Illumination invariant single face image recognition under heterogeneous lighting condition

Jun-Yong Zhu[a,b,d], Wei-Shi Zheng[a,e,*], Feng Lu[b], Jian-Huang Lai[a,c]

[a] *School of Data and Computer Science, Sun Yat-sen University, PR China*
[b] *Institute of Criminal Science and Technology, Guangzhou, Guangdong, PR China*
[c] *Guangdong Key Laboratory of Information Security Technology, Sun Yat-sen University, PR China*
[d] *Collaborative Innovation Center of High Performance Computing, National University of Defense Technology, Changsha 410073, China*
[e] *Key Laboratory of Machine Intelligence and Advanced Computing (Sun Yat-sen University), Ministry of Education, China*

A B S T R A C T

Illumination problem is still a bottleneck of robust face recognition system, which demands extracting illumination invariant features. In this field, existing works only consider the variations caused by lighting direction or magnitude (denoted as *homogeneous lighting*), but the effect of spectral wavelength is always ignored and thus existing illumination invariant descriptors have its limitation on processing face images under different spectral wavelengths (denoted as *heterogeneous lighting*). We propose a novel gradient based descriptor, namely Logarithm Gradient Histogram (LGH), which takes the illumination direction, magnitude and the spectral wavelength together into consideration, so that it can handle both homogeneous and heterogeneous lightings. Our proposal contributes in three-folds: (1) we incorporate LMSN-LoG filter to eliminate the lighting effect of each image and extract two illumination invariant components, namely logarithm gradient orientation (LGO) and logarithm gradient magnitude (LGM); (2) we propose an effective post-processing strategy to make our model tolerant to noise and generate a histogram representation to integrate both LGO and LGM; (3) we present solid theoretical analysis on the illumination invariant properties of our proposed descriptors. Extensive experimental results on CMU-PIE, Extended YaleB, FRGC and HFB databases are reported to verify the effectiveness of our proposed model.

1. Introduction

Face recognition has attracted significant attention in the last decades owing to its wide range of applications, including information security, law enforcement, video surveillance, cooperative user applications, etc. Many research efforts have focused on the face recognition problem under relatively well-controlled conditions using sufficient training data [1–4]. However, in practical application, the performance of face recognition system is greatly affected by the illumination condition. Face recognition under wide range of lighting variations is still an open issue, especially when there is only one sample available for each person, which is a common scenario in many security system.

This work focuses on single-image-based face recognition model for handling illumination variations. As indicated by early research [5–7], the accuracy of a recognition system heavily depends on the number of training samples for each person. However, in many large-scale identification applications, such as law enforcement, driver license or passport card identification, there is usually only one sample per

person stored in the database. The main challenge is how to guarantee robust performance under this extremely small sample size condition for each person. Solutions to this problem can be roughly divided into two aspects [8], i.e., the holistic methods [9–11] and the local methods [12–15]. The former one extracted globally stable facial features and the latter one described each face image as a batch of local features which are assumed to be invariant to lighting variation.

The illumination problem, as one of the main challenges for existing face recognition systems, has become a barrier in many face related applications. The well-known face recognition vendor test (FRVT) 2006 [2] has revealed that large variation in illumination would probably affect the performance of face recognition algorithms. A variety of works have been proposed to address this issue and they mainly fall into three categories [16]: preprocessing and normalization techniques [17,18], illumination modeling based approaches [19,20] and invariant feature extraction [14,21–25].

More specifically, preprocessing and normalization methods like histogram equalization (HE) [17], gamma correction [26] and homo-

morphic filtering [26] attempt to take holistic normalization on face images such that the restored images appear to be consistent with those under normal illumination. However, most of these methods are ad hoc and hard to obtain satisfactory results when suffering uneven lighting conditions, although their visual effect appears acceptably in some cases. To further investigate the cause of illumination problem, the modeling based approaches turn to exploring the mechanism of face imaging. Based on the assumption that the surface of face is Lambertian, images of the same face under varying lighting conditions can be approximated by a low dimensional linear subspace [19,27]. The relationships between 2D illuminated images and corresponding 3D facial surface are investigated in some advanced work, e.g., morphing face [28] and shape from shading (SFS) [29]. In theory, these methods model the illumination variation quite well, but they need a great deal of training samples to learn the variation and easily suffer from the over-fitting problem, which largely restricts their use in real applications, especially when only single image per person is available.

Compared to the above two categories, most invariant feature based methods are more effective and do not demand learning. Classical methods such as Local Binary Pattern (LBP) [14], Gabor [30] and their variations [30–32] are widely conceived to be robust to slight illumination change, but their performance is no longer guaranteed when the lighting condition becomes severe. To overcome this problem, a variety of state-of-the-art methods have been proposed by extracting the reflectance component [21,22,16] or alleviating the illumination component based on the Lambert's reflectance model [25,24], and great success using these effective methods has been reported.

The illumination problem mentioned above is mainly due to the variations caused by either varying lighting direction or varying lighting magnitude. In those methods, a main assumption they make is that the wavelengths of light are the same, and this case is called the *homogeneous lighting* in this paper. However, it is not always the case in realistic applications. For example, the lighting wavelengths of indoor and outdoor conditions are always different, so is the case of visible (VIS) and near infrared (NIR) spectral face images [33]. As a result, the reflectance component which is determined by the albedo and the normal direction of facial surface will change, since the albedo is related to the spectral wavelength. Some related works like [33,34] consider the VIS–NIR face matching as a multi-modality face recognition problem rather than an illumination related task. For convenience, we denote the lighting condition with different spectral wavelengths as *heterogeneous lighting*. As far as we know, there is still lack of work on solving the heterogeneous lighting problem, and theoretical development on a valid image descriptor in this aspect is also limited.

In this paper, we formulate a new illumination invariant feature extraction model for both homogeneous and heterogeneous lightings. In particular, we propose a novel gradient based descriptor, namely *logarithm gradient histogram* (LGH), to eliminate the illumination variations. We have provided an in-depth analysis on its illumination invariant property. In addition, we also introduce a multi-scale bank-pass filtering as a preprocessing to constrain the illumination effect and enhance facial information. Experiments on several public face databases have been conducted to verify the effectiveness of our proposed method.

The rest of this paper is organized as follows: Section 2 further discusses the invariant feature extraction approaches. Section 3 introduces our proposed descriptor in detail. After that, the theoretical proof of illumination invariant properties of our proposed method is presented in Section 4. Experiments on CMU-PIE, Extended YaleB, FRGC v2.0 and HFB databases and further analysis will be carried out in Section 5 to evaluate our proposed LGH. Finally, the conclusion of this paper is drawn in Section 6.

2. Related work

In this section, we mainly review related work on illumination invariant descriptors for face recognition. In addition, some work about face recognition using single training image under illumination will be discussed at the end of this section.

According to the Lambertian Law, the intensity of the illuminated image I can be formulated as $I(x, y) = F(x, y)L(x, y)$, which is a product of the illumination component $L(x, y)$ and the reflectance component $F(x, y)$. As commonly assumed, the $L(x, y)$ in the Lambert's reflectance model changes very slowly, and $F(x, y)$ is independent of lighting condition [35,36,24]. Thus, $F(x, y)$ is commonly regarded as illumination invariant feature.

In order to extract the illumination insensitive component only related to $F(x, y)$, self-quotient image (SQI) [21] alleviates the effect of illumination by dividing itself with its blurred version. Logarithmic Total Variation (LTV) [22] incorporates TV model to preserve the edge information and obtain a more elaborate representation. Following a similar way, Xie et al. [16] computed a better reflectance component by using logarithmic non-subsampled contourlet transform and obtained significant improvement on the cost of time consuming. There are also some other work on investigating the illumination invariant property by considering the normalized local intensity contrast, such as relative gradient [37] and Weber face (WF) [25]. Inspired by Weber's Law [38], the authors in [25] show that the ratio between local difference and the center degree is insensitive to the illumination change and encouraging results were reported.

Apart from empirically predefined methods, Cao et al. designed a learning-based descriptor to handle face recognition in the wild, which is also suitable for addressing illumination problem [39]. Taking a further step, Lei et al. proposed a method named discriminant face descriptor (DFD) [40] to learn a discriminant descriptor in a data-driven way. This work is effective in handling both homogeneous and heterogeneous lighting conditions. In addition, deep learning has shown its great ability in solving wild face recognition involving lighting, pose, occlusion and some other variations [4,3,41]. However, most of these methods are based on the learning mechanism that demands multiple training samples for each person. To tackle this problem, one-shot technique could be applied. For example, a typical solution is the one-shot similarity kernel (OSSK) approach to incorporate label information from few training samples. OSSK computes similarity score that describes the likelihood of each vector belonging to the same class as the other vectors but not in a class defined by a fixed set of negative examples, which can naturally use unlabeled data to help define the class boundary [42]. However, for face recognition, it is more effective to learn a descriptor even when single image is available for each person. Recently, Lu et al. proposed an effective binary face descriptor, named compact binary face descriptor (CBFD), by mapping the pixel difference vectors (PDVs) in local patches into low-dimensional binary vectors, which achieved advanced results on multiple face recognition tasks [43]. Moreover, the CBFD method is applicable even in the single face image recognition problem.

Albeit effective, most of the aforementioned methods do not provide theoretical proofs for the illumination invariant properties. One exception is that, Zhang et al. [24] proved that the gradient orientation is somehow insensitive to the illumination influence both in theoretical analysis and experimental validation. Though great success has been achieved in [24], there are still some limitations. First, the gradient orientations are compared in a pixel-wise manner, making it sensitive to noise and misalignment. Meanwhile, some important information in gradient domain like the gradient magnitude is ignored. Also, it does not discuss the variation caused by the wavelength of light. Compared to [24], all these issues will be addressed in our proposed method.

One common assumption is that the face images captured under various lighting conditions are homogeneous, i.e. having the same

spectral wavelength; it means that the reflectance component $F(x, y)$ is independent of lighting variance $L(x, y)$. However, the reflectance component indeed will be affected by lighting wavelength in reality. Hence the extracted reflectance components of the same subject under heterogeneous lighting such as sunlight, electric lamp, near-infrared camera, etc., are different from each other. To tackle this problem, [33] suggests encoding the local pattern via LBP followed by Difference-of-Gaussian (DoG) filtering, and [34] applies Scale-invariant feature transform (SIFT) [44] descriptor instead. Liu et al. [45] applied multi-scale DoG filtering and three local descriptors to generate over-complete feature pool, and then Gentle AdaBoost is used to select the discriminant features. However, these methods are empirically designed and do not provide theoretical analysis on the illumination invariant property. In addition, as processed in other learning-based heterogeneous face recognition methods [46,47], the final feature representations of the above methods are learned based on multiple labeled data, which cannot be applicable for single face image recognition.

Apart from the local image descriptors, there are also related works addressing this problem in other aspects. For example, Bischof et al. [48] incorporated a gradient based filter bank into the eigenspace and they found that the eigenimage coefficients are invariant to linear filtering. To achieve further illumination insensitivity, they designed a robust procedure for coefficient recovery. Although their method can be applied when one sample per person is available, the recovered face images are easily blurred by the eigenspace trained by all people, so that the details containing discriminant information for each person may loss in recognition. Based on 3D Morphable Model [28] and the spherical harmonic illumination representation [27], Zhang and Samaras [49] proposed two methods to estimate the spherical harmonic basis images from just one image taken under arbitrary illumination conditions, and each face image can be represented as a weighted combination of basis images. Recently, Lu et al. [50] divided a single image into small patches and formulated the single sample per person problem as a manifold–manifold matching problem, and the performance turned out well against illumination changes.

