumination-invariant image representation called local extremum logarithm difference ($LELD$), which is a modification of the work explained in [33]. $LELD$ gives only high-frequency image representation at a coarse level. A local micro level feature representation is also important to capture local texture information. Motivated by the superior output results of the local binary pattern (LBP) [34] and LBP-like features in face and texture recognition, we propose one novel local zig zag binary pattern ($LZZBP$). $LZZBP$ measure the binary relation between pixels, which are in a zigzag position in a local square window. $LZZBP$ captures more edge and texture patterns than LBP. Finally, the combination of $LELD$ and $LZZBP$ gives the proposed modality-invariant feature representation for HFR and we call it a zigzag pattern of local extremum logarithm difference ($ZZPLELD$).

Experimental results on different HFR databases show the excellent performance of the proposed methodology. The major contributions are:

1. *LELD* is proposed for capturing illumination-invariant image representation.
2. *LZZBP* is developed to capture the relation between pixels in a zigzag position of a square window.
3. *ZZPLELD* is developed to capture the local texture and edge patterns of the modality-invariant key facial features.
4. Weights learning model is developed to enhance the discrimination power of the proposed *ZZPLELD*.

This paper is organized as follows: in Sect. 2, the proposed *ZZPLELD* is described in detail. Experimental results and comparisons are presented in Sect. 3, and finally, the paper concludes with Sect. 4.

2   Proposed Work

In this section, we introduce the way we extract the modality-invariant features for HFR. We start with a detailed idea about illumination-invariant *LELD* based image representation. Finally, we conclude with a detail description of the proposed *ZZPLELD* feature.

2.1   Local Extremum Logarithm Difference Based Image Representation

In any face recognition system, one of the main problems is the presence of illumination variations. Due to the presence of illumination variations, the intra-class variation between faces also increase heavily. At the same time, we consider edges as our modality-invariant feature and edges are also sensitive towards illumination. Therefore, we need an illumination-invariant image representation for extracting better edge information. According to the Illumination-Reflectance Model (IRM) [35,36], a gray face image $I(x,y)$ at each point (x,y) is expressed as the product of the reflectance component $R(x,y)$ and the illumination component $L(x,y)$, as shown in Eq. 1

$$I(x,y) = R(x,y) \times L(x,y) \tag{1}$$

Here, the R component consists of information about key facial points, and edges, whereas the L component represents only the amount of light falling on the face. Now, after the elimination of the L component from a face image, the R component is still able to represent the key facial features and edges, which are the most important information for our modality-invariant feature representation. Moreover, the L component corresponds to the low-frequency part of an image, whereas the R component corresponds to the high-frequency part. One widely accepted assumption in the literature [27,37] is that L remains approximately constant over a local 3×3 neighborhood.

In literature, a wide range of approaches has been proposed to reduce the illumination effect. In those methods, mainly two different mathematical operations are used: division and subtraction. Methods like [37,38] used division operation and methods like [33,42] used subtraction operation. In the case of subtraction operation, at first, the image is converted to the logarithmic domain to convert the multiplicative IRM into an additive one, as shown in Eq. 3. Since, the division operation is "ill-posed and not robust in numerical calculations" [33] due to the problem of divided by zero, it is better to apply subtraction operation to eliminate the L component.

Lai *et al.* [33] proposed a multiscale logarithm difference edgemaps (MSLDE), which used logarithmic domain and subtraction operation to compensate the illumination effect. They calculated a local logarithm difference using the following equation:

$$MSLDE = \sum_{(\hat{x},\hat{y})\epsilon\aleph_{(x,y)}} \left(I_{log}(x,y) - I_{log}(\hat{x},\hat{y})\right), \forall x, \forall y \tag{2}$$

where $\aleph_{(x,y)}$ is a square neighborhood surrounding a center pixel at (x,y). In MSLDE, only the R component is present. Therefore, MSLDE is an illumination-invariant method. In this method authors considered a long range of square neighborhood from 3×3 to 13×13 and all the logarithm difference values are added together. Now, the question is does the L component be constant for such a long neighborhood i.e. 13×13. We know that R component belongs to high-frequency. Similarly edge and noise also belong to high-frequency. There is no doubt that MSLDE increases the edge information, which is very important in face recognition, by adding all logarithm differences. However, it is also increasing the noise, which causes degradation in true edge detection. The effect of noise is clearly visible in Fig. 2(d), where the proposed MSLDE provides too much false edge information and which are nothing but noise. To solve both the problems i.e. large neighborhood size and presence of noise, we consider the maximum and minimum logarithm difference in a 3×3 neighborhood (as shown in Fig. 1) and we call it local extremum logarithm difference ($LELD$).

