Face Recognition Under Varying Illumination

2 authors:

Zhichao Lian
Yale University
18 PUBLICATIONS   **134** CITATIONS

Meng Joo Er
Nanyang Technological University
334 PUBLICATIONS   **7,365** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   Progressively Learning Universal Classifier for Binary, Multi-class and Multi-label Problems View project

Project   Study III-V Material enhance the switching ratio View project

Face Recognition under Varying Illumination

Lian Zhichao and Er Meng Joo
Nanyang Technological University
Singapore

1. Introduction

Face Recognition by a robot or machine is one of the challenging research topics in the recent years. It has become an active research area which crosscuts several disciplines such as image processing, pattern recognition, computer vision, neural networks and robotics. For many applications, the performances of face recognition systems in controlled environments have achieved a satisfactory level. However, there are still some challenging issues to address in face recognition under uncontrolled conditions. The variation in illumination is one of the main challenging problems that a practical face recognition system needs to deal with. It has been proven that in face recognition, differences caused by illumination variations are more significant than differences between individuals (Adini et al., 1997). Various methods have been proposed to solve the problem. These methods can be classified into three categories, named face and illumination modeling, illumination invariant feature extraction and preprocessing and normalization. In this chapter, an extensive and state-of-the-art study of existing approaches to handle illumination variations is presented. Several latest and representative approaches of each category are presented in detail, as well as the comparisons between them. Moreover, to deal with complex environment where illumination variations are coupled with other problems such as pose and expression variations, a good feature representation of human face should not only be illumination invariant, but also robust enough against pose and expression variations. Local binary pattern (LBP) is such a local texture descriptor. In this chapter, a detailed study of the LBP and its several important extensions is carried out, as well as its various combinations with other techniques to handle illumination invariant face recognition under a complex environment. By generalizing different strategies in handling illumination variations and evaluating their performances, several promising directions for future research have been suggested.

This chapter is organized as follows. Several famous methods of face and illumination modeling are introduced in Section 2. In Section 3, latest and representative approaches of illumination invariant feature extraction are presented in detail. More attentions are paid on quotient-image-based methods. In Section 4, the normalization methods on discarding low frequency coefficients in various transformed domains are introduced with details. In Section 5, a detailed introduction of the LBP and its several important extensions is presented, as well as its various combinations with other face recognition techniques. In Section 6, comparisons between different methods and discussion of their advantages and disadvantages are presented. Finally, several promising directions as the conclusions are drawn in Section 7.

2. Face and illumination modelling

Here, two kinds of modeling methods named face modeling and illumination modeling will be introduced. Regarding face modeling, because illumination variations are mainly caused by three-dimension structures of human faces, researchers have attempted to construct a general 3D human face model in order to fit different illumination and pose conditions. One straight way is to use specific sensors to obtain 3D images representing 3D shape of human face. A range image, a shaded model and a wire-frame mesh are common alternatives representations of 3D face data. Several detailed surveys on this area can be referred to Bowyer et al. (2006) and Chang et al. (2005). Another way is to map a 2D image onto a 3D model and the 3D model with texture is used to produce a set of synthetic 2D images, with the purpose of calculating the similarity of two 2D images on the 3D model. The most representative method is the 3D morphable model proposed by Blanz & Vetter (2003) which describes the shape and texture of a human face under the variations such as poses and illuminations. The model is learned from a set of textured 3D scans of heads and all parameters are estimated by maximum a posteriori estimator. In this framework, faces are represented by model parameters of 3D shape and texture. High computational load is one of the disadvantages for this kind of methods.

For illumination variation modeling, researchers have attempted to construct images under different illumination conditions. Modeling of face images can be based on a statistical model or a physical model. For statistical modeling, no assumptions concerning the surface is needed. The principal component analysis (PCA) (Turk & Pentland, 1991) and linear discriminant analysis (LDA) (Etemad & Chellappa, 1997; Belhumeur et al., 1997) can be classified to the statistical modeling. In physical modeling, the model is based on the assumption of certain surface reflectance properties, such as Lambertian surface (Zou et al., 2007). The famous Illumination Cone, 9D linear subspace and nine point lights all belong to the illumination variation modeling.