In this paper, we study a more general case for the illumination problem in face recognition, where the lighting direction, magnitude, and the spectral wavelength will change in different lighting conditions. We propose a new and efficient illumination invariant descriptor based on the gradient information in the logarithm domain, which can work across heterogeneous lighting conditions. Parts of this work have been first reported in our preliminary conference version [51]. In this work, we have improved the algorithm so as to gain much better performance, and we also offer more detailed solid theoretical and experimental analysis on the illumination invariance of the proposed descriptor.

3. The approach

In this section, we aim to extract gradient based illumination insensitive components. Specifically, we will elaborate our proposed LGH in three folds: (i) the multi-scale bank-pass filtering used for constraining the illumination effect and enhancing facial information; (ii) two illumination invariant components, i.e., logarithm gradient orientation (LGO) and logarithm gradient magnitude (LGM), and the theoretical analysis on the illumination invariant property; (iii) post-processing for integrating LGO and enhanced LGM into a histogram based feature representation. The whole procedure is illustrated in Fig. 1.

3.1. Scale space representation

Multi-scale analysis [52,30,16] can capture both characteristics in the spatial space and scale space. According to scale-space theory [52], given any image $f: \mathbf{R}^2 \to \mathbf{R}$, its scale-space representation
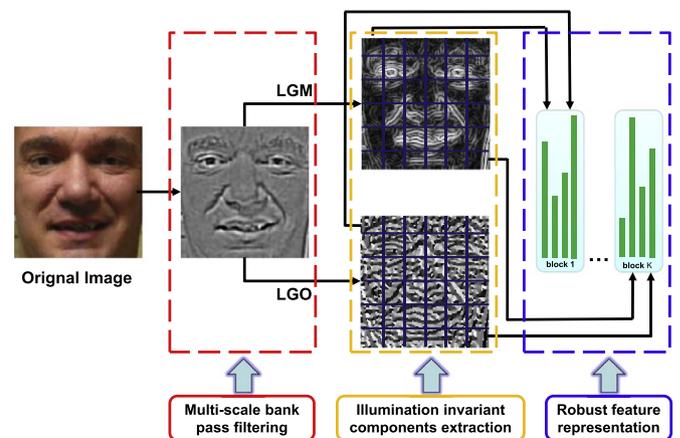
Fig. 1. Flowchart of the whole process, including multi-scale bank pass filtering, illumination invariant component extraction and robust feature representation.

$M: \mathbf{R}^2 \times \mathbf{R}_+ \to \mathbf{R}$ is defined by

$$M(\cdot;t) = g(\cdot;t)*f(\cdot),\tag{1}$$

where $g: \mathbf{R}^2 \times \mathbf{R}_+ \to \mathbf{R}$ denotes the Gaussian kernel given by

$$g(x;t) = \frac{1}{2\pi t}e^{-(\|x\|^2)/(2t)},\tag{2}$$

and the variance t is the scale parameter. Hence, the scale-space derivatives are defined as

$$M_{x^\alpha}(\cdot;t) = \partial_{x^\alpha}M(\cdot;t) = g_{x^\alpha}(\cdot;t)*f,\tag{3}$$

where α denotes the order of differentiation. The induced operators are widely used in a large number of visual operations. Interestingly, neurophysiological studies have found that some profiles in the mammalian retina and visual cortex can be well modeled by superpositions of Gaussian derivatives [53].

Although the scale-space theory provides a well-founded framework for representing image structures (e.g. edges, blobs) at multiple scales, the amplitude of spatial derivatives will decrease as scale increases according to Eq. (3). To overcome this problem, Lindeberg [53] suggested using γ-normalized derivatives $\partial_\xi = t^{\gamma/2}\partial_x$ to obtain the normalized measure of feature strength among different scales. A commonly used normalized derivative is $t\nabla^2 L$, which is sensitive to the structure like edges and blobs [52,53].

3.2. Logarithmic multi-scale normalized Laplacian-of-Gaussian filtering

According to the Lambertian reflectance function, the intensity of a 2D surface I can be described as

$$I(x, y) = R(x, y)L(x, y),\tag{4}$$

where I, L, R represent the intensity, illumination component and reflectance components, respectively. It is assumed that L changes very slowly [35,36,24], so the illumination effect mainly lies in the low-frequency domain. Thus, it is possible to alleviate the illumination effect by taking a bank-pass filtering in the frequency domain. However, as we can see in Eq. (4), it is not feasible to apply filtering on L since the reflectance component R and illumination component L are combined in the multiplicative form. To overcome this problem, taking the logarithm transformation at the first step is able to make the two components combined additively, so that the filtering on I can be separated into the sum of the filtering on L and that on R in the logarithm domain as follows

$$D(x, y)\widetilde{I}(x, y) = D(x, y)\widetilde{R}(x, y) + D(x, y)\widetilde{L}(x, y),\tag{5}$$

where $\widetilde{I}(x, y)$, $\widetilde{R}(x, y)$ and $\widetilde{L}(x, y)$ represent $\ln(I(x, y))$, $\ln(R(x, y))$ and

(a) CMU-PIE

(b) Extended YaleB

(c) HFB VIS

(d) HFB NIR
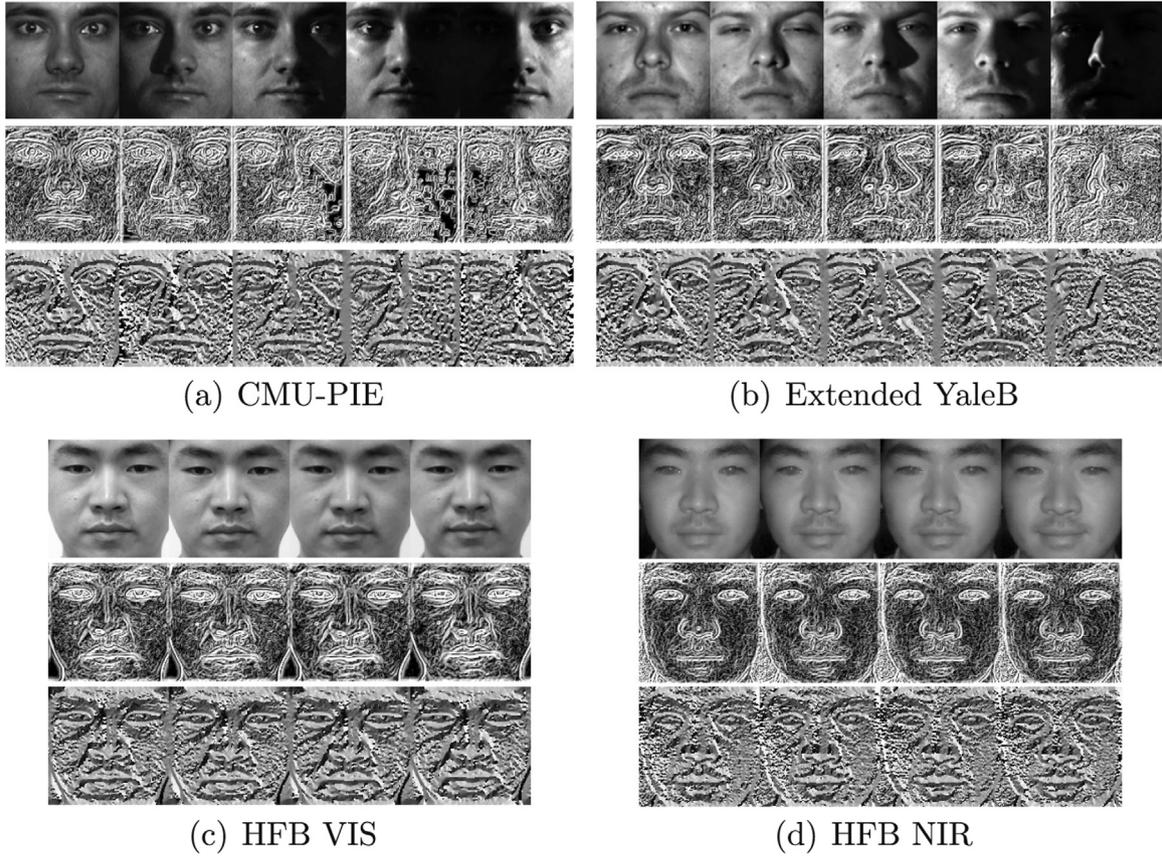
Fig. 2. Illustration of the logarithm gradient components. From top to bottom: original face images, the corresponding logarithm gradient magnitudes (LGM) and logarithm gradient orientations (LGO).

$\ln(L(x, y))$, respectively, and $D(x, y)$ denotes the filtering function.

Always, the very high-frequency components of image I are mostly conceived to be noise, and the illumination effect is assumed to mainly lie in the low-frequency domain [26]. Therefore, bank-pass filtering can be applied here to suppress low frequency illumination effect and high frequency noise, and meanwhile enhance the mid-frequency facial characteristics.

As a kind of approximate second derivative measurement, Laplacian filter is able to capture the rapid intensity change. Since the second derivative is also sensitive to noise, a Gaussian smoothing will be always taken before applying the Laplacian filtering. Such a processing is well-known as the Laplacian-of-Gaussian (LoG) filter, which is indeed a kind of bank-pass filtering used to reduce high frequency noise components. The 2D LoG function with zero center and standard Gaussian deviation has the following form:

$$LoG(x, y) = -\frac{1}{\pi\sigma^4}\left[1 - \frac{x^2 + y^2}{2\sigma^2}\right]e^{-\frac{x^2+y^2}{2\sigma^2}}, \tag{6}$$

However, taking the LoG filtering on multi-scale image representation is somehow time consuming when the number of scale becomes larger. A more efficient choice is calculating the differences of Gaussian smoothed images, which is known as Difference-of-Gaussian (DoG), an approximation of the scale-normalized LoG filtering. The relationship between DoG and LoG is based on the (heat) diffusion equation:

$$\frac{\partial L}{\partial \sigma} = \sigma\nabla^2 L, \tag{7}$$

where σ is the scale parameter having $\sigma = \sqrt{t}$. Hence, we can get the approximation as follows:

$$\sigma\nabla^2 L = \lim_{c \to 1} \frac{L(x, y, c\sigma) - L(x, y, \sigma)}{c\sigma - \sigma} \approx \frac{L(x, y, c\sigma) - L(x, y, \sigma)}{c\sigma - \sigma}. \tag{8}$$

Finally, rearranging Eq. (8) yields

$$(c - 1)\sigma^2\nabla^2 L \approx L(x, y, k\sigma) - L(x, y, \sigma). \tag{9}$$

That is,

$$(c - 1)\sigma^2 LoG = DoG. \tag{10}$$

As shown, DoG approximates the scale normalized LoG up to a (negligible) multiplicative constant $c - 1$. In another word, we can efficiently generate the LoG results via suitable multi-scale representation. The DoG filtering, as a kind of classical bank-pass filter, has already shown its ability in eliminating illumination variations [23,33]. We denote the multi-scale normalized LoG filtering as MSN-LoG. In the following section, we will further analyze the illumination invariant property in detail.