Theorem 1. *Given an arbitrary image $I(x,y)$ with illumination effect. The local extremum logarithm difference (LELD) between central pixel and its 3×3 neighborhood is an illumination invariant feature.*

Proof. Let us consider a local 3×3 window and the center pixel is 'c', as shown in Fig. 1. Applying logarithmic domain, the IRM (1) multiplicative model is converted to additive model as follows

$$I_{log}(x,y) = log\left(R\left(x,y\right) \times L\left(x,y\right)\right)$$
$$\Rightarrow I_{log}(x,y) = log\left(R\left(x,y\right)\right) + log\left(L\left(x,y\right)\right) \tag{3}$$

where, $I_{log}(x,y)$ is the logarithm representation of the image pixel $I(x,y)$. Now, let us measure the difference between central pixel against all its 3×3 neighbors and the logarithm differences (LD) are represented as follows

$$LD = \{I_{log}(x_i,y_i) - I_{log}(x_c,y_c)\}, (x_i,y_i)\epsilon\aleph_{(x_c,y_c)} \tag{4}$$

I (x_4, y_4)	I (x_3, y_3)	I (x_2, y_2)
I (x_5, y_5)	I (x_c, y_c)	I (x_1, y_1)
I (x_6, y_6)	I (x_7, y_7)	I (x_8, y_8)

Fig. 1. Central pixel and its 3×3 neighborhood.

where $\aleph_{(x,y)}$ is a 3×3 square neighborhood surrounding a center pixel at (x_c, y_c). Putting the representation of $I_{log}(x, y)$ Eq. (3) into Eq. 4, we have:

$$LD = \{(log\,(R\,(x_i, y_i)) + log\,(L\,(x_i, y_i))) - (log\,(R\,(x_c, y_c)) + log\,(L\,(x_c, y_c)))\}$$
$$\Rightarrow LD = \{(log\,(R\,(x_i, y_i)) - log\,(R\,(x_c, y_c))) + (log\,(L\,(x_i, y_i)) - log\,(L\,(x_c, y_c)))\}\,(5)$$

Based on the widely accepted assumption that L varies slowly, we can say L is almost equal to a local 3×3 window. Therefore, we can write:

$$L(x_i, y_i) = L(x_c, y_c)$$
$$\Rightarrow log\,(L\,(x_i, y_i)) = log\,(L\,(x_c, y_c))$$
$$\Rightarrow log\,(L\,(x_i, y_i)) - log\,(L\,(x_c, y_c)) = 0 \qquad (6)$$

Now, the Eq. 5 is modified using Eq. 6 as follows:

$$LD = \{log\,(R\,(x_i, y_i)) - log\,(R\,(x_c, y_c))\} \qquad (7)$$

The maximum and minimum logarithm differences are calculated as follows:

$$LD_{max} = max\,\{log\,(R\,(x_i, y_i)) - log\,(R\,(x_c, y_c))\}, \forall x_c, \forall y_c \qquad (8)$$
$$LD_{min} = min\,\{log\,(R\,(x_i, y_i)) - log\,(R\,(x_c, y_c))\}, \forall x_c, \forall y_c \qquad (9)$$

Therefore, the local extremum logarithm difference ($LELD$) is as follows:

$$LELD = \{LD_{max}, LD_{min}\}, \forall x_c, \forall y_c \qquad (10)$$

Now, in Eq. 10, we can see that '$LELD$' consists of R-part of the IRM. According to this model, R is the illumination invariant feature. Thus, it is proved that the local extremum logarithm difference between central pixel and its 3×3 neighborhood is an illumination invariant feature.
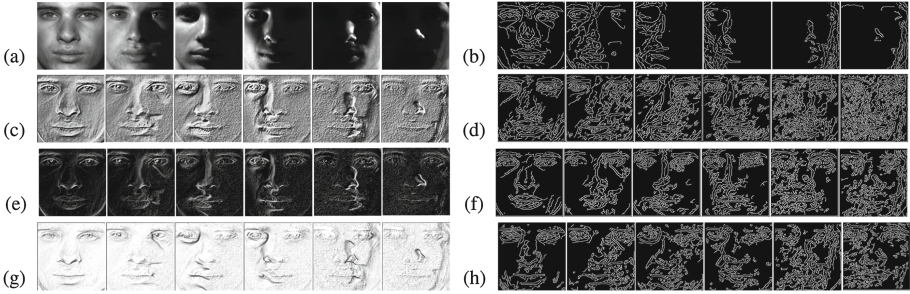
Fig. 2. (a) The original images with illumination variations, (b) The canny edge images for (a) images, (c) The corresponding MSLDE images of (a) images as proposed in [33], (d) The canny edge images of MSLDE images, (e) The proposed *LELD* images of (a) images with maximum difference, (f) The canny edge images of (e) images, (g) The proposed *LELD* images of (a) images with minimum difference, (h) The canny edge images of (g) images.

Since we are considering the smallest square neighborhood i.e. 3×3, the assumption that L is constant in the neighborhood is theoretically true. Again, we are avoiding total sum of differences, therefore, presence of noise is reduced. The results of proposed *LELD* are shown in Fig. 2(e)–(h) and it gives better edge detection result than MSLDE. Finally, we have two different *LELD* methods to convert the different modality face images into an illumination-invariant domain, where important facial key values and edges are almost intact.