In (Belhumeur & Kriegman 1998), an illumination model illumination cone is proposed for the first time. The authors proved that the set of n-pixel images of a convex object with a Lambertian reflectance function, under an arbitrary number of point light sources at infinity, formed a convex polyhedral cone in IR^n named as illumination cone (Belhumeur & Kriegman 1998). If there are k point light sources at infinity, the image X of the illuminated object can be modeled as

$$X = \sum_{i=1}^{k} \max(B_{S_i}, 0) \tag{1}$$

where S_i is a single light source whose magnitude $|S_i|$ represents the intensity of the source and the unit normal $S_i / |S_i|$ represents the direction, B is a convex object with surface normals and albedo. Any image in the illumination cone can be determined as a convex combination of extreme rays (images) given by

$$X_{ij} = \max(B_{S_{ij}}, 0) \tag{2}$$

where $S_{ij} = b_i \times b_j$ are rows of B with $i \neq j$. The paper proved that the dimension of this illumination cone equaled the number of distinct surface normals. Furthermore, the illumination cone is proved that it can be constructed from as few as three images. In addition, the set of n-pixel images of an object of any shape and with a more general

reflectance function, seen under all possible illumination conditions, still forms a convex cone in IR^n. The paper also extends these results to colour images. Based on the illumination cone model, Georghiades et al. (2001) presented a generative appearance-based method for recognizing human faces under variations in lighting and viewpoint. Their method exploits the illumination cone model and uses a small number of training images of each face taken with different lighting directions to reconstruct the shape and albedo of the face. As a result, this reconstruction can be used to render images of the face under novel poses and illumination conditions. The pose space is then sampled, and for each pose the corresponding illumination cone is approximated by a low-dimensional linear subspace whose basis vectors are estimated using the generative model.

Basri and Jacobs (2003) showed that a simple 9D linear subspace could capture the set of images of Lambertian objects under distant, isotropic lighting. Moreover, they proved that the 9D linear space could be directly computed from a model, as low-degree polynomial functions of its scaled surface normals. Spherical harmonics is used to represent lighting and the effects of Lambertian materials are considered as the analog of a convolution. The results help them to construct algorithms for object recognition based on linear methods as well as algorithms that use convex optimization to enforce non-negative lighting functions.

Lee et al. (2005) showed that linear superpositions of images acquired under a few directional sources are likely to be sufficient and effective for modeling the effect of illumination on human face. More specifically, the subspace obtained by taking k images of an object under several point light source directions with k typically ranging from 5 to 9, is an effective representation for recognition under a wide range of lighting conditions. Because the subspace is constructed directly from real images, the proposed methods has the following advantages, 1)potentially complex steps can be avoided such as 3D model of surface reconstruction; 2) large numbers of training images are not required to physically construct complex light conditions.

In addition to the assumption that the human face is Lambertian object, another main drawback of these illumination modeling methods is that several images are required for modeling.

3. Illumination invariant feature extraction

The purpose of these approaches is to extract facial features that are robust against illumination variations. The common representations include edge map, image intensity derivatives and Gabor-like filtering image (Adini et al., 1997). However, the recognition experiment on a face database with lighting variation indicated that none of these representations was sufficient by itself to overcome the image variation due to the change of illumination direction (Zou et al., 2007).

Recently, quotient-image-based methods are reported to be a simple and efficient solution to illumination variances and have become one active research direction. Quotient Image (QI) (Amnon & Tammy, 2001) is defined as image ratio between a test image and linear combinations of three unknown independent illumination images. The quotient image depends only on the relative surface texture information and is free of illumination. However, the performance of QI depends on the bootstrap database.