3.3. The logarithm gradient histogram

The shapes, contours and the relative small-scale facial objects such as eyes, nose and mouths are key features for face recognition [22]. The gradient information (e.g., magnitude and orientation) around these components contain much more valuable information than facial skin area. More importantly, as shown in Fig. 1, the gradient magnitudes of these parts are always large and fluctuant, while those of the skin area are relatively small. As a result, it is helpful to take the gradient magnitudes as important measurement of facial components. However, as mentioned in Gradient Face [24], the gradient magnitudes of face image do not satisfy the illumination invariant requirement. In this work, we find that this problem can be solved by transferring the derivation into logarithmic domain so that the multiplicative combination in the original input domain will become the additive form. Thus, different from Gradient Face, which only retains gradient orientations as the illumination invariant features, we incorporate both gradient

magnitude and orientation in the logarithmic domain to form our illumination insensitive features.

Until now, we have obtained the local filtered face images where edges and contours are enhanced and meanwhile the partial lighting effect is alleviated. We further claim that the gradient orientation and gradient magnitude in the logarithmic domain are insensitive to both homogeneous and heterogeneous illumination variations. Detailed inferences can be referred to Section 4. In the following part, we present extracting the *logarithmic gradient orientation* (LGO) and *logarithmic gradient magnitude* (LGM) as illumination insensitive components. Some examples are shown in Fig. 2.

To obtain robust description, postprocessing is taken on the two components before generating the histogram representation. Since the gradient orientation is somehow sensitive to the quality of image, a local smoothing operation is implemented on LGO in order to alleviate the impulse responses caused by discrete noises and make the gradient direction changes smoothly. For such a postprocessing, taking inter-polation or using a suitable σ for Gaussian kernel as similarly done in [24] is applicable.

After that, we quantify the values in LGO into several bins to achieve fault-toleration. Since the Lambertian assumption does not strictly hold everywhere, there are cast shadows in face images. As a result, the pixels with dominating values in LGM may belong to the boundaries of shadows, and meanwhile the edges of facial objects (e.g., eyes, mouths) may become less significant especially when the lighting condition becomes severe. In this case, taking a local normalization operation may help extracting illumination invariant features. On one hand, it enhances the weak edges lying in dark areas which belongs to the facial objects, and on the other hand it also alleviates the global domination caused by the strong edges of shadows. More importantly, the multiplicative constant coefficient shown in Eq. (10) can be eliminated in the local normalized LGM. As a result, the sub-histogram generated in each scale can be treated equally.

Finally, we obtain the quantified gradient orientation and normal-ized gradient magnitude for each pixel in a face image. Since conduct-ing pixel-wise matching on these two components between different face images is quite unreliable, we integrate LGO and LGM to form a unified histogram based feature representation. The histogram is generated in a block-wise form; that is the post-processed gradient magnitudes of all pixels in the block will accumulate according to the orientation bins they belong to. At last, we concatenate histograms of all blocks into a long vector to form our histogram based feature representation. The whole procedure is illustrated in Algorithm 1.

Algorithm 1. Generation of logarithm gradient histogram.

Input: LGO, LGM, k (the number of bins in histogram for en-coding each block)
Output: H (histogram of a face image)
1: Quantify LGO in k bins $\{b_1, \ldots, b_k\}$ to increase the fault-tolerant capability;
2: Divide LGM and LGO into small blocks evenly, and denote M^i and O^i as the ith block in LGM and LGO separately;
3: For each M^i, calculate the weighted normalized gradient
    magnitude by $\widetilde{M}^i(p, q) = \frac{M^i(p,q)W(p,q)}{\sum_{(p',q') \in N(p,q)} M^i(p',q')W(p',q')}$, or else
$\widetilde{M}^i(p, q) = M^i(p, q)$, where W denotes the Gaussian weight ma-trix.
4: Generate the histogram for each block,
$H^i(t) = sum\{\widetilde{M}^i(j)|O^i(j) == b_t\}$, for $t = 1, \ldots, k$, and then con-catenate them into a long vector $H = [H^i]$ to represent a face image.

---

4. Illumination invariant analysis

In the former section we claim that LGO and LGM are two illumination insensitive components. In this section, we will prove the illumination invariant property of our proposed features based on the following theorems.

4.1. Homogeneous lighting

Lemma 1. *Assume that $I(x, y)$ is an illuminated image of Lambertian face, which is captured under the homogeneous lighting (i.e., with arbitrary lighting direction and magnitude but with the same light wavelength). Let $\tilde{I}(x, y)$ denote the logarithmic image of $I(x, y)$, i.e., $\tilde{I}(x, y) = \ln(I(x, y))$, then the partial derivation $\partial_x \tilde{I}(x, y)$ and $\partial_y \tilde{I}(x, y)$ are insensitive to the illumination variation.*

Proof. For any function $f(x, y) > 0$, we denote $\tilde{f}(x, y) = \ln(f(x, y))$. Let us consider two neighboring points (x, y) and $(x + \Delta x, y)$. According to the Lambertian Law, we have

$$I(x, y) = R(x, y)L(x, y), \tag{11}$$

$$I(x + \Delta x, y) = R(x + \Delta x, y)L(x + \Delta x, y), \tag{12}$$

and then

$$\tilde{I}(x, y) = \tilde{R}(x, y) + \tilde{L}(x, y), \tag{13}$$

$$\tilde{I}(x + \Delta x, y) = \tilde{R}(x + \Delta x, y) + \tilde{L}(x + \Delta x, y). \tag{14}$$

Hence we have

$$\tilde{I}(x + \Delta x, y) - \tilde{I}(x, y) = (\tilde{R}(x + \Delta x, y) - \tilde{R}(x, y)) + (\tilde{L}(x + \Delta x, y) - \tilde{L}(x, y)). \tag{15}$$

It is commonly assumed that L varies very slowly while R can change abruptly [22,16]. The same assumption holds for $\tilde{L}$ and $\tilde{R}$. Therefore, as inspired by [35,36,24], it is reasonable to draw the conclusion that the difference between $\tilde{L}(x + \Delta x, y)$ and $\tilde{L}(x, y)$ can be discarded as compared to that between $\tilde{R}(x + \Delta x, y)$ and $\tilde{R}(x, y)$ when Δx is small enough. As a result, taking the limit of both sides in Eq. (15), we have the following approximation:

$$\partial_x \tilde{I}(x, y) \approx \partial_x \tilde{R}(x, y). \tag{16}$$

Also, by following the same procedure we can obtain

$$\partial_y \tilde{I}(x, y) \approx \partial_y \tilde{R}(x, y). \tag{17}$$

In a word, the partial derivation $\partial_x \tilde{I}(x, y)$ and $\partial_y \tilde{I}(x, y)$ are dominated by the reflectance component R instead of the illumination component L and thus insensitive to the illumination variation. □

Theorem 1. *Let $I(x, y)$ be an illuminated image of Lambertian face, which is captured under the homogeneous lighting (i.e., with arbitrary lighting direction and magnitude but with the same light wavelength). Let $\tilde{I}(x, y) = \ln(I(x, y))$, and then the gradient orientation and gradient magnitude of $\tilde{I}(x, y)$ are both illumination invariant components.*

Proof. We use the same denotation in Lemma 1. It is easy to calculate the gradient orientation and magnitude of $\tilde{I}(x, y)$ as

$$LGO(x, y) = \arctan(\partial_y \tilde{I}(x, y)/\partial_x \tilde{I}(x, y)), \tag{18}$$

and

$$LGM(x, y) = \sqrt{(\partial_x \tilde{I}(x, y))^2 + (\partial_y \tilde{I}(x, y))^2}. \tag{19}$$

As proved in Lemma 1, both $\partial_x \tilde{I}(x, y)$ and $\partial_y \tilde{I}(x, y)$ are insensitive to illumination change, so that it is straightforward to have the conclusion that both $LGO(x, y)$ and $LGM(x, y)$ are illumination invariant components. □

So far we have known that the gradient orientation $LGO(x, y)$ and

gradient magnitude $LGM(x, y)$ of $\tilde{I}(x, y)$ are invariant to the illumination change caused by varying lighting direction and magnitude. However, as mentioned, "the same light wavelength" assumption in homogeneous lighting can hardly hold in real world applications, since the spectral wavelengths will change in different environments. In the next section, we will further investigate the illumination invariant property of our proposed LGO and LGM when relaxing the assumption on spectral wavelength.

4.2. Heterogeneous lighting

According to the Lambertian model, the reflectance component, which is determined by the albedo (related to spectral wavelength) and facial normal direction, will turn out to be different when suffering from heterogeneous lighting.[1] Generally, it is not feasible to tackle this problem under wild assumption. For example, the facial skin reflectance under visible and far infrared light differ so much that the obtained face images are hardly the same. Nevertheless, according to the work on skin reflectance spectra simulation [54], we find that the response of skin reflectance spectra changes smoothly as the lighting wavelength increases within the visible and near-infrared spectral range (450–1100 nm). Inspired by this observation, we assume that within a small patch of facial skin, the reflectance components under two different light of different wavelengths are approximately proportional. We denote it as the *locally proportional reflectance assumption* (see Theorem 2). Therefore, the key of solving such a general illumination problem is to extract illumination invariant features under this assumption. In the following part, we prove that our proposed LGO and LGM are also invariant components in such a scenario.

Theorem 2. *Assume that* $I_1(x, y) = R_1(x, y)L_1(x, y)$ *and* $I_2(x, y) = R_2(x, y)L_2(x, y)$ *are Lambertian face images of the same subject captured under two heterogeneous light (i.e., with different lighting directions, magnitudes and even spectral wavelengths). Let* $\tilde{I}(x, y) = \ln(I(x, y))$. *If* $R_1(x, y)$ *is locally proportional to* $R_2(x, y)$, *i.e.,* $R_1(x, y) = wR_2(x, y)$ *for some constant w which is determined by* $\mathcal{N}(x, y)$ *(the neighborhood of (x,y)), then the gradient orientation and gradient magnitude of* $\tilde{I}_1(x, y)$ *and* $\tilde{I}_2(x, y)$ *are equal, i.e., they are both invariant components for the illumination problem.*

Proof. As indicated in Theorem. 1, both the gradient orientation and gradient magnitude of $\tilde{I}(x, y)$, i.e., $LGO(x, y)$ and $LGM(x, y)$ are invariant to varying illumination directions and magnitudes. Therefore, in this part we only need to prove the invariant property against the varying wavelength. By substituting Eq. (16) and (17) into Eq. (18) and (19), we have

$$LGO(x, y) = \arctan\left(\frac{\partial_y \tilde{I}(x, y)}{\partial_x \tilde{I}(x, y)}\right) = \arctan\left(\frac{\partial_y \tilde{R}(x, y)}{\partial_x \tilde{R}(x, y)}\right)$$
$$= \arctan\left(\frac{\partial_y R(x, y)/R(x, y)}{\partial_x R(x, y)/R(x, y)}\right) = \arctan\left(\frac{\partial_y R(x, y)}{\partial_x R(x, y)}\right),$$

(20)

and

$$LGM(x, y) = \sqrt{(\partial_x \tilde{I}(x, y))^2 + (\partial_y \tilde{I}(x, y))^2} = \sqrt{(\partial_x \tilde{R}(x, y))^2 + (\partial_y \tilde{R}(x, y))^2}$$
$$= \frac{\sqrt{(\partial_x R(x, y))^2 + (\partial_y R(x, y))^2}}{R(x, y)}.$$

(21)