2.2   Local ZigZag Binary Pattern for *ZZPLELD* Generation

Local binary pattern (LBP) [34] has been used successfully in many fields of image processing and pattern classification problems. It is capable to represent local features in micro structures. LBP implements the binary relation of each and every neighboring pixels with respect to center pixel i.e. if neighboring pixel is greater than or equal to center pixel, then binary value '1' otherwise '0'. Although LBP captures the binary relations between surrounding neighboring pixels with the center pixel, it is not able to capture edge information properly, mainly in diagonal direction. Since our modality-invariant feature is edge related, a local pattern having good edge capturing capability is important. Inspired from the zigzag scanning pattern used for MPEG data compression in discrete cosine transform (DCT) domain, we developed a zigzag binary pattern for the pixels in a zigzag position of a square mask. Figure 3 shows the positions of 3×3 and 5×5 zigzag scanning used in our experiment. We use a left to right zigzag scanning (considering top left pixel as the starting point) and right to left zigzag scanning (considering top right pixel as the starting point) in each square mask. Figure 3(a) shows the 3×3 left zigzag scanning, Fig. 3(b) shows the 3×3 right zigzag scanning and Fig. 3(c) shows the 5×5 left zigzag scanning. Again, we consider only 8 bits binary patterns to make the histogram feature vector length

up to 256 bins. In case of 5×5 image pixels we can have a total of 24 bits binary string. To make it with in a small range, we divide it into three 8 bits binary string. Different arrows in Fig. 3(c) shows different sequences of binary string.

Let, the image pixels in a square window are collected in a zigzag pattern, as shown in Fig. 3(a). Then, the pixels are stored in a linear array $Z_P = \{g_1, g_2, g_3, \ldots, g_P\}$, where P is the number of pixels in a square window. For 3×3 window P is 9 and for 5×5 window it is 25. Then, we calculate the binary relation between those consecutive pixels according to the following equation (for $P = 8$):

$$LZZBP = \sum_{i=1}^{P-1} 2^{(i-1)} \times f\left(g_{i+1} - g_i\right)$$

$$f(a) = \begin{cases} 1, & if \quad a \geq 0 \\ 0, & otherwise \end{cases} \qquad (11)$$

In case of 5×5 window the 24 bits binary string is first broken into 3 parts of 8 bits string and then converted to three separate patterns.
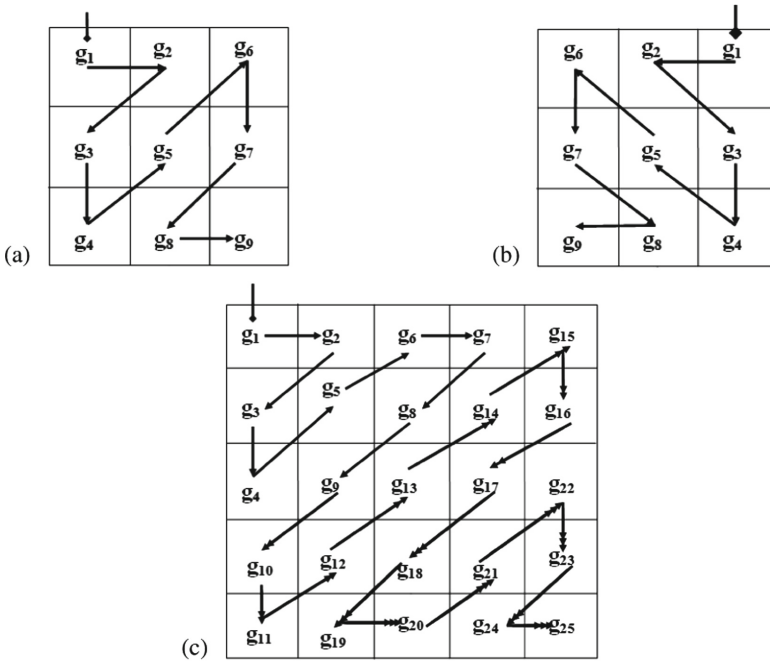

Fig. 3. (a) Zigzag scanning in a 3×3 window starting at top left corner, (b) Zigzag scanning in a 3×3 window starting at top right corner, (c) Zigzag scanning in a 5×5 window starting at top left corner. Here 3 different arrows are used to represent 3 different sequences of 8 bits binary strings.

Fig. 4. (a) Sample image and its corresponding LBP and *LZZBP* images. The 2nd image is the 3×3 LBP image and rest 3rd–4th are 3×3 *LZZBP* images. (b) The 2nd image is the 5×5 LBP image and rest 3rd–8th are 5×5 *LZZBP* images.

The proposed zigzag binary pattern gives better edge preserving texture information. The different LBP images and *LZZBP* images in Fig. 4 shows the edge preserving properties of *LZZBP*. The key facial features are preserved far better than normal LBP in *LZZBP*.

Finally, the proposed *LZZBP* is applied on *LELD* image representation to measure the local patterns of *LELD* image and which give our proposed *ZZPLELD*. For two different extremum *LELD* i.e. maximum and minimum differences, we get 4 different *ZZPLELD* results after applying 3×3 *LZZBP* and 12 different *ZZPLELD* results after applying 5×5 *LZZBP*. Therefore, altogether 16 different *ZZPLELD* images. The image features are represented in the form of histogram bins.