Without the bootstrap database and known lighting conditions, Wang et al. (2004a; 2004b) proposed Self-Quotient Image (SQI) to solve the illumination variation problem. The salient feature of the method is to estimate luminance using the image smoothed up by a weighted Gaussian filter. The SQI is defined as

$$Q=\frac{I}{\hat{I}} = \frac{I}{F*I} \tag{3}$$

where $\hat{I}$ is the smoothed version of I, F is the smoothing kernel, and the division is point-wise as in the original quotient image. A detailed analysis is presented to prove that the self-quotient image is robust against illumination variation under: 1) the regions without shadow and with small surface normal variation and 2) the shadow regions. However, the method is still illumination dependent in the regions without shadows but with large surface normal variation. Another advantage of the SQI is that although the analysis is carried out based on Lambertian model of single illumination point, it is also effective of other types of illumination model such as a linear combination of several illumination sources. Furthermore, the SQI does not need an alignment procedure compared to other quotient methods.

Inspired by the SQI, Chen et al. (2005; 2006) developed Logarithmic Total Variation (LTV) model and Total Variation Quotient Image (TVQI) model. The proposed methods employ the TV-L image decomposition model because the TV-L is multiscale but intensity-independent decomposition and has easy parameter selection compared to other TV-based models. The TV-L decomposes the input image *I* into two parts, naming one with large-scale features *u* and the other with small-scale features *v*, as the following formula

$$u=\mathrm{aremin}_u\int_\Omega|\nabla u|+\lambda\,|\log(I)\text{-}u|\,dx \tag{4}$$

$$v=\log(I)\text{-}u \tag{5}$$

Because the logarithm of the input image *I* can be modeled as

$$\mathrm{LogI}(x,y)=\log\rho(x,y)+\log S \tag{6}$$

Where $\log\rho(x,y)$ is the logarithm function of the albedo and $\log S$ is related to the illumination. In the LTV, we can simply estimated $\log\rho(x,y)$ and $\log S$ as

$$u\approx\log(S') \tag{7}$$

and

$$v\approx\log(\rho') \tag{8}$$

As $\log\rho(x,y)$ is the intrinsic feature which is robust against illumination variations, only small-scale features *v* will be used for recognition. The TVQI is one variation of the LTV model. Without the logarithm transform, the original image *I* is directly divided by *u* which is as the same as that in the LTV. The only difference between them is that the log operator removes noise from the image more and hence improves the performance. The model has better edge-preserving ability and simpler parameter selection and promising results have been achieved. However, it is very time consuming because it needs to find an optimal solution to decomposition step by step.

In Morphological Quotient Image (MQI) proposed by He et al. (2007), morphology operation is employed to smooth up the image and obtain luminance estimation. In the proposed method, closing operator is employed for the illumination estimation. Closing operator is a nonlinear filter, and it is carried out by Dilation operator followed by Erosion

operator. The author claims that it can remove small-scale features and retain large-scale features so that it is less destructive of the original boundary shape of foreground (He et al., 2007). Considering the effect of singularity noise, the author also proposed a method based on local maximum and minimum to remove the singularity noise before the closing operator. The method is modeled as

$$\text{Denoise}(I(i,j))=\begin{cases} \max_{(n,m)\in\text{neighbor}(i,j)}(I(m,n)), & \forall(m,n)\in\text{neighbor}(i,j),I(i,j) > I(m,n) \\ \min_{(n,m)\in\text{neighbor}(i,j)}(I(m,n)), & \forall(m,n)\in\text{neighbor}(i,j),I(i,j) < I(m,n) \\ I(i,j), & \text{otherwise} \end{cases} \quad (9)$$

where neighbor(i,j) is the 8-neighbor set of point (i, j). As most of the quotient methods, the size of closing operation is a key parameter for the performance. The paper had a detailed discussion on this parameter and presented a method named "dynamic morphological quotient image (DMQI)" based on the following formula:

$$\text{Dclose}(i,j)=\begin{cases} \text{Close}^l(i,j), & \text{Close}^l(i,j) > \alpha\text{Close}^S(i,j) \\ \text{Close}^m(i,j), & \text{Close}^S(i,j) \geq \text{Close}^l(i,j) \geq \beta\text{Close}^S(i,j) \\ \text{Close}^S(i,j), & \text{Close}^S(i,j) > \alpha\text{Close}^l(i,j) \end{cases} \quad (10)$$