Assume that $R_1(x, y)$ and $R_2(x, y)$ are locally proportional. That is,

$$R_1(x, y) = wR_2(x, y)$$

(22)

for some constant $w > 0$ which is determined by $\mathcal{N}(x, y)$. Thus,

$$\arctan\left(\frac{\partial_y R_1(x, y)}{\partial_x R_1(x, y)}\right) = \arctan\left(\frac{\partial_y R_2(x, y)}{\partial_x R_2(x, y)}\right),$$

(23)

i.e.,

$$LGO_1(x, y) = LGO_2(x, y).$$

(24)

Also, we have

$$LGM_1(x, y) = \frac{\sqrt{(\partial_x R_1(x, y))^2 + (\partial_y R_1(x, y))^2}}{R_1(x, y)}$$
$$= \frac{\sqrt{(k\partial_x R_2(x, y))^2 + (k\partial_y R_2(x, y))^2}}{kR_2(x, y)}$$
$$= \frac{\sqrt{(\partial_x R_2(x, y))^2 + (\partial_y R_2(x, y))^2}}{R_2(x, y)} = LGM_2(x, y).$$

(25)

Hence, we now can draw the conclusion that $LGO(x, y)$ and $LGM(x, y)$ are both insensitive to the illumination variation caused by the change of spectral wavelength under the local proportional assumption. □

5. Experiments

In this section, we conducted a series of experiments to evaluate the proposed illumination invariant descriptor for single-image-based face recognition. Three scenarios will be considered in our experiments: (1) the case when images were approximately captured under controlled homogeneous lighting, i.e., with different lighting directions and magnitudes but with the same spectral wavelength; (2) the case when images were captured in uncontrolled lighting conditions, e.g., outdoor environment; (3) the case when images were captured under heterogeneous lighting, i.e., with varying lighting direction, magnitude and wavelength. For the first and second cases, we compared our method with HE [17], SQI [21], LTV [22], Weber Face [25], Gradient Face [24] and TT-LTP [23] on CMU-PIE [55], Extended YaleB [19] and FRGC v2.0 [1] databases. For the third case, we particularly consider the heterogeneous face recognition and compared our method with DoG +LBP [33] and SIFT descriptor used in [34] on the heterogeneous face biometric (HFB) database [56]. In our experiments, we mainly evaluated the performance of different image descriptors and thus simply adopted the cosine distance as the similarity measurement and used nearest neighborhood classifier for classification.

5.1. Controlled homogeneous lighting

5.1.1. Databases

In this experiment, we evaluated the performance of various methods on CMU-PIE and Extended YaleB databases. To form the set of frontal face images, 1428 frontal face images from 68 individuals captured under 21 different illumination conditions were selected from CMU-PIE database. For Extended YaleB, face images from 38 individuals of nine poses were captured under 64 different lighting conditions, and we only used $64 \times 38 = 2432$ frontal face images here. All images were simply aligned according to eyes coordinates and resized to 128×128.

5.1.2. Compared methods

We compared our LGH with several classical and state-of-the-art methods, including HE [17], SQI [21], LTV [22], Weber Face (WF) [25], Gradient Face (GF) [24] and TT-LTP [23] on both CMU-PIE and Extended YaleB databases. All methods were implemented with the parameters as suggested in the references. For our LGH, the kernel size of DoG filtering was fixed as 7×7, the array of σ was empirically set as [1, 2, 4, 8, 16] (i.e., the number of scales was 4), and 5 bins were adopted to represent the block histogram. PCA was conducted to obtain a compact representation and at last 500 dimensions were reserved. Unless otherwise stated, the block size is set as 4×4, unsigned gradient

---

[1] Especially, the heterogeneous lighting here refers to the light with different wavelengths.
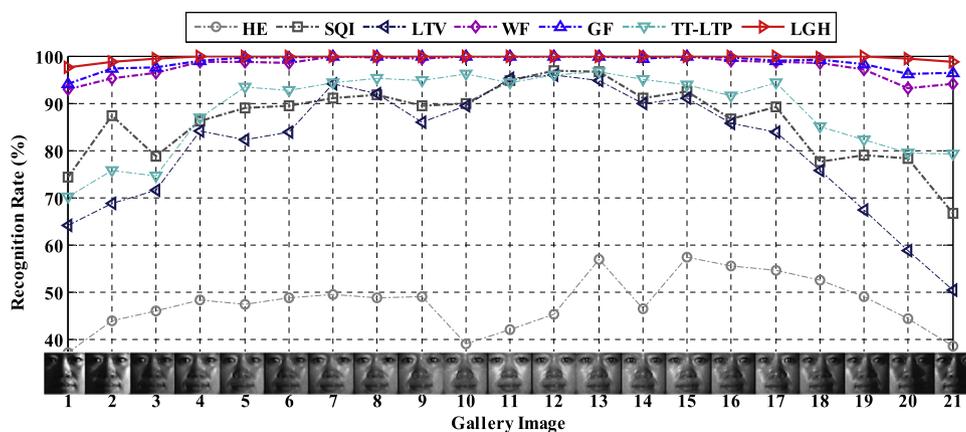
Fig. 3. Recognition rates with different galleries on CMU-PIE.

orientation was adopted and 5 bins were used for the quantization procedure.

5.1.3. Results on CMU-PIE

The subset used here consists of 21 images for each person, which were captured under 21 different illumination conditions as shown in Fig. 3. Only one image per individual was chosen as the gallery and the other formed the probe. We varied the gallery from the 1st image to the 21st one for each person in order to ensure that all illumination conditions were tested. The final results are illustrated in Fig. 3. As shown, in general, the performances of different algorithms degenerate as the lighting orientation changes from the frontal direction to much more non-frontal ones. All seven methods except HE achieved the best performance when using the frontal lighting images (No. 12) as galleries, and the corresponding results for various methods are as follows: 96.84% for SQI, 95.88% for LTV, 100% for WF, 99.93% for GF, 95.88% for TT-LTP and 100% for LGH. What is more, some approaches such as SQI, LTV and WF turned out less effectively when the illumination condition becomes severe, while both GF and LGH performed well even under the most extreme situation. Also, our proposed LGH outperformed all other methods almost in all lighting conditions and achieved the best average recognition rate at 99.66%. The average results of HE, SQI, LTV, WF, GF and TT-LTP are 47.59%, 86.47%, 81.17%, 98.06%, 98.76% and 88.62%, respectively.

5.1.4. Results on Extended YaleB

Different from CMU-PIE, the face images in Extended YaleB database were captured in more complex/challenging environments. Some examples can be found in Fig. 4(a). The whole dataset has been divided into 5 sets according to the angle between lighting source direction and the frontal face direction. To better explore the performance of our proposed illumination invariant descriptor, we conducted experiments using all frontal lighting images as gallery and the rest to form the probe set. Recognition accuracies were reported on set 1 to set

5 separately.

It is worth mentioning that the angle between frontal face direction and lighting orientations increases from set 1 to set 5. That is, it is generally more challenging to handle the face recognition task as the angle increases. This is confirmed by the results in Table 1, where the "*All*" in Table 1 means the results were calculated by taking all other 63 images per subject except the gallery one as probe. As shown in Table 1, it is interesting to find that most methods performed well when the angle is small but dramatically degenerate when the angle is large. For example, HE obtained the best accuracy 97.81% on set 1, even higher than those of WF and GF, but only achieved 13.43% recognition rate in set 5. The reduction in accuracy is notable about 84%. However, WF, GF and our proposed method turned out to be consistently robust, achieving relative high performances in all subsets. The general recognition accuracies of WF, GF and LGH on YaleB-Extended database are 86.88%, 86.93% and 99.37% respectively. Specially, we observe that our proposed LGH outperformed all other methods and achieves consistent highest performance in all subsets. The accuracies of LGH on set 1 to set 5 are 100%, 100%, 99.12%, 99.44% and 98.89%, respectively, obtaining 4.82%, 0%, 15.13%, 19.74% and 15.65% improvement as compared to the second best approach (i.e., GF).

5.2. Uncontrolled homogeneous lighting

5.2.1. Databases and setting

For a more challenging test, we evaluated our proposed method on the Face Recognition Grand Challenge version 2 data set (FRGC v2.0) [1], which is a large public data set involving illumination changes. It contains about 50,000 images captured from 625 subjects, under controlled environments, uncontrolled indoor and outdoor settings, and all images contain diverse variations in illumination, expression, and ornaments (glasses). To be more detailed, the controlled images with two facial expressions were taken under two lighting conditions, and the uncontrolled images were taken in the varying illumination

Fig. 4. Image samples in (a) YaleB-Extended and (b) FRGC datasets. For FRGC dataset, the ones in the upper row come from the controlled subset, and those in the bottom row come from the uncontrolled subset.

Table 1
Results on Extended YaleB (in accuracy (%)).

Methods	Set1	Set2	Set3	Set4	Set5	All
HE [17]	97.81	92.76	36.18	10.90	13.43	40.35
SQI [21]	83.33	100.00	69.30	75.75	72.99	79.03
LTV [22]	87.28	99.78	66.45	45.30	43.07	62.03
WF [25]	78.95	99.78	78.51	89.29	84.76	86.88
GF [24]	95.18	100.00	83.99	79.70	83.24	86.93
TT-LTP [23]	100.00	100.00	95.61	61.28	33.38	70.47
LGH	**100.00**	**100.00**	**99.12**	**99.44**	**98.89**	**99.37**

conditions, e.g., hallways, atriums, and outdoor. Following the protocol used in [57], we selected the subset containing at least five controlled images and five uncontrolled images with neutral expressions. At last, we formed a dataset containing 339 identities where each identity has 5 controlled images and 5 uncontrolled images. All images were aligned according to eyes coordinates, cropped and resized to 128×128. Some examples are shown in Fig. 4(b).

5.2.2. Results on FRGC

Since there are two subsets (i.e., controlled and uncontrolled) in FRGC database, we evaluated our proposed method according to the following two protocols:

- Protocol I: In each subset, one image per subject was chosen as gallery and the other four images per subject were used as probes.
- Protocol II: One image per subject in the controlled subset (denoted as Controlled) was chosen as gallery and all five images per subject in the uncontrolled subset (denoted as Uncontrolled) were used as probes.

Protocol I: According to Protocol I, we chose one out of five images per subject to form the gallery set at each time. The whole process was repeated five times and we reported the average recognition accuracies. The comparisons were carried out on the two subsets separately. All results are illustrated in Fig. 5(a) and (b). As shown in Fig. 5, all different methods performed similarly to the case on CMU-PIE and YaleB-E in general, but their average accuracies reduced. It is because there are various variations including illumination changes on this database, so that most of the compared methods cannot model the non-illumination effect. Nevertheless, as a kind of histogram based descriptors, TT-LTP and our proposed LGH outperformed other descriptors in this setting, achieving relative high recognition rates, i.e., 89.70% and 92.24%, respectively.

On the more challenging uncontrolled subset, WF and GF, which performed well in the controlled setting, degenerated significantly here and achieved average rank-1 accuracies less than 50%. TT-LTP and our proposed method again obtained better performance than other methods, but the difference between TT-LTP and LGH is clear. The average recognition rate of LGH is 72.01%, obtaining more than14% improvement when compared to TT-LTP.

Protocol II: Following Protocol II, we conducted experiments using one sample per subject from the controlled subset to form the gallery and all samples in the uncontrolled subset form the probe set. Similar to the case in Protocol I, we repeated five times to ensure that each image in the controlled subset has been involved in the experiment. The final average cumulative match score curves are reported in Fig. 5(c). As expected, the results of traditional methods such as HE, SQI, LTV and GF are not satisfactory in this setting. In addition to the various variations in the dataset itself, the reason is because the face images captured in the uncontrolled environment were affected by other interference factors. Also, the commonly used Lambertian reflectance model may not strictly hold in this complex uncontrolled environment. In comparison, our proposed LGH also outperformed all other methods and achieved much better results, where LGH gained

more than 25% matching rate at rank-1 recognition as compared to the second best approach.