2.3   Similarity Measure

The whole *ZZPLELD* image is divided into a set of non-overlapping square blocks with dimension $w_b \times w_b$. Then, the histogram of each square block is measured. Finally, all the histograms measured from all the blocks of all the different (16) *ZZPLELD* images are concatenated to obtain the final face feature vector. Here, we use the nearest neighbor (NN) classifier with histogram intersection distance measure. Therefore, for a query image (I_q) with a concatenated histogram $H_q^{i,j}$ and a gallery image (I_G) with a concatenated histogram $H_G^{i,j}$, the similarity measure for a particular level of *ZZPLELD* is given as follows:

$$S^k(I_q, I_G) = -\sum_{i,j} min\left(H_q^{i,j}, H_G^{i,j}\right) \qquad (12)$$

where, $S^k(I_q, I_G)$ is the similarity score of kth *ZZPLELD* level of both query and gallery image; (i, j) is the jth bin of ith block. Square block selection is another essential task. In this paper, we chose 3×3 and 5×5 windows pixels for *LZZBP* and different block sizes $w_b = 6, 8, 10, 12$ for histogram matching.

2.4   Weighted ZZPLELD Model

To enhance the performance of *ZZPLELD*, we assign a weight value to each *ZZPLELD* patterns. A set of weight values is measured depending on the discriminating ratio proposed in the linear discriminating analysis (LDA). This weight calculation needs a training set consisting of few people and their corresponding different modality images. After calculating the weights from the

training set, the values are applied to rest of the database images. A detail idea is given below.

Let, there are 'C' numbers of different persons i.e. 'C' number of different classes in the training set. We have 1 to N numbers of different $ZZPLELD$ for one face image (where N is the number of different $ZZPLELD$ patterns). Since there are two different $LELD$ images, for each $LELD$ two different 3×3 $LZZBP$ and two different 5×5 $LZZBP$. Again, each 5×5 $LZZBP$ is partitioned into three different 8-bits patterns. Therefore, altogether $10(2 + 2 + 3 + 3 = 10)$ different 8-bits patterns. Thus the value of N is 10. Now, we calculate the class-wise mean (μ_i^C) and overall mean (μ_i) for the ith $ZZPLELD$.

$$\mu_i^C = \frac{1}{n} \sum_{j=1}^{n} ZZPLELD_i^{C,j} \tag{13}$$

where, 'j' denotes the jth number of different modality images (total number of images n, which is either 2 for sketch-photo and 4 for NIR-VIS) of a particular person 'C' and $i \in (1, 2, \ldots, N)$.

$$\mu_i = \frac{1}{C \times n} \sum_{k=1}^{C} \sum_{j=1}^{n} ZZPLELD_i^{C,j} \tag{14}$$

Then, we calculate within-class error (e_w) and between-class error (e_b):

$$e_w = \sum_{k=1}^{C} \sum_{j=1}^{n} |ZZPLELD_i^{C,j} - \mu_i^C| \tag{15}$$

$$e_b = n \times \sum_{k=1}^{C} |\mu_i^C - \mu_i| \tag{16}$$

If the $ZZPLELD$ has a high discriminating ability, then e_b should be relatively large compared with e_w. Hence, we set the ratio between this two as our weights (α_i). The normalized weights (α_i^{norm}) are

$$\alpha_i = \frac{e_b}{e_w} \tag{17}$$

$$\alpha_i^{norm} = \frac{\alpha_i}{\sum_{i=1}^{P} \alpha_i} \tag{18}$$

Therefore, for each different $ZZPLELD$ pattern we have different α_i^{norm}. Now, the similarity between two images, as given in Eq. 12, is modified into the following equation:

$$Sim(I_q, I_G) = \sum_{k=1}^{P} \alpha_k \times S^k(I_q, I_G) \tag{19}$$

The value of α also varies from database to database. Different α values measured during training in different databases are shown in Table 1. The results of weighted $ZZPLELD$ are superior than normal $ZZPLELD$. The recognition accuracy results on different databases are given in a form of tables in the next section and weight learning $ZZPLELD$ is represented as $ZZPLELD$ (weighted).

Table 1. Different level weight values for ZZPLELD, measured from training data of CMU-PIE, Extended Yale B, CUFSF, and CASIA NIR-VIS databases

Levels of weight values	CMU-PIE	Extended Yale B	CUFSF	CASIA NIR-VIS
α_1	0.1021	0.0734	0.0471	0.0841
α_2	0.0371	0.0576	0.0501	0.0819
α_3	0.1156	0.1008	0.0718	0.0812
α_4	0.1097	0.0995	0.0695	0.0964
α_5	0.1015	0.1134	0.1094	0.1101
α_6	0.1807	0.1693	0.1103	0.1103
α_7	0.1503	0.1412	0.1378	0.1145
α_8	0.0962	0.1139	0.1230	0.1039
α_9	0.1013	0.1200	0.1301	0.1151
α_{10}	0.0055	0.0109	0.1509	0.1025

3 Experimental Results

In this section, *ZZPLELD* is evaluated on one illumination variation scenario and two different HFR scenarios, i.e. face sketch vs. photo recognition and NIR image vs. VIS image recognition respectively on the existing benchmark databases. For illumination variation scenario, we tested our proposed method on CMU-PIE [46], and Extended Yale B Face Database [39,40]. For face sketch vs. photo recognition, we tested the proposed method on the CUHK Face Sketch FERET Database (CUFSF) [21]. CASIA-HFB Face Database [41] is used for NIR face image vs VIS face recognition testing.