where α and β are the threshold parameters while $\alpha > \beta > 1.0$ l, m, and n are three different sizes while $l>m>s>1$. The basic idea is to choose suitable size by the comparison of the closing operator of different sizes. The details can be referred to (He et al., 2007). The proposed methods obtain superior performances compared to the SQI but close to the LTV. In fact, the DMQI transfers the problem of selecting the size to the threshold. It is still a problem to determine the appropriate values. Furthermore, the DQMI will increase computational burden because of searching an appropriate size.
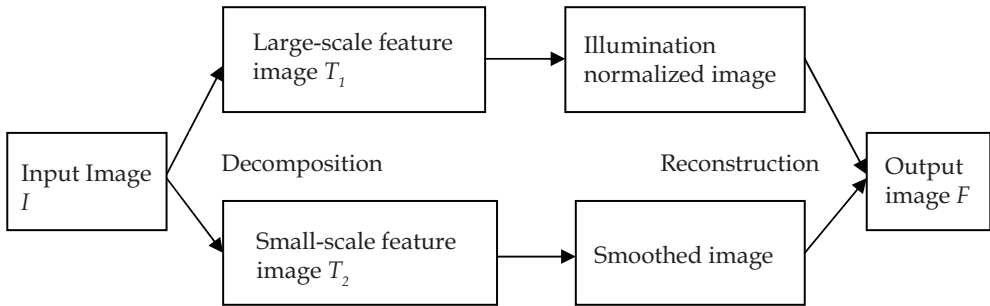

Fig. 1. Framework of the proposed method

In (Xie et al., 2008), an illumination compensation frame similar to the LTV was proposed, as shown in Fig. 1. In the frame, the image I is decomposed into two parts: small-scale and large-scale similar to the LTV. Normalization is performed in large-scale features images T_1 rather than the entire image while small-scale features image T_2 should be smoothed up. After that, an image F combining the normalized large-scale image and the smoothed small-scale image, instead of only using small-scale features in the LTV, will be used for recognition. In the smoothing, a 3×3 mask is employed for convolution if the gray value of

the center point in the convolution region is larger than an empirical threshold, and each image is convolved twice. For the normalization on large-scale image, two different methods, the DCT (Chen et al., 2006) and the non-point light quotient image (NPL-QI), are separately applied to large-scale image. The experimental results on the CMU PIE and Yale B as well as the Extended Yale B database show that the proposed framework outperforms existing methods.

In addition to the quotient-image-based methods, local binary pattern (LBP) is another attractive representation of facial features. Local descriptors of human faces have gained attentions due to their robustness against the variations of pose and expression. The LBP operator is one of the best local texture descriptors. Besides the robustness against pose and expression variations as common texture features, the LBP is also robust to monotonic gray-level variations caused by illumination variations. The details of the LBP and several important extensions of the operator will be introduced in the following section, as well as its various combinations with other techniques to handle illumination invariant face recognition under a complex environment.

4. Pre-processing and normalization

In this kind of approaches, face images under illumination variations are preprocessed so that images under normal lighting can be obtained. Further recognition will be performed based on the normalized images. Histogram Equalization (Gonzales & Woods, 1992) is the most commonly used method. By performing histogram equalization, the histogram of pixel intensities in the resulting images is flat. Therefore improved performances can be achieved after histogram equalization. Adaptive histogram equalization (Pizer & Amburn, 1987), region-based histogram equalization (Shan et al., 2003), and block-based histogram equalization (Xie & Lam, 2005) are several important variants of the HE and obtain good performances.

Recently, the methods on discarding low frequency coefficients in various transformed domains are reported to be simple and efficient solutions to tackle illumination variations and have become one active research direction. In (Chen et al., 2006), the image gray value level $f(x,y)$ is assumed to be proportional to the product of the reflectance $r(x,y)$ and the illumination $e(x,y)$ i.e.,

$$f(x,y) = r(x,y) \cdot e(x,y) \tag{11}$$

The reflectance is a stable characteristic of facial features. The goal is to recover the reflectance of faces under illumination variations. After logarithm transform, we have

$$\log f(x,y) = \log r(x,y). \log e(x,y) \tag{12}$$

In a face image, illumination usually changes slowly compared with the reflectance except some casting shadows and specularities on the face. As a result, illumination variations mainly lie in the low-frequency band. Therefore, we can remove the low frequency part to reduce illumination variation. In (Chen et al., 2006), the low frequency DCT coefficients are set to zero to eliminate illumination variations. Normally, the 2D DCT is defined as follows:

$$A(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \alpha(u)\alpha(v)f(x,y)\cos[\pi(2x+1)u/2M]\cos[\pi(2y+1)v/2N] \tag{13}$$

and the inverse transform is defined as

$$f(x,y) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \alpha(u)\alpha(v)A(x,y)\cos[\pi(2x+1)u/2M]\cos[\pi(2y+1)v/2N] \qquad (14)$$

where $\alpha(u) = \begin{cases} 1/\sqrt{M}, u = 0 \\ 2/\sqrt{M}, u = 1,2,...,M-1 \end{cases}$ and $\alpha(v) = \begin{cases} 1/\sqrt{N}, v = 0 \\ 2/\sqrt{N}, u = 1,2,...,N-1 \end{cases}$

Following (14), after setting the DCT coefficients to zero, we have

$$f'(x,y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} E(u,v) - \sum_{i=1}^{n} E(u_i, v_i) \qquad (15)$$

where $E(u,v) = \alpha(u)\alpha(v)A(u,v)\cos[\pi(2x+1)u/2M]\cos[(2y+1)v/2N]$ and n is the number of discarded DCT coefficients. Because illumination variations mainly lie in the low-frequency band, the term $E(u,v)$ can be approximately considered as the illumination $\log e(x,y)$. By comparing (12) and (15), we can see that $f'(x,y)$ obtained by (15) is the expected reflectance part in (12).

Therefore, zeroing the low frequency DCT coefficients in the logarithm domain can eliminate the effects of illumination variation and obtain the reflectance, which can be used in further illumination invariant recognition task. The method does not require multiple images to be trained and good experimental results both in the Yale Face database B and the CMU database are obtained.

Based on Chen's idea, Vishwakarma et al. (2007) proposed to rescale low-frequency DCT coefficients to lower values instead of zeroing them in Chen's. In (Vishwakarma et al., 2007), the first 20 low frequency DCT coefficients are divided by a constant 50, and the AC component is then increased by 10%. Although they presented some comparisons of figures to demonstrate the effect of the proposed method compared to the original DCT method, there is no experimental result to prove the effect of the proposed method for recognition task. Besides, it is difficult to choose the value of rescale parameter only by experiences. Perez and Castillo (2008) also proposed a similar method which applied Genetic Algorithms to search appropriate weights to rescale low-frequency DCT coefficients. Two different strategies of selecting the weights are compared. One is to choose a weight for each DCT coefficient and the other is to divide the DCT coefficients into small squares, and choose a weight for each square. The latter strategy reduces the computational cost because fewer weights are required to choose. However, the GA still takes a large computational burden and the obtained weights depend on the training figures.

As mentioned in the last section, the local binary patterns (LBP) operator is one of the best local texture descriptors. Because of its robustness against pose and expression variations as well as monotonic gray-level variations caused by illumination variations, Heydi et al. (2008) presented a new combination way of using the DCT and LBP for illumination normalization. Images are divided into blocks and the DCT is applied in each block. After that, the LBP is used to represent facial features because the LBP can represent well facial structures when variations in lights are monotonic. The experimental results demonstrate the proposed method can achieve good performances when several training samples per person are used. However, in the case where only one frontal image per person is used for training, which is common in some practical applications, the performance cannot be satisfactory especially in the cases with larger illumination variations.

Besides the DCT, discrete wavelet transform (DWT) is another common method in face recognition. There are several similarities between the DCT and the DWT: 1) They both transform the data into frequency domain; 2) As data dimension reduction methods, they are both independent of training data compared to the PCA. Because of these similarities, there are also several studies on illumination invariant recognition based on the DWT. Similar to the idea in (Chen et al., 2006), a method on discarding low frequency coefficients of the DWT instead of the DCT was proposed (Nie et al., 2008). Face images are transformed from spatial domain to logarithm domain and 2-dimension wavelet transform is calculated by the algorithm. Then coefficients of low-low subband image in n-th wavelet decomposition are discarded for face illumination compensation in logarithm domain. The experimental results prove that the proposed method outperforms the DCT and the quotient images. The kind of wavelet function and how many levels of the DWT need to carry out are the key factors for the performance of the proposed method. Different from the method in (Nie et al., 2008), Han et al. (2008) proposed that the coefficients in low-high, high-low and high-high subband images were also contributed to the effect of illumination variation besides the low-low subband images in n-th level. Based on the assumption, a homomorphic filtering is applied to separate the illumination component from the low-high, high-low, and high-high subband images in all scale levels. A high-pass Butterworth filter is used as the homomorphic filter. The proposed method obtained promising results on the Yale B and CMU PIE databases.