5.3. Heterogeneous lighting

5.3.1. Databases and setting

To further investigate the effectiveness of our proposed LGH in handling heterogeneous lighting, we selected VIS–NIR subset of HFB database [56] for evaluation. The HFB VIS–NIR dataset contains 100 people, and each one has 4 VIS and 4 NIR face images. Note that the spectral wavelength of NIR light used here is around 850 nm and that of VIS light is smaller than 700 nm. It is more challenging since they were captured under two distinct heterogeneous light sources with no overlapping wavelength. As in FRGC dataset, variations caused by expression and occlusion (by glasses) also exist in HFB, making the recognition more challenging. All images were aligned according to the eyes coordinates, cropped and resized to 128×128. For comparison, the results of DoG+LBP [33] and SIFT descriptor used in [34] accompanied with methods mentioned in the former section are also reported.

5.3.2. Results on HFB

In this part, the gallery set consisted of one VIS face images per individual and the probe one contains all NIR face images. We report the means and standard deviations of recognition accuracies using different galleries in Fig. 6. As shown, the traditional approaches which are designed to overcome the homogeneous illumination problem failed in this scenario. In Fig. 6, the overall performance of all methods degenerated dramatically as compared to the case under controlled homogeneous lighting. The average recognition rates of traditional approaches such as HE, SQI, LTV, WF, GF and TT-LTP are only 3.25%, 4.13%, 7.56%, 15.56%, 9.88% and 12.50%, respectively. Meanwhile, the two task specific methods, namely DoG+LBP [33] and SIFT [34], did not show their superiorities here without the learning-based feature selection procedure. However, our proposed LGH obtained significant better results due to its relative robust and effective histogram based representation. As analyzed, it is not so sensitive to the heterogeneous illumination variation and thus is able to achieve the best results among all compared approaches.

5.4. Further analysis

5.4.1. Toleration to pose variation

We claimed that our local patch histogram based representation is more robust against misalignment or small pose variation as compared to the pixel-wised matching. In this section, we verify it on CMU Multi-PIE dataset [58].

The CMU Multi-PIE face database contains more than 750,000 images of 337 people recorded up to four sessions over five months. Subjects were imaged under 15 view points and 19 illumination conditions, and facial expressions exist. Since this work concerns illumination and pose variation, we only used the subset captured in session01, which contains images captured under 5 view points (+00, −15, +15, −30, +30) and under all illumination conditions with natural expression. Note that the denotation of view point +00 represents the angle between frontal face direction and the camera view; that is, '+00' indicates the frontal setting as used in all former databases. For convenience of description, we also used "+00", "−15", "+15", "−30", and "+30" to denote the subsets in session01. The frontal illuminated images in subset +00 were used as galleries, and the rest images in +00 and those in other four subsets formed five probe sets separately. For comparison, two local patch histogram based descriptors, i.e. LBP [14] and HOG [59], were used as baselines. All results are shown in Table 2.

Most of the compared methods turned out to be saturated in this setting, so the distinction between pixel-wise representation and local histogram based representation is not clear. However, when the angle increases to 15̄, the recognition rates of pixel-wise methods signifi-

(a) Controlled

(b) Uncontrolled

(c) Controlled-Uncontrolled
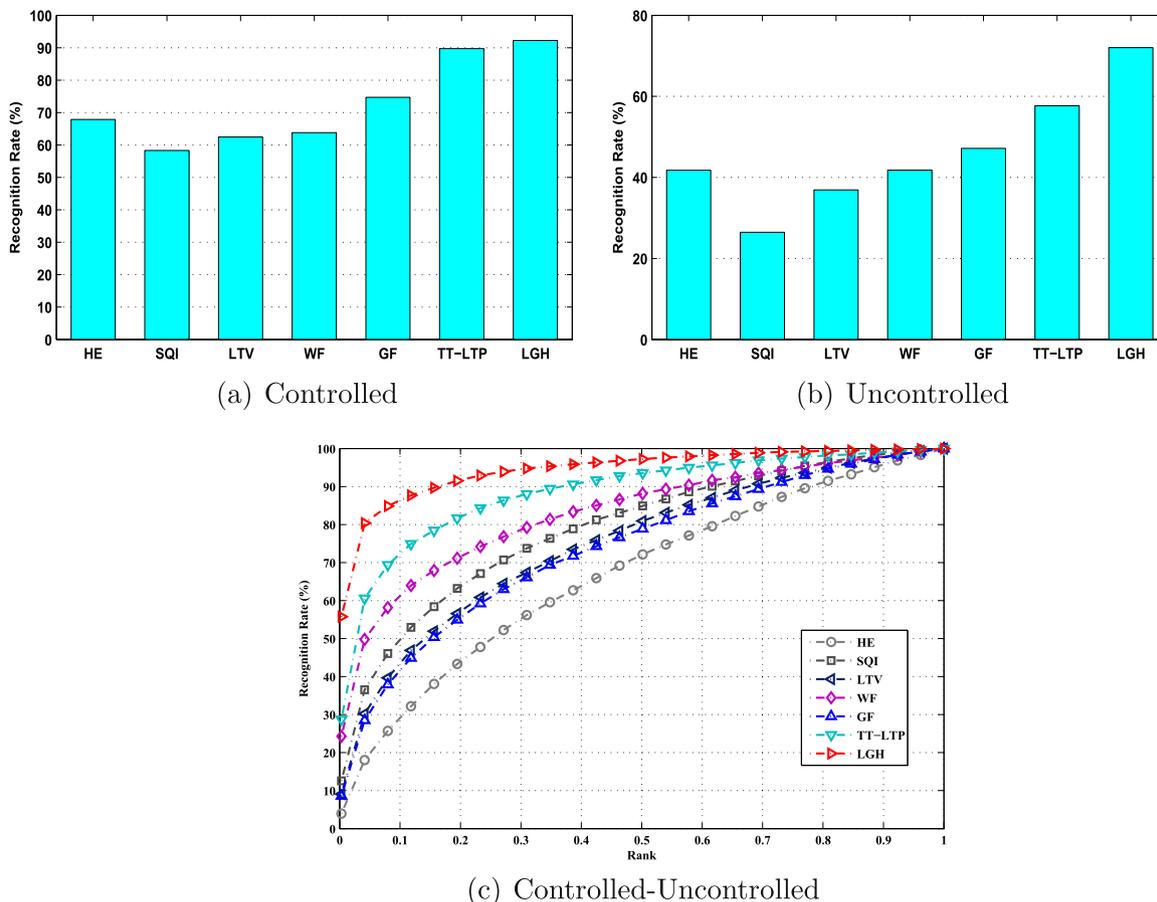
Fig. 5. Recognition accuracies of various methods under one sample per person scenario on FRGC v2.0 database. (a) and (b) are results under controlled lighting and uncontrolled lighting conditions, respectively. (c) is the cumulative match score curves (CMC) using controlled images to form the gallery set and using uncontrolled images to form the probe set.
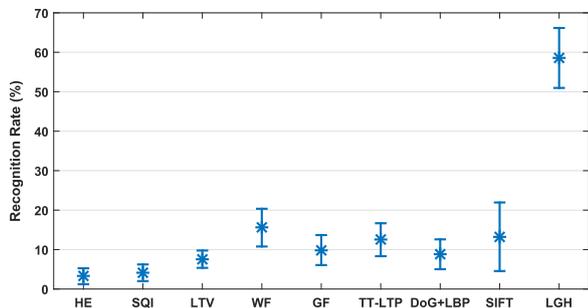
Fig. 6. Means and standard deviations of recognition accuracies (%) on HFB VIS–NIR dataset.

Table 2
Results on CMU Multi-PIE dataset (in accuracy (%)).

Methods	+00	−15	+15	−30	+30
HE [17]	66.96	48.49	45.26	22.89	23.45
SQI [21]	99.79	26.33	26.55	11.53	9.26
LTV [22]	94.44	5.24	4.58	2.49	2.53
WF [25]	100.00	48.49	45.78	20.76	19.18
GF [24]	99.66	55.84	54.86	23.49	25.60
TT-LTP [23]	97.95	70.38	66.37	31.39	31.71
LBP [14]	66.96	48.49	45.26	22.89	23.45
HOG [59]	98.63	64.38	65.20	32.07	31.69
LGH	100.00	90.48	87.95	46.49	39.50

cantly dropped. For instance, the recognition accuracy of SQI reduced from 99.79% to around 26%, and the situation is even worse for LTV which dropped from 94.44% to about 5%. A small disturbance in alignment or pose variation would probably impose great impact on the recognition accuracy of pixel-wise matching. Fortunately, owing to the illumination insensitive components extraction and local patch histogram representation, our proposed LGH has shown its great toleration in small pose variations. The recognition rates of LGH in subset −15 and +15 are 90.48% and 87.95%, with 20% better than the second best approach (i.e., TT-LTP). However, as the angle increases to 30°, pose variation becomes dominant and thus all methods turned out unsatisfactory. Nevertheless, the overall performance of local patch histogram based methods is still better than that of using pixel-wise representation, and meanwhile our LGH outperformed all other methods.