At first, proposed *ZZPLELD* is tested on illumination-invariant face recognition. We compared *ZZPLELD* with several other methods, namely LBP, Gradient-face [37], TVQI [38], HF+HQ [42], MSLDE [33] on CMU-PIE and Extended Yale B Face Databases. All the methods mentioned above are well tuned according to their respective published papers.

We compared *ZZPLELD* with several state-of-the-art methods, namely, PLS [15], CITE [21], MCCA [25], TFSPS [10], KP-RS [23], MvDA [18], G-HFR [31], and LGFP [28] on viewed sketch database. We also compared *ZZPLELD* with several state-of-the-art methods, namely, KP-RS, LCKS-CSR [17], and THFM [24] on NIR face image vs VIS face image database. Experimental setups (training and testing samples) and accuracies of the methods mentioned above, except LGFP, are taken from the published papers.

3.1 Rank-1 Recognition Results on CMU-PIE

This database contains 41368 images of 68 different subjects. We have tested our proposed method on the illumination subset "C27" with 1428 images of 68 subjects. All the images are in frontal face with pose 27 and 21 different lighting
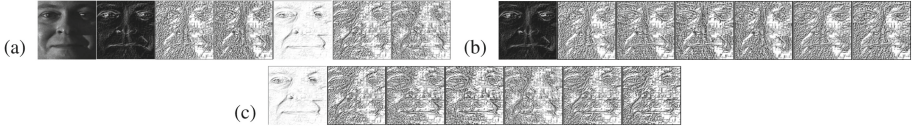
Fig. 5. (a) Sample *ZZPLELD* images after applying 3×3 *LZZBP* on both extremum *LELD* from CMU-PIE Face database. (b) Sample *ZZPLELD* images after applying 5×5 *LZZBP* on maximum *LELD* from CMU-PIE Face database. (c) Sample *ZZPLELD* images after applying 5×5 *LZZBP* on minimum *LELD* from CMU-PIE Face database.

Table 2. Average rank-1 recognition rates for different methods on CMU-PIE database of 68 subjects

State-of-the-art methods	Average rank-1 accuracy (%)
LBP	69.15
Gradient-face	93.66
TVQI	92.94
HF+HQ	93.86
MSLDE	94.85
ZZPLELD	**95.23**
ZZPLELD (weighted)	**96.93**

conditions. In each turn, one image per subject is chosen as the gallery, and the others are tested as query images. All total 21 rank-1 recognition rates for 21 turns. Figure 5 shows few sample face images and their corresponding *ZZPLELD* from CMU-PIE database at different illuminations.

Average rank-1 recognition is shown in Table 2 on CMU-PIE database of 68 subjects. Proposed method is better than other methods in average recognition.

3.2 Rank-1 Recognition Results on Extended Yale B Database

This database contains total 2432 images of 38 subjects under 64 different illumination conditions and in a cropped form with a size of 192×168 pixels. Again, the database is divided into five different subsets according to the illumination angle: Subset 1 ($0°$ to $12°$, 7 images per subject), Subset 2 ($13°$ to $25°$, 12 images per subject), Subset 3 ($26°$ to $50°$, 12 images per subject), Subset 4 ($51°$ to $77°$, 14 images per subject), and Subset 5 ($78°$ and above, 19 images per subject). Figure 6 shows one sample face images under different illumination conditions from the Extended Yale B database and their corresponding *ZZPLELD* images. For the experiment, the image with the most neutral light condition without illumination for each subject from Subset 1 were defined as the gallery, and the remaining images from Subset 1 to Subset 5 were used as query images. A comparison on of the rank 1 accuracy achieved on this database of 38 subjects
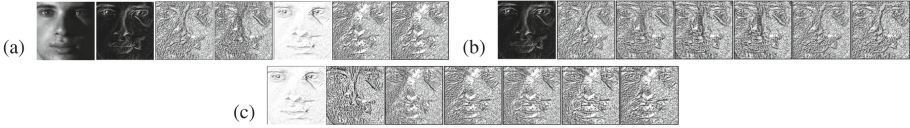
Fig. 6. (a) Sample *ZZPLELD* images after applying 3×3 *LZZBP* on both extremum *LELD* from Extended Yale B Face database. (b) Sample *ZZPLELD* images after applying 5×5 *LZZBP* on maximum *LELD* from Extended Yale B Face database. (c) Sample *ZZPLELD* images after applying 5×5 *LZZBP* on minimum *LELD* from Extended Yale B Face database.

Table 3. Rank-1 recognition rates for different methods on different subsets of Extended Yale B database of 38 subjects.

State-of-the-art methods	Rank-1 accuracy (%)					
	S1	S2	S3	S4	S5	Avg
LBP	84.81	89.91	80.04	72.18	74.99	80.39
Gradient-face	94.74	**100**	83.33	75.94	84.65	87.73
TVQI	92.28	95.20	90.27	81.39	84.32	88.69
HF+HQ	94.81	98.76	93.18	82.90	84.43	90.82
MSLDE	96.61	**100**	93.20	86.66	89.33	93.16
ZZPLELD	**97.01**	**100**	**94.56**	**90.62**	**90.37**	**94.51**
ZZPLELD (weighted)	**99.01**	**100**	**96.06**	**91.93**	**92.07**	**95.81**

on the individual subset and after averaged over all subsets is shown in Table 3.