Different from the above methods, Gu et al. (2008) modified the face model (2) into

$$F'(m,n) = G(m,n) \cdot F(m,n) + B(m,n) \tag{16}$$

where $F'(m,n)$ is the gray level of the point under the illumination $G(m,n)$, $F(m,n)$ is the original face image, and $B(m,n)$ is the background noise which is assumed to change slowly. In (Gu et al., 2008), the background noise is estimated and eliminated by multi-level wavelet decomposition followed by spline interpolation. After that, the effect of illumination is removed by the similar multi-level wavelet decomposition in the logarithm domain. Experimental results on Yale B face database show that the proposed method achieves superior performance compared to the others. However, by comparing the results of the proposed method and the DCT, we find the result of the proposed method is worse than that of the DCT. Hence, the proposed method is not as effective as the DCT for illumination invariant recognition. The novelty of the proposed method is that the light variation in an image can be modeled as multiplicative noise and additive noise, instead of only the multiplicative term in (11), which may be instructive in modeling the face under illumination variations in future.

5. Illumination invariant face recognition under complex environment

In fact, there is seldom the case in practical applications that only illumination variations exist. Illumination variations are always coupled with pose and expression variations in practical environments. To deal with such complex environment, an ideal feature representation of human face should not only be illumination invariant, but also robust enough against pose and expression variations. Local descriptors of human faces have gained attentions due to their robustness against the variations of pose and expression. The local binary patterns (LBP) operator is one of the best local texture descriptors. The operator

has been successfully applied to face detection, face recognition and facial expression analysis. In this section, we will present a detailed introduction of the LBP and several important extensions of the operator, as well as its various combinations with other techniques to handle illumination invariant face recognition under a complex environment.

The original LBP operator was proposed for texture analysis by Ojala et al. (1996). The main idea in LBP is to compare the gray value of central point with the gray values of other points in the neighborhood, and set a binary value to each point based on the comparison. After that, a binary string is transformed to a decimal label as shown in the following equation

$$LBP_{P,R}(x,y)=\sum_{i=0}^{P-1}s(g_i-g_c)2^i \tag{17}$$

Where $LBP_{P,R}(x,y)$ is the decimal label of point (x, y), P is the number of sampling points, R is the radius of the patch, g_c is the gray level of central point (x, y), g_i is the gray level of neighborhood sampling point around central point (x, y) and

$$s(x)=\begin{cases}1, x > 0 \\ 0, x \le 0\end{cases}. \tag{18}$$

A histogram of the decimal label is calculated and can be used as a texture feature. The histogram can be defined as

$$H_i = \sum_{x,y} I\{LBP_{P,R}(x,y) = i\}, i = 0,...n-1 \tag{19}$$

Where n is the number of different labels produced by the LBP operator and

$$I\{A\} = \begin{cases}1, & A \ is \ ture \\ 0, & A \ isfalse\end{cases} \tag{20}$$

Later, the operator is extended to use circular neighborhoods. The values of pixels are bilinearly interpolated when the point is not in the center of a pixel. Furthermore, a local binary pattern is called uniform if it contains at most two bitwise transitions from 0 to 1 or vice verse when the binary string is considered circular (Ojala et al., 2002). For example, 11111111, 11110000 and 00111100 are uniform patterns. The authors noticed that the uniform patterns accounted for most of all patterns. Therefore, in many applications all non-uniform patterns are classified into a single pattern and more attentions are paid in uniform patterns.