5.4.2. Contribution of each component in LGH

Since LGH consists of three parts including LMSN-LoG filtering (denoted as LoG), logarithm gradient magnitude (LGM) and logarithm gradient orientation (LGO), we explored the contribution of each part. In this section, we conducted experiments on CMU-PIE, Extended YaleB, FRGC and HFB VIS-NIR databases and report the recognition rates of each part separately in Fig. 7. Note that, only one near frontal illuminated face image was used as gallery in all four databases, and the LoG filtering used here was only conducted using a single scale. The final results validate our previous analysis that (i) the LMSN-LoG filtering is able to restrain the illumination effect; (ii) gradient magnitude and gradient orientation in logarithmic domain are insensitive to the illumination change; (iii) our proposed LGH integrates the above two components in an effective way and achieves great success
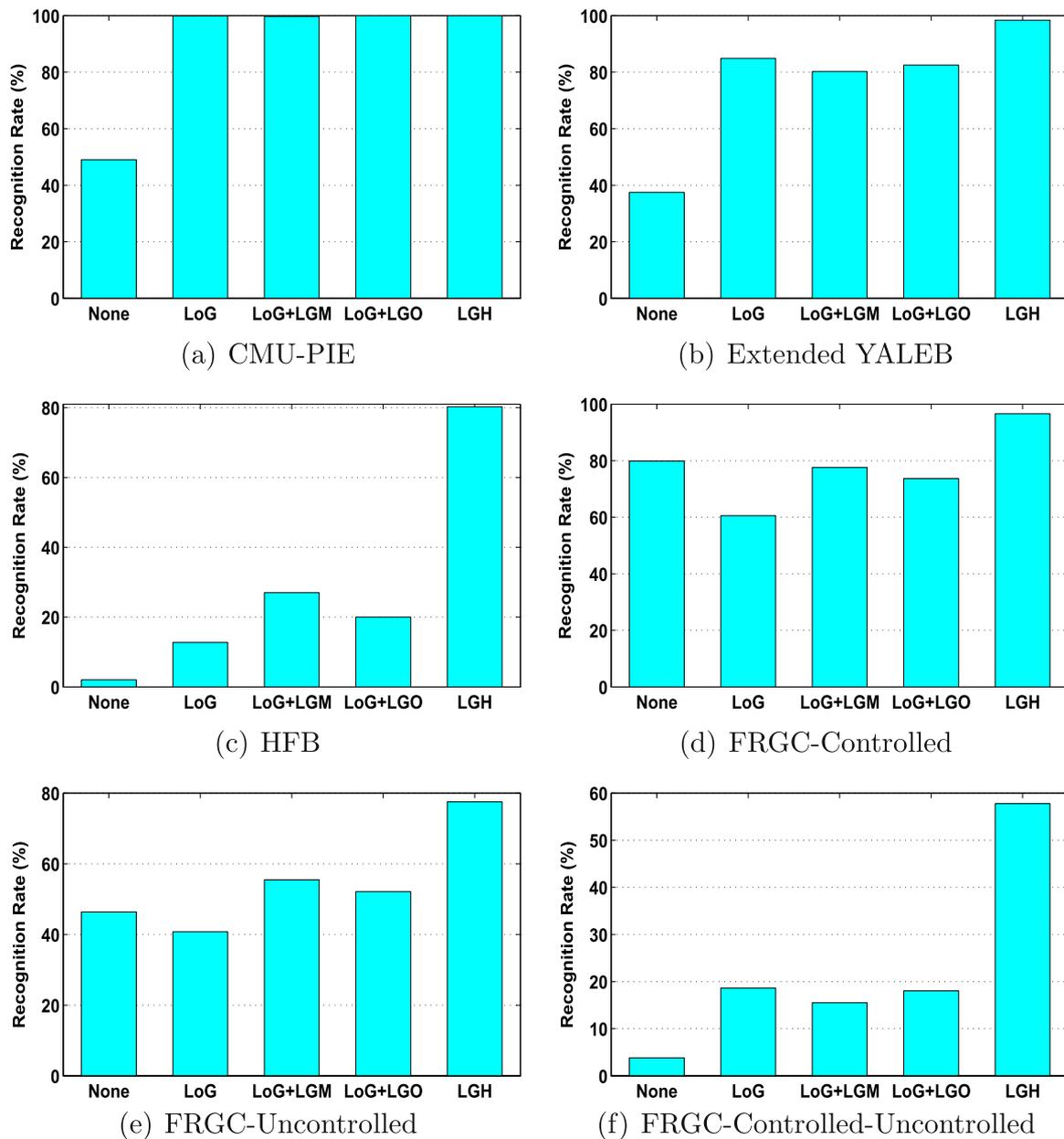
Fig. 7. Contribution of each part in LGH. Only one near frontal illuminated face image was used as gallery. Notations used in the *x*-axis from left to right indicate original image (None), image after LoG filtering (LoG), logarithm gradient magnitude followed by LoG filtering (LoG+LGM), gradient orientation followed by LoG filtering (LoG+LGO) and the whole LGH (LGH), respectively.

on tolerance of lighting change, especially when the change is severe.

5.4.3. Parameter selection and time complexity

In this section, we evaluated the effect of parameters and the time consumption empirically of the proposed LGH. Two main parameters of LGH are the number of scale *s* in multi-scale representation and the number of bins used in histogram generation. In the following part, we evaluated the effect of parameters by varying the number of scale and bin at the same time. Specifically, *s* was set to increase from 1 to 6 with step 1, and the bin was chosen in the set of *bins* = [2, 3, 5, 7, 9, 11, 13]. Since the size of face image is 128×128 only, the maximum number of scale we used was set to 6. All results are shown in Fig. 8. As observed, using a suitable number of scales such as $s \in \{2, 3, 4\}$ may lead to better recognition results when compared to using single scale. But using a too large scale (e.g. *s*=6) may do harm to the recognition performance, e.g., the cases as shown in Fig. 8(c–f). This is caused probably because the features extracted from large scale belong to the

very low frequency domain, and thus they are easily corrupted by the illumination change. Regarding the number of bins, we find that significant improvement is gained when the number of bins varies from 2 to 5, and the performance still increases slightly as more bins are incorporated, but finally the performance may degenerate when larger bin is used (e.g., *bin*=13 in Fig. 8(a), (b) and (f)).

We evaluated the time consumption of LGH feature extraction on different databases using different scales and bins. All experiments were conducted on a server with Intel i5 2.2 GHz cores CPU and the average time consumption is reported in Fig. 9. It can be found that the time cost of LGH is linear to the number of probe images and the number of scales. It took about 16.7 ms to calculate LGH feature for each 128×128 face image using 1 scale when *bin*=5, and the time cost for 4 scales is 66.7 ms, exactly 4 times of the 1 scale's. What is more, the more bins we use the more time it needs. Hence, a balance between time cost and performance is necessary.

(a) CMU-PIE

(b) YaleB Extended

(c) FRGC Controlled

(d) FRGC Uncontrolled

(e) FRGC Controlled-Uncontrolled

(f) HFB
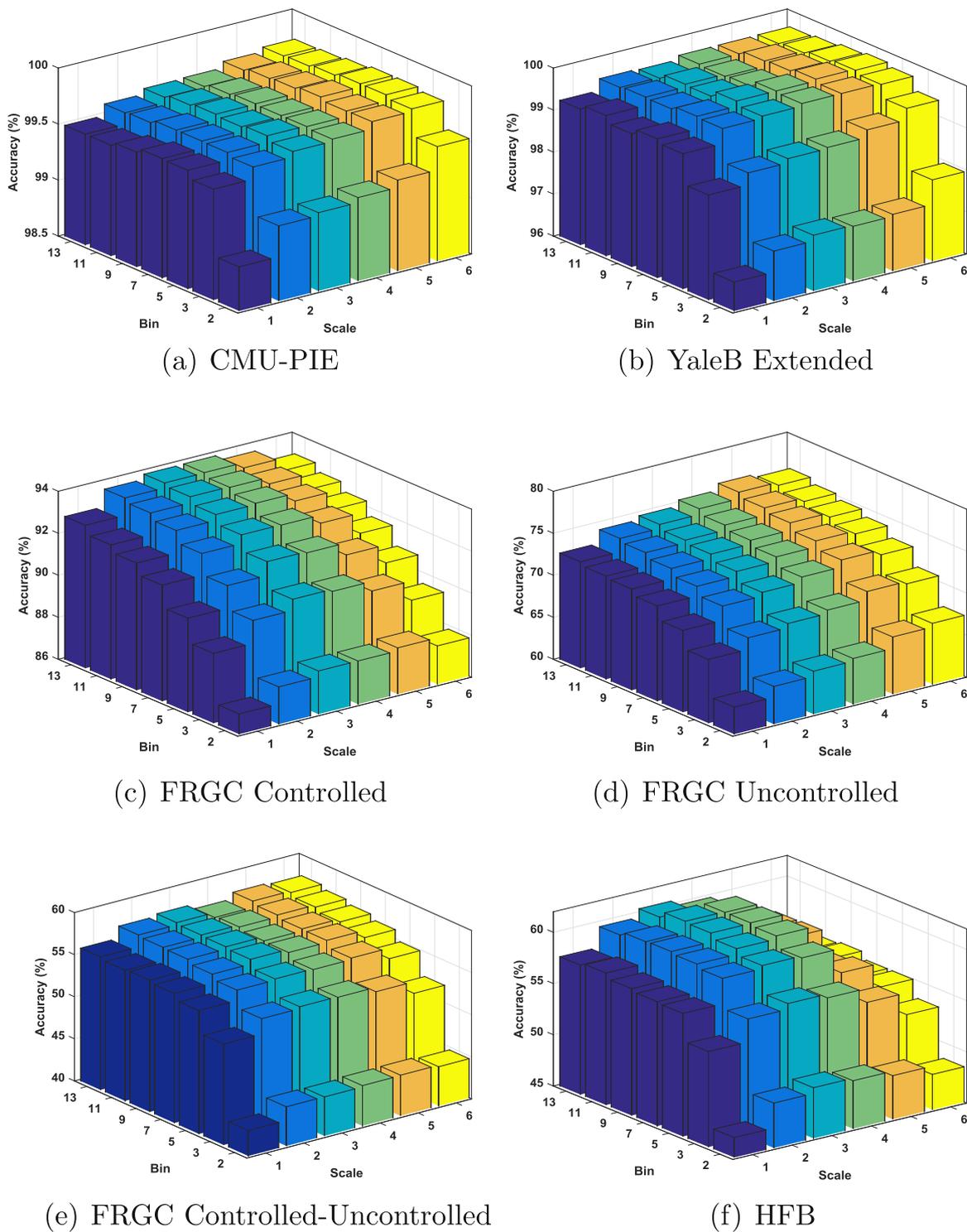
Fig. 8. Performances of the proposed method under different parameters. The number of scales increases from 1 to 6 with step 1, and the number of bins is chosen from the set $BinSet$ = [2, 3, 5, 7, 9, 11, 13].

5.4.4. Comparison with the state-of-the-art face descriptor

Recently, learning based approaches, especially the deep learning, have achieved great success in face recognition [4,3,41,43]. However, these results were reported on Internet images, in which there can be multiple images available for each person and the lighting change is not serious, while our focus (including our model) is specific to face recognition against serious lighting change. Hence, we would like to compare our face representation to state-of-the-art learning based features. In this part, by following the one sample per person protocol, we compared our method to VGG-Face CNN descriptor (VGG) [41] and

a recently proposed learning-based compact binary face descriptor (CBFD) [43]. Note that, we conducted LMSN-LoG filtering on both four scales and one single scale in order to see the difference of scale selection. All results can be found in Table 3. On FRGC database, VGG and CBFD achieved slightly better results in homogeneous lighting condition on both the controlled subset (denoted as FRGC-C) and uncontrolled subset (denoted as FRGC-U), but the case changes when matching uncontrolled lighting face images with the controlled ones (denoted as FRGC-C-U). In the latter case, our proposed LGH and the CBFD obtained significant better performance than VGG. The same

(a) Time v.s. Scale

(b) Time v.s. Bin
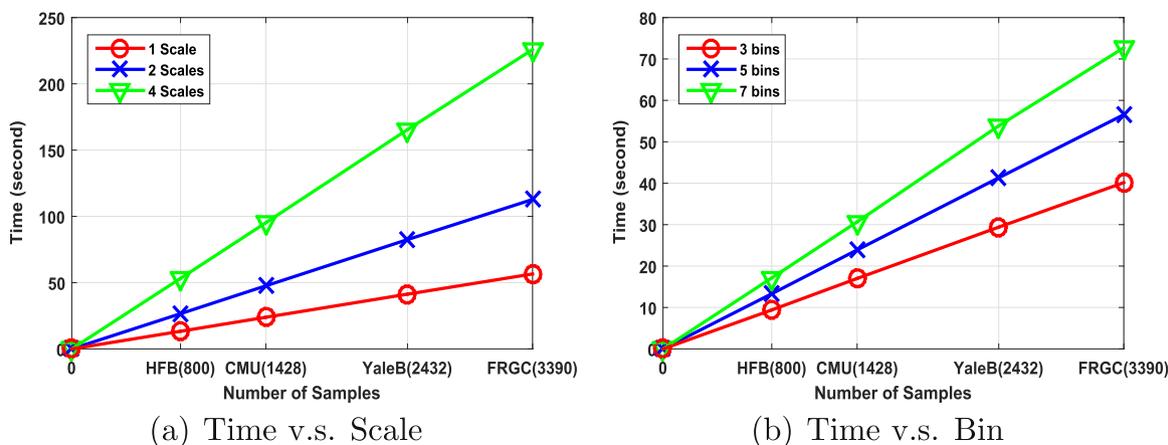
Fig. 9. Time consumption of the proposed method. (a) Average computation time under different scales when fixing bin=5 on four databases. (b) Average computation time under different bins when fixing scale=1 on four databases.