3.3   Rank-1 Recognition Results on Viewed Sketch Databases

A CUHK Face Sketch FERET (CUFSF) database has been used for the experimental study, which includes 1194 different subjects from the FERET database. For each person, there is a sketch with shape exaggeration drew by an artist when viewing this photo and a face photo with lighting variations. All the frontal faces in the database are cropped manually by setting approximately the same eye levels and resized to 120×120 pixels. The proposed method is tested with the existing state-of-the-art methods (PLS, CITE, MCCA, TFSPS, KP-RS, MvDA, G-HFR, and LGFP). Experimental setups and results of other state-of-the-art methods are collected from the published papers. The rank-1 recognition result of proposed *ZZPLELD* on CUFSF database is 96.35% at rank-1, and it is shown in the Table 4. Proposed method outperforms other state-of-the-art methods. Figure 7 shows one sample face photo and sketch image from the CUFSF database and their corresponding *ZZPLELD* images.

Table 4. Rank-1 recognition rates for different methods on CUFSF database.

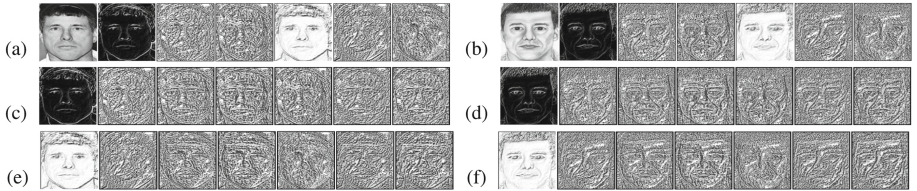| State-of-the-art methods	Number of training samples (subjects)	Number of testing samples (subjects)	Rank-1 accuracy (%)
PLS	300	894	51.00
CITE	500	694	89.54
MCCA	300	894	92.17
TFSPS	300	894	72.62
KP-RS	500	694	83.95
MvDA	500	694	55.50
G-HFR	500	694	96.04
LGFP	0	1194	94.47
ZZPLELD	0	1194	**96.35**
ZZPLELD (weighted)	500	694	**98.05**

Fig. 7. Sample Sketch and photo images and their corresponding *ZZPLELD* images from CUFSF database. (a) Photo-*ZZPLELD* images after applying 3×3 *LZZBP* on both extremum *LELD*, (b) Sketch-*ZZPLELD* images after applying 3×3 *LZZBP* on both extremum *LELD*, (c) Photo-*ZZPLELD* images after applying 5×5 *LZZBP* on maximum *LELD*, (d) Sketch-*ZZPLELD* images after applying 5×5 *LZZBP* on maximum *LELD*, (e) Photo-*ZZPLELD* images after applying 5×5 *LZZBP* on minimum *LELD*, (f) Sketch-*ZZPLELD* images after applying 5×5 *LZZBP* on minimum *LELD*.

3.4 Rank-1 Recognition Results on NIR-VIS CASIA-HFB Database

This database has 200 subjects with probe images captured in the near-infrared and gallery images captured in the visible light. Each and every subject has 4 NIR images and 4 VIS images with pose and expression variations. All the frontal faces in the database are cropped manually by setting approximately the same eye levels and resized to 120×120 pixels. This database follows standard evaluation protocols.

The output result is also tested against other state-of-the-art methods (KP-RS, LCKS-CSR, THFM). Table 5 shows the rank-1 accuracy of the proposed method and other state-of-the-art methods. The rank-1 recognition of all those methods, mentioned above, is found from different published papers. The rank-1 recognition accuracy of the proposed method is 99.39%, and it is

Table 5. Rank-1 recognition rates for different methods on CASIA-HFB database of 200 subjects.

State-of-the-art methods	Number of training samples (subjects)	Number of testing samples (subjects)	Rank-1 accuracy (%)
KP-RS	133	67	87.80
LCKS-CSR	150	150	81.43
THFM	100	100	99.28
ZZPLELD	0	200	**96.01**
ZZPLELD (weighted)	100	100	**99.69**
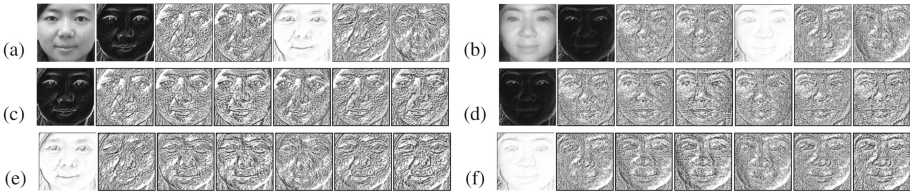
Fig. 8. Sample VIS and NIR images and their corresponding *ZZPLELD* images from CASIA-HFB database, (a) VIS-*ZZPLELD* images after applying 3×3 *LZZBP* on both extremum *LELD*, (b) NIR-*ZZPLELD* images after applying 3×3 *LZZBP* on both extremum *LELD*, (c) VIS-*ZZPLELD* images after applying 5×5 *LZZBP* on maximum *LELD*, (d) NIR-*ZZPLELD* images after applying 5×5 *LZZBP* on maximum *LELD*, (e) VIS-*ZZPLELD* images after applying 5×5 *LZZBP* on minimum *LELD*, (f) NIR-*ZZPLELD* images after applying 5×5 *LZZBP* on minimum *LELD*.

better than other methods. One sample NIR-VIS pair image from CASIA-HFB database and its different levels of *ZZPLELD* is shown in Fig. 8.