In (Hadid et al., 2004; Ahonen et al., 2004), LBP based methods were presented to deal with face detection and recognition for the first time, and achieved good performance. After an image was divided into non-overlapped blocks, a LBP operator was applied to each block and a histogram of different labels was calculated for each block. All the histograms of each block were concatenated to an entire histogram to build a global description of the image. The salient features of the LBP description are that the description contains three levels information: the labels for the histogram contain information about the patterns on a pixel-level, the labels are summed over a small region to produce information on a regional level and an entire histogram concatenated by regional histograms presents a global description of the image (Ahonen et al., 2004). Because of the features, the LBP operator is considered as one of the best local texture descriptors. Besides the robustness against pose and expression variations as common texture features, the LBP is also robust to monotonic gray-level

variations caused by illumination variations. It is evident that different regions on the face have different discriminative capacities. Therefore, different regions are assigned different weights in similarity measurement. In (Hadid et al., 2004), with the purpose of dealing with face detection and recognition in low-resolution images, the LBP operator was applied in two-level hierarchies, regions and the whole image. In the first level, LBP histograms were extracted from the whole image as a coarse feature representation. Then, a finer histogram, extracted from smaller but overlapped regions, was used to carry out face detection and recognition further.

In (Jin et al., 2004), the author pointed out that the original LBP missed the local structure under some certain circumstance because the central point was only taken as a threshold. In order to obtain all the patterns in a small patch such as 3*3, the mean value of the patch was taken as a threshold instead of the gray level of the central point. Because the central point provides more information in most cases, a largest weight is set to it as following equation

$$LBP_{P,R}(x,y)=\sum_{i=0}^{P-1}s(g_i-m)2^i + s(g_c-m)2^P \tag{21}$$

Where

$$m=\frac{1}{P+1}(\sum_{i=0}^{P-1}g_i + g_c) \tag{22}$$

and other variables are the same as defined in (17). Although the proposed extension obtains good result, the uniform pattern argument cannot apply in the extension. In (Tan & Triggs, 2010), a local ternary pattern (LTP), another important extension of original LBP is proposed. The most important difference between the LTP and LBP is that the LTP use 3-valued codes instead 2-valued codes in the LBP. A width parameter t is proposed in the following equation instead of (18)

$$s(i)=\begin{cases} 1, i \geq t \\ 0, |i| < t \\ -1, i \leq -t \end{cases} \tag{23}$$

Because of the extension, the LTP is more discriminant and less sensitive to noise. To apply the uniform pattern in the LTP, a coding scheme that split each ternary pattern into its positive and negative halves is also proposed in (Tan & Triggs, 2010). The resulted halves can be treated as two separated LBPs and used for further recognition task.

From another point of view, the LBP can be considered as the descriptor of the first derivation information in local patch of the image. However they only reflect the orientation of local variation and could not present the velocity of local variation. In order to solve the problem, Huang et al. (2004) proposed to apply the LBP to Sobel gradient-filtered image instead of original image. Jabid et al. (2010) proposed local directional pattern which was obtained by computing the edge response value in all eight direction at each pixel position and generating a code from the relative strength magnitude. The local directional pattern is more robust against noise and non-monotonic illumination changes. Moreover, a high-order local pattern descriptor, local derivative pattern (LDP), was proposed by Zhang et al. (2010). LDP is a general framework which describes high-order local derivative direction variations instead of only the first derivation information in the LBP. Based on the experimental results, the third-

order LDP can capture more detailed discriminative information than the second-order LDP and the LBP. The details of the LDP can be referred to (Zhang et al., 2010).

The experimental results in (Ahonen et al., 2006) proved that the LBP outperformed other texture descriptors and several existing methods for face recognition under illumination variations. However, the LBP is still not robust enough against larger illumination variations in practical applications. Several other techniques are proposed to combine with the LBP to tackle face recognition under complex variations. In addition to the DCT as mentioned in the last section, Gabor wavelets are also promising candidates for combination. The LBP is good at coding fine details of facial appearance and texture, while Gabor features provide a coarse representation of face shape and appearance. In (Zhang et al., 2005), a local Gabor binary pattern histogram sequence (LGBPHS) method was proposed in which Gabor wavelet filters were used as a preprocessing stage for LBP feature extraction. The LBP was applied in different Gabor wavelets filtered image instead of the original images and only Gabor magnitude pictures were used because Gabor phase information were considered sensitive to position variations. To overcome the problem, Xie et al. (2010) proposed a novel framework to fuse LBP features of Gabor magnitude and phase images. A local Gabor XOR patterns (LGXP) was developed whose basic idea was that two phases were considered to reflect similar local features if two phases belonged to the same interval. Furthermore, the paper presented two methods to combine local patterns of Gabor magnitude and phase, feature-level and score-level. In the feature-level, two different local pattern histograms were simply concatenated into one histogram and the resulting histogram was used for measuring similarity. In the score-level, two different kinds of histograms were used to compute similarities respectively and then two similarity scores were fused together based on a weighted sum rule.

6. Comparisons and discussions

In this section, we will compare different methods and discuss their advantages and disadvantages. To evaluate the performances of different methods under varying lighting conditions without other variances, there are three popular databases, the Yale B, Extended Yale B and CMU PIE database. In the Yale Face database B, there are 64 different illumination conditions for nine poses per person (Georghiades et al., 2001). To study the performances of methods under different light directions, the images are divided into 5 subsets based on the angle between the lighting direction and the camera axis. The Extended Yale B database consists 16128 images of 28 subjects with the same condition as the Original Yale B (Lee et al., 2005). In the CMU PIE, there are altogether 68 subjects with pose, illumination and expression variations (Sim et al., 2003). Because we are concerned with the illumination variation problem, only 21 frontal face images per person under different illumination conditions are chosen, totally 1421 images.

The performances of several representative approaches of each category are shown in Table 1. The results are directly referred from their papers since they are based on the same database. It can be seen that several methods achieved satisfactory performance. However, each technique still has its own drawbacks. High computational load is one of the main disadvantages for face modeling. For illumination modeling methods, most of them require several training images. Besides, as mentioned before, physical illumination modeling generally is based on the assumption that the surface of the object is Lambertian, which is not consistent with the real human face. Regarding the illumination invariant features, most

of them even the QI-based methods are still not robust enough against larger illumination variation. The LTV and TVQI obtain the best performances among all the mentioned methods. However, they are very time consuming because they needs to find an optimal solution to decompose face images step by step. Compared to QI-based methods, the methods on discarding low frequency coefficients in various transformed domains are easier to implement and usually have lower computational costs because QI-based methods need to estimate albedo point by point. The performances of the methods discarding low frequency coefficients are also good but still not satisfactory as QI-based methods. In fact, illumination variations and facial features cannot be perfectly separated based on frequency components, because some facial features also lie in the low-frequency part as illumination variations. Therefore, some facial information will be lost when low-frequency coefficients are discarded. Furthermore, the performance of illumination normalization methods generally depends on the choices of parameters, most of which are determined only by experience and cannot be suitable for different cases.

Methods	Error Rate (%)						
	Yale B			Yale B+Extended Yale B			CMU
	Subset3	Subset4	Subset5	Subset3	Subset4	Subset5	
Without any methods	10.8	51.4	77.4	n/a	n/a	n/a	43.0
Eigenface	25.8	75.7	n/a	n/a	n/a	n/a	n/a
Eigenface w/o 1st 3	19.2	66.4	n/a	n/a	n/a	n/a	n/a
Linear Space	0	15.0	n/a	n/a	n/a	n/a	n/a
Cones-attached	0	8.6	n/a	n/a	n/a	n/a	n/a
Cones-cast	0	0	n/a	n/a	n/a	n/a	n/a
9PL	0	2.8	n/a	n/a	n/a	n/a	1.9
QI	38.1	65.9	76.7	n/a	n/a	n/a	n/a
QIR	0	9.4	21.2	n/a	n/a	n/a	n/a
SQI	0	3.6	2.1	n/a	n/a	n/a	n/a
MQI	0	0	2.1	n/a	n/a	n/a	n/a
LTV	0	0	0	20.6	23.9	21.7	0
TVQI	0	0	0	n/a	n/a	n/a	0
RLS(LOG