Table 3
Comparison between LGH and learning-based features.

Dataset	Chi-square distance	Recognition accuracy (%)				
		Intensity	LGH (s=1)	LGH (s=4)	VGG [41]	CBFD [43]
CMU-PIE	0.1565	37.54	99.57	99.66	92.38	98.61
YaleB-E						
ALL	0.1395	37.80	98.25	99.37	62.91	97.45
SET1	0.8193	99.12	100	100	100	100
SET2	0.6347	97.15	100	100	99.56	100
SET3	0.1148	37.28	98.46	99.12	85.31	100
SET4	0.0396	7.89	99.62	99.44	48.68	96.80
SET5	0.0179	3.32	97.09	98.89	24.38	93.91
FRGC-C	0.6229	70.18	89.00	92.24	95.08	98.08
FRGC-U	0.3587	43.42	67.89	72.01	73.11	72.46
FRGC-C-U	0.0309	4.22	51.60	55.75	36.85	53.03
HFB	0.0358	5.19	56.69	58.56	38.00	59.00

comparison results can be found on HFB database, where LGH obtained great improvement as compared to VGG due to its robustness in handling heterogeneous lighting, which is not considered by VGG. We also note that although the proposed LGH is non-learning-based and CBFD is learning-based, the overall performance of LGH is similar to that of CBFD, which verifies the effectiveness of LGH. Some failure cases of CBFD and LGH are shown in Fig. 10. Since the proposed LGH is learning free and only requires basic photometric operations, it is a competitive tool in handling face recognition involving both homogeneous and heterogeneous lighting variations.

5.4.5. Further discussion

In order to have a deep insight into the selected four publically available datasets, we attempt to quantify the difficulty level of face datasets. For convenience, we summarize a brief description of all datasets used in the experiment in Table 4. We first calculated the Chi-square distance between samples from the same class, denoted as matched score, and the distance between samples from different classes, denoted as non-matched score, using the original intensities as image feature. We used this overlap to measure the difficulty level of face classification (some examples are shown in Fig. 11). A smaller Chi-square distance indicates that the classification is more difficult. At the same time, we report the recognition accuracies using the original image intensity feature as baselines. All results are shown in Table 3. From the results, we find that:

(1) The CMU-PIE and Yale-B Extended datasets (denoted as Group A) only contain variations caused by illumination changes, but the samples in FRGC and HFB datasets (denoted as Group B) suffer from diverse variations, such as facial expression, glasses occlusion and illumination changes. For each group, a small Chi-square distance between the distribution of matched scores and that of non-matched scores will probably lead to low recognition accuracies, indicating that the classification on that dataset is more difficult. Typically, in the YaleB-Extended dataset, the angle between the frontal view direction and the lighting direction increases from set 1 to set 5, and the Chi-square distance decreases correspondingly. And it is the fact that the classification task becomes more challenging as the set number increases.

(2) The performance on FRGC and HFB datasets (denoted as

(a) CBFD failed

(b) LGH failed
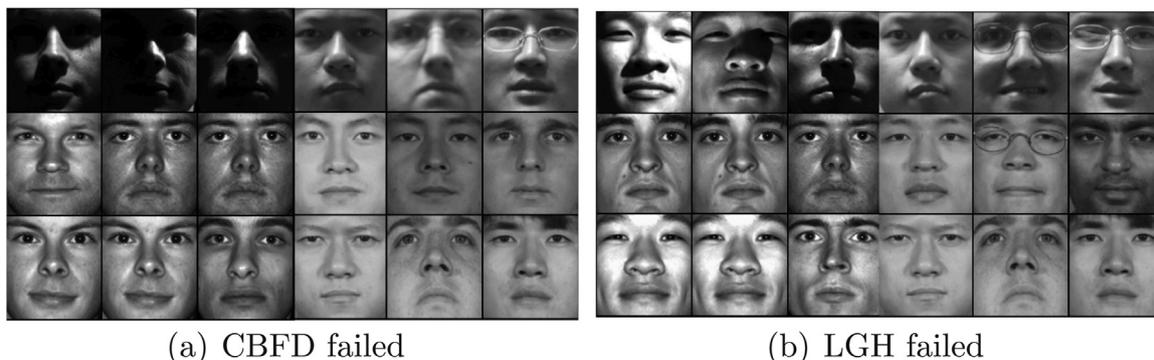
Fig. 10. Failure cases of CBFD and LGH. The upper row represents probe images to be recognized, while the middle row illustrates misclassified results for CBFD and LGH methods, and the ground truth gallery images are shown in the bottom row. (a) LGH makes correct match but CBFD does not; (b) CBFD makes correct match but LGH does not.

Table 4
Description of the datasets employed in the experiments.

Dataset	Description
CMU-PIE	68 individuals, 21 different lighting conditions; all images captured in a few seconds, mainly caused by illumination variation
YaleB-E	
ALL	38 individuals, 38 different lighting conditions; all images captured in a few seconds, mainly caused by illumination variation
SET1	The angle between lighting direction and frontal direction, denoted as S, is smaller than 12 degree
SET2	$12° < S \leq 25°$
SET3	$25° < S \leq 50°$
SET4	$50° < S \leq 77°$
SET5	$S > 77°$
FRGC	
Controlled	339 identities; controlled indoor environment; suffering from diverse variations in illumination, expression and glasses occlusion
Uncontrolled	339 identities; uncontrolled indoor and outdoor environment; suffering from diverse variations in illumination, expression and glasses occlusion
Controlled–uncontrolled	1 controlled image used as gallery and 5 uncontrolled images used as probes
HFB	100 individuals, each having 4 VIS images and 4 NIR images; variations caused by heterogeneous lighting, expression and glasses occlusion; 1 VIS image used as gallery and 4 VIS images used as probes

Group B) is inferior to that on CMU-PIE and Yale-B Extended datasets when the Chi-square distances are similar. The reason could be that, on one hand the samples in FRGC and HFB suffered from variations caused by other factors but not just the illumination change, and on the other hand, the complex lighting condition in FRGC and HFB datasets involves various spectral wavelengths, so that the classification on FRGC and HFB is more difficult.

6. Conclusions

In this paper, we have proposed a novel illumination invariant descriptor LGH to address the illumination problem in face recogni-

tion. Different from the existing models, we consider variations caused by the lighting direction, magnitude and the spectral wavelength together, so that the new proposed descriptor is able to handle face recognition under both homogeneous and heterogeneous lighting conditions. We have proposed two illumination invariant components in the logarithm domain after bank-pass filtering, i.e., LGO and LGM. After that, we integrate them into a histogram based feature representation followed by post-processing to enhance the fault-tolerant ability. We have provided solid theoretical analysis on illumination invariant property of the proposed descriptors. Experimental results verify the effectiveness of our proposed model on tackling single-image-based face recognition under serious illumination problems from the homogeneous lighting to heterogeneous lighting.
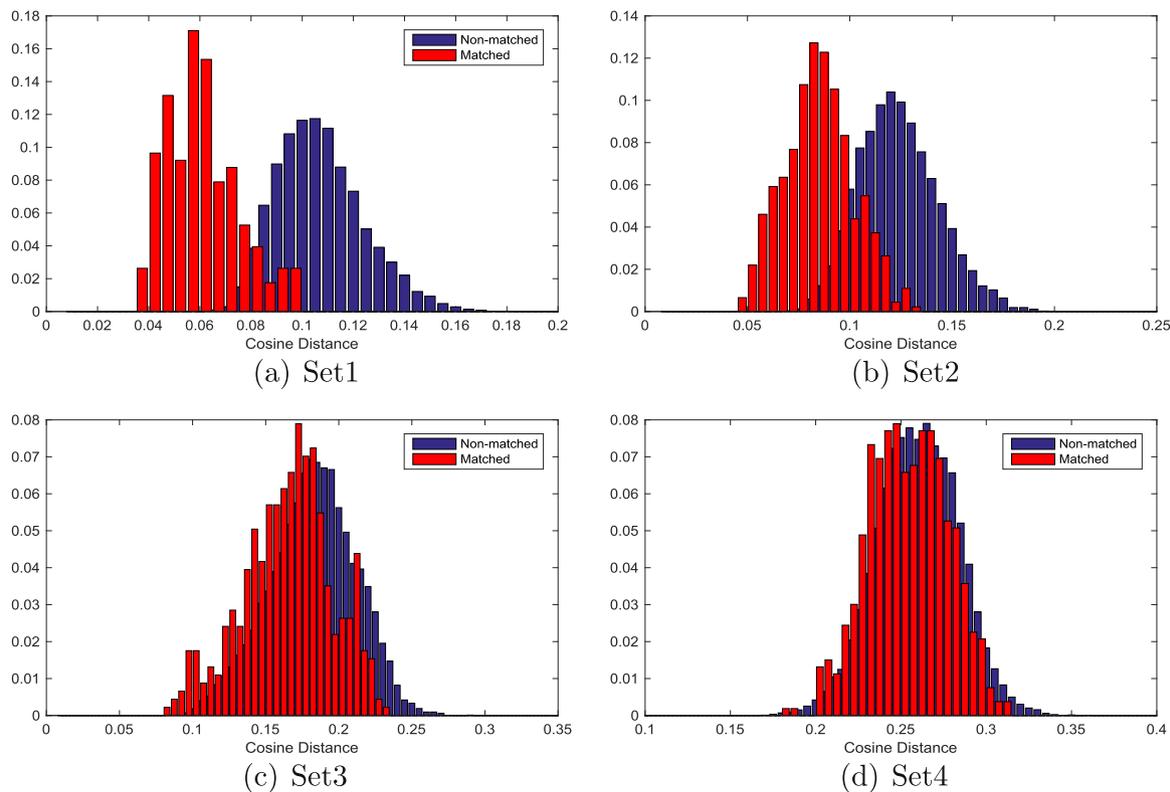
Fig. 11. Distributions of matched scores and non-matched scores on YaleB-E dataset.

References

[1] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, vol. 1, 2005, pp. 947–954.

[2] P. Phillips, W. Scruggs, A. O'Toole, P. Flynn, K. Bowyer, C. Schott, M. Sharpe, Frvt 2006 and ice 2006 large-scale experimental results, IEEE Trans. Pattern Anal. Mach. Intell. 32 (5) (2010) 831–846.

[3] Y. Sun, X. Wang, X. Tang, Deep learning face representation from predicting 10, 000 classes, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 2014, pp. 1891–1898.

[4] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, Deepface: closing the gap to human-level performance in face verification, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701–1708.

[5] A.K. Jain, B. Chandrasekaran, Dimensionality and sample size considerations in pattern recognition practice, in: Handbook of Statistics, vol. 2, 1982, pp. 835–855.

[6] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cogn. Neurosci. 3 (1) (1991) 71–86.

[7] P. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.

[8] X. Tan, S. Chen, Z.-H. Zhou, F. Zhang, Face recognition from a single image per person: a survey, Pattern Recognit. 39 (9) (2006) 1725–1745.