4   Conclusion

We have presented a novel modality-invariant feature representation *ZZPLELD* for HFR. It is a combination of *LELD* and *LZZBP* to boost the performance of HFR. The proposed *LELD* is an illumination-invariant image representation. To capture the local patterns of *LELD* a novel zigzag binary pattern (*LZZBP*) is proposed. Since the entire database images used in the experiment are in a frontal mode without rotation, we have not thought about whether *LZZBP* will work good in huge rotation variations or not.

Experimental results on illumination variations, sketch-photo and NIR-VIS databases, it shows the supremacy in rank-1 recognition than other compared methods. The result shows *ZZPLELD* has a good verification and discriminating ability in heterogeneous face recognition.

LELD can easily be used as a preprocessing stage to remove illumination variations, and to enhance edge features. Therefore, it has a long range of applications for illumination variations to heterogeneous face recognition. Since the

entire database images used in the experiment are in a frontal mode without rotation, we have not thought about whether *LZZBP* will work well on huge rotation variations or not. In future a rotation-invariant *LZZBP* can be thought of texture analysis. Again, how *ZZPLELD* will work on other facial variations like pose, expression, etc. that also need a further investigation. At the same time, a further investigation is also required to search other application domains for *ZZPLELD*.

References

1. Li, S., Chu, R., Liao, S., Zhang, L.: Illumination invariant face recognition using NIR images. IEEE Trans. Pattern Anal. Mach. Intell. **29**(4), 627–639 (2007)
2. Li, S.: Encyclopaedia of Biometrics. Springer, Boston (2009)
3. Tang, X., Wang, X.: Face sketch recognition. IEEE Trans. Circ. Syst. Video Technol. **14**(1), 50–57 (2004)
4. Chen, J., Yi, D., Yang, J., Zhao, G., Li, S., Pietikainen, M.: Learning mappings for face synthesis from near infrared to visual light images. In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 156–163 (2009)
5. Gao, X., Zhong, J., Li, J., Tian, C.: Face sketch synthesis algorithm on e-hmm and selective ensemble. IEEE Trans. Circ. Syst. Video Technol. **18**(4), 487–496 (2008)
6. Wang, X., Tang, X.: Face photo-sketch synthesis and recognition. IEEE Trans. Pattern Anal. Mach. Intell. **31**(1), 1955–1967 (2009)
7. Li, J., Hao, P., Zhang, C., Dou, M.: Hallucinating faces from thermal infrared images. In: Proceedings of the IEEE International Conference on Image Processing, pp. 465–468 (2008)
8. Gao, X., Wang, N., Tao, D., Li, X.: Face sketchphoto synthesis and retrieval using sparse representation. IEEE Trans. Circ. Syst. Video Technol. **22**(8), 1213–1226 (2012)
9. Wang, N., Li, J., Tao, D., Li, X., Gao, X.: Heterogeneous image transformation. Elsevier J. Pattern Recogn. Lett. **34**, 77–84 (2013)
10. Wang, N., Tao, D., Gao, X., Li, X., Li, J.: Transductive face sketch-photo synthesis. IEEE Trans. Neural Netw. **24**(9), 1364–1376 (2013)
11. Peng, C., Gao, X., Wang, N., Tao, D., Li, X., Li, J.: Multiple representation-based face sketch-photo synthesis. IEEE Trans. Neural Netw. **27**(11), 1–13 (2016)
12. Lin, D., Tang, X.: Inter-modality face recognition. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3954, pp. 13–26. Springer, Heidelberg (2006). https://doi.org/10.1007/11744085_2
13. Yi, D., Liu, R., Chu, R., Lei, Z., Li, S.: Face matching between near infrared and visible light images. In: Proceedings of International Conference on Biometrics, pp. 523–530 (2007)
14. Lei, Z., Li, S.: Coupled spectral regression for matching heterogeneous faces. In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1123–1128 (2009)
15. Sharma, A., Jacobs, D.: Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch. In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 593–600 (2011)