[9] J. Yang, D. Zhang, A.F. Frangi, J.-y. Yang, Two-dimensional pca: a new approach to appearance-based face representation and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 26 (1) (2004) 131–137.

[10] D. Beymer, T. Poggio, Face recognition from one example view, in: Proceedings of the 5th IEEE International Conference on Computer Vision, 1995, pp. 500–507.

[11] A.M. Martínez, Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class, Pattern Anal. Mach. Intell. 24 (6) (2002) 748–763.

[12] R. Brunelli, T. Poggio, Face recognition: features versus templates, IEEE Trans. Pattern Anal. Mach. Intell. 15 (10) (1993) 1042–1052.

[13] L. Wiskott, J.-M. Fellous, N. Kuiger, C. Von Der Malsburg, Face recognition by elastic bunch graph matching, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 775–779.

[14] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (12) (2006) 2037–2041.

[15] S. Chen, J. Liu, Z.-H. Zhou, Making flda applicable to face recognition with one sample per person, Pattern Recognit. 37 (7) (2004) 1553–1555.

[16] X. Xie, J. Lai, W. Zheng, Extraction of illumination invariant facial features from a single image using nonsubsampled contourlet transform, Pattern Recognit. 43 (12) (2010) 4177–4189.

[17] S. Pizer, E. Amburn, J. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. Zimmerman, K. Zuiderveld, Adaptive histogram equalization and its variations, Comput. Vis. Graph. Image Process. 39 (3) (1987) 355–368.

[18] S. Shan, W. Gao, B. Cao, D. Zhao, Illumination normalization for robust face recognition against varying lighting conditions, in: Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures, 2003, pp. 157–164.

[19] A. Georghiades, P. Belhumeur, D. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 643–660.

[20] H. Shim, J. Luo, T. Chen, A subspace model-based approach to face relighting under unknown lighting and poses, IEEE Trans. Image Process 17 (8) (2008) 1331–1341.

[21] H. Wang, S. Li, Y. Wang, Generalized quotient image, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004, pp. 498–505.

[22] T. Chen, W. Yin, X. Zhou, D. Comaniciu, T. Huang, Total variation models for variable lighting face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (9) (2006) 1519–1524.

[23] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, in: Proceedings of the 3rd IEEE International Conference on Analysis and Modeling of Faces and Gestures, 2007, pp. 168–182.

[24] T. Zhang, Y. Tang, B. Fang, Z. Shang, X. Liu, Face recognition under varying illumination using gradientfaces, IEEE Trans. Image Process. 18 (11) (2009) 2599–2606.

[25] B. Wang, W. Li, W. Yang, Q. Liao, Illumination normalization based on Weber's law with application to face recognition, IEEE Signal Process. Lett. (2011) 462–465.

[26] M. Sonka, V. Hlavac, R. Boyle, Image Processing: Analysis and Machine Vision, Chapman & Hall, London, 1993.

[27] R. Basri, D. Jacobs, Lambertian reflectance and linear subspaces, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2) (2003) 218–233.

[28] V. Blanz, T. Vetter, Face recognition based on fitting a 3d morphable model, IEEE Trans. Pattern Anal. Mach. Intell. 25 (9) (2003) 1063–1074.

[29] W.Y. Zhao, R. Chellappa, Symmetric shape-from-shading using self-ratio image, Int. J. Comput. Vis. 45 (1) (2001) 55–75.

[30] W. Zhang, S. Shan, W. Gao, X. Chen, H. Zhang, Local Gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition, in: Proceedings of the 10th IEEE International Conference on Computer Vision, vol. 1, 2005, pp. 786–791.

[31] B. Zhang, S. Shan, X. Chen, W. Gao, Histogram of Gabor phase patterns (hgpp): a novel object representation approach for face recognition, IEEE Trans. Image Process. 16 (1) (2007) 57–68.

[32] Z. Lei, S. Liao, M. Pietikäinen, S.Z. Li, Face recognition by exploring information jointly in space, scale and orientation, IEEE Trans. Image Process. 20 (1) (2011) 247–256.

[33] S. Liao, D. Yi, Z. Lei, R. Qin, S.Z. Li, Heterogeneous face recognition from local structures of normalized appearance, in: Proceedings of the IAPR/IEEE International Conference on Biometrics, 2009, pp. 209–218.

[34] B. Klare, A. Jain, Heterogeneous face recognition: matching nir to visible light images, in: Proceedings of the IEEE International Conference on Pattern Recognition, 2010, pp. 1513–1516.

[35] D. Jobson, Z. Rahman, G. Woodell, A multiscale retinex for bridging the gap between color images and the human observation of scenes, IEEE Trans. Image Process. 6 (7) (1997) 965–976.

[36] R. Gross, V. Brajovic, An image preprocessing algorithm for illumination invariant face recognition, in: Proceedings of the 4th IEEE International Conference on Audio- and Video-Based Biometrie Person Authentication, 2003, pp. 10–18.

[37] Z. Hou, W. Yau, Relative gradients for image lighting correction, in: Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing, 2010, pp. 1374–1377.

[38] A. Jain, Fundamentals of Digital Image Processing, Prentice-Hall, 1989.

[39] Z. Cao, Q. Yin, X. Tang, J. Sun, Face recognition with learning-based descriptor, in: Proceedings IEEE International Conference on Computer Vision and Pattern Recognition, 2010, pp. 2707–2714.

[40] Z. Lei, M. Pietikäinen, S.Z. Li, Learning discriminant face descriptor, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2) (2014) 289–302.

[41] O.M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition, in: British Machine Vision Conference, 2015.

[42] L. Wolf, T. Hassner, Y. Taigman, The one-shot similarity kernel, in: IEEE International Conference on Computer Vision, 2009, pp. 897–902.

[43] J. Lu, V.E. Liong, X. Zhou, J. Zhou, Learning compact binary face descriptor for face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 37 (10) (2015) 2056–2410.

[44] K. Mikolajczyk, C. Schmid, Scale & affine invariant interest point detectors, Int. J. Comput. Vis. 60 (1) (2004) 63–86.

[45] S. Liu, D. Yi, Z. Lei, S. Li, Heterogeneous face image matching using multi-scale features, in: International Conference on Biometrics, 2012, pp. 79–84.

[46] J.-Y. Zhu, W.-S. Zheng, J.-H. Lai, Matching nir face to vis face using transduction, IEEE Trans. Inf. Forensics Secur. 9 (3) (2014) 501–514.

[47] B.F. Klare, A.K. Jain, Heterogeneous face recognition using kernel prototype similarities, IEEE Trans. Pattern Anal. Mach. Intell. 35 (6) (2013) 1410–1422.

[48] H. Bischof, H. Wildenauer, A. Leonardis, Illumination insensitive recognition using eigenspaces, Comput. Vis. Image Underst. 95 (1) (2004) 86–104.

[49] L. Zhang, D. Samaras, Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics, IEEE Trans. Pattern Anal. Mach. Intell. 28 (3) (2006) 351–363.

[50] J. Lu, Y.-P. Tan, G. Wang, Discriminative multimanifold analysis for face recognition from a single training sample per person, IEEE Trans. Biom. Compend. 35 (1) (2012) 39–51.

[51] J.Y. Zhu, W.S. Zheng, J.H. Lai, Logarithm gradient histogram: a general illumination invariant descriptor for face recognition, in: IEEE International Conference on Automatic Face and Gesture Recognition, 2013, pp. 1–8.

[52] T. Lindeberg, Edge detection and ridge detection with automatic scale selection, Int. J. Comput. Vis. 30 (2) (1998) 117–156.

[53] R.A. Young, The Gaussian derivative model for spatial vision: I. Retinal mechanisms, Spat. Vis. 2 (4) (1987) 273–293.

[54] I. Meglinski, S. Matcher, Quantitative Assessment of Skin Layers Absorption and Skin Reflectance Spectra Simulation in the Visible and Near-infrared Spectral Regions, vol. 23, Iop Publishing, 2002.

[55] T. Sim, S. Baker, M. Bsat, The cmu pose, illumination, and expression (pie) database, in: Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition, 2002, pp. 46 – 51.

[56] S. Li, Z. Lei, M. Ao, The hfb face database for heterogeneous face biometrics research, in: Proceedings IEEE Computer Society Workshop on Computer Vision and Pattern Recognition, 2009, pp. 1–8.

[57] X. Xie, W.-S. Zheng, J. Lai, P.C. Yuen, C.Y. Suen, Normalization of face illumination based on large-and small-scale features, IEEE Trans. Image Process. 20 (7) (2011) 1807–1821.

[58] R. Gross, I. Matthews, J.F. Cohn, T. Kanade, S. Baker, Multi-pie, Image Vis. Comput. 28 (5) (2010) 807–813.

[59] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, pp. 886–893.

Jun-Yong Zhu received his M.S. and Ph.D. degrees in applied mathematics in the School of Mathematics and Computational Science from Sun Yat-sen University,

Guangzhou, PR China, in 2010 and 2014, respectively. Currently, he is working toward the post-doctoral research in the Department of Information Science and Technology, Sun Yat-Sen University. His current research interests include heterogeneous face recognition, visual transfer learning using partial labeled or unlabeled auxiliary data and non-linear clustering. He has published several papers in international conferences such as ICIP, AMFG and ICDM. His cooperative ICDM 2010 paper won the Honorable Mention for Best Research Paper Awards and his CCBR 2012 paper won the Best Student Paper Awards.

Wei-Shi Zheng received the Ph.D. degree in applied mathematics from Sun Yat-Sen University in 2008. He is now a professor at Sun Yat-sen University. He had been a postdoctoral researcher on the EU FP7 SAMURAI Project at Queen Mary University of London and an associate professor at Sun Yat-sen University after that. He has now published more than 80 papers, including more than 40 publications in main journals (TPAMI, TNN, TIP, TSMC-B, PR) and top conferences (ICCV, CVPR, IJCAI, AAAI). He has joined the organization of four tutorial presentations in ACCV 2012, ICPR 2012, ICCV 2013 and CVPR 2015 along with other colleagues. His research interests include person/object association and activity understanding in visual surveillance. He has joined Microsoft Research Asia Young Faculty Visiting Programme. He is a recipient of excellent young scientists fund of the national natural science foundation of China, and a recipient of Royal Society-Newton Advanced Fellowship.

Feng Lu received his M.S. and Ph.D. degrees in computer application in the School of Mathematics and Computational Science from Sun Yat-sen University, Guangzhou, PR China, in 1999 and 2002. He had been a postdoctoral researcher in the Department of Finance Bureau, Guangzhou, Guangdong during 2002–2004. Currently, he is the office director of Department of Public Security of Guangdong Province. His research interests include scientific and technological information in public security, automatic crime data analysis, intelligent analysis and application technology for surveillance video.

Jian-Huang Lai received his M.S. degree in applied mathematics in 1989 and his Ph.D. in mathematics in 1999 from Sun Yat-Sen University, China. He joined Sun Yat-sen University in 1989 as an assistant professor, where currently, he is a professor with the Department of Data and Computer Science. His current research interests are in the areas of digital image processing, pattern recognition, multimedia communication, wavelet and its applications. He has published over 80 scientific papers in international journals and conferences on image processing and pattern recognition, e.g. IEEE TPAMI, IEEE TNN, IEEE TIP, IEEE TSMC (Part B), Pattern Recognition, ICCV, CVPR and ICDM. Prof. Lai serves as a standing member of the Image and Graphics Association of China and also serves as a standing director of the Image and Graphics Association of Guangdong.