16. Mignon, A., Jurie, F.: CMML: a new metric learning approach for cross modal matching. In: Proceedings of Asian Conference on Computer Vision, pp. 1–14 (2012)
17. Lei, Z., Liao, S., Jain, A.K., Li, S.Z.: Coupled discriminant analysis for heterogeneous face recognition. IEEE Trans. Inf. Forensics Secur. **7**(6), 1707–1716 (2012)
18. Kan, M., Shan, S., Zhang, H., Lao, S., Chen, X.: Multi-view discriminant analysis. IEEE Trans. Pattern Anal. Mach. Intell. **38**(1), 188–194 (2016)
19. Liao, S., Yi, D., Lei, Z., Qin, R., Li, S.: Heterogeneous face recognition from local structure of normalized appeaaranceshared representation learning for heterogeneous face recognition. In: Proceedings of IAPR International Conference on Biometrics (2009)
20. Klare, B.F., Li, Z., Jain, A.K.: Matching forensic sketches to mug shot photos. IEEE Trans. Pattern Anal. Mach. Intell. **33**(3), 639–646 (2011)
21. Zhang, W., Wang, X., Tang, X.: Coupled information-theoretic encoding for face photo-sketch recognition. In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 513–520 (2011)
22. Bhatt, H.S., Bharadwaj, S., Singh, R., Vatsa, M.: Memetically optimized MCWLD for matching sketches with digital face images. IEEE Trans. Inf. Forensics Secur. **7**(5), 1522–1535 (2012)
23. Klare, B.F., Jain, A.K.: Heterogeneous face recognition using kernel prototype similarities. IEEE Trans. Pattern Anal. Mach. Intell. **35**(6), 1410–1422 (2013)
24. Zhu, J., Zheng, W., Lai, J., Li, S.: Matching NIR face to VIS face using transduction. IEEE Trans. Inf. Forensics Secur. **9**(3), 501–514 (2014)
25. Gong, D., Li, Z., Liu, J., Qiao, Y.: Multi-feature canonical correlation analysis for face photo-sketch image retrieval. In: Proceedings of ACM International Conference on Multimedia, pp. 617–620 (2013)
26. Roy, H., Bhattacharjee, D.: Heterogeneous face matching using geometric edge-texture feature (getf) and multiple fuzzy-classifier system. Elsevier J. Appl. Soft Comput. **46**, 967–979 (2016)
27. Roy, H., Bhattacharjee, D.: Local-gravity-face (LG-face) for illumination-invariant and heterogeneous face recognition. IEEE Trans. Inf. Forensics Secur. **11**(7), 1412–1424 (2016)
28. Roy, H., Bhattacharjee, D.: Face sketch-photo matching using the local gradient fuzzy pattern. IEEE J. Intell. Syst. **31**(3), 30–39 (2016)
29. Roy, H., Bhattacharjee, D.: Face sketch-photo recognition using local gradient checksum: LGCS. Springer Int. J. Mach. Learn. Cybern. **8**(5), 1457–1469 (2017)
30. Roy, H., Bhattacharjee, D.: A novel quaternary pattern of local maximum quotient for heterogeneous face recognition. Elsevier Pattern Recogn. Lett. (2017). https://doi.org/10.1016/j.patrec.2017.09.029
31. Peng, C., Gao, X., Wang, N., Li, J.: Graphical representation for heterogeneous face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **39**(2), 1–13 (2016)
32. Sinha, P., Balas, B., Ostrovsky, Y., Russell, R.: Face recognition by humans: Nineteen results all computer vision researchers should know about. Proc. IEEE **94**, 1948–1962 (2006)
33. Lai, Z., Dai, D., Ren, C., Huang, K.: Multiscale logarithm difference edgemaps for face recognition against varying lighting conditions. IEEE Trans. Image Process. **24**(6), 1735–1747 (2015)
34. Ojala, T., Pietikinen, M., Menp, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)

35. Land, E.H., McCann, J.J.: Lightness and retinex theory. J. Opt. Soc. Am. **61**(1), 1–11 (1971)
36. Horn, B.K.P.: Robot Vision. MIT Press, Cambridge (2011)
37. Zhang, T., Tang, Y.Y., Fang, B., Shang, Z., Liu, X.: Face recognition under varying illumination using gradientfaces. IEEE Trans. Image Process. **18**(11), 2599–2606 (2009)
38. An, G., Wu, J., Ruan, Q.: An illumination normalization model for face recognition under varied lighting conditions. Elsevier J. Pattern Recogn. Lett. **31**, 1056–1067 (2010)
39. Belhumeur, P., Georghiades, A., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Learn. **23**(6), 643–660 (2001)
40. Lee, K.C., Ho, J., Kriegman, D.: Acquiring linear subspaces for face recognition under variable lighting. IEEE Trans. Pattern Anal. Mach. Learn. **27**(5), 684–698 (2005)
41. Li, S.Z., Lei, Z., Ao, M.: The HFB face database for heterogeneous face biometrics research. In: Proceedings of IEEE International Workshop on Object Tracking and Classification Beyond and in the Visible Spectrum, Miami (2009)
42. Fan, C.N., Zhang, F.Y.: Homomorphic filtering based illumination normalization method for face recognition. Elsevier J. Pattern Recogn. Lett. **32**, 1468–1479 (2011)
43. Turk, M., Pentland, A.: Eigenfaces for recognition. J. Cogn. Neurosci. **3**(1), 71–86 (1991)
44. Lawrence, S., Giles, C.L., Tsoi, A., Back, A.: Face recognition: a convolution neural-network approach. IEEE Trans. Neural Netw. **8**(1), 98–113 (1997)
45. Wiskott, L., Fellous, J.M., Kruger, N., Malsburg, C.V.: Face recognition by elastic bunch graph matching. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 775–779 (1997)
46. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression database. IEEE Trans. Pattern Anal. Mach. Intell. **25**(12), 1615–1618 (2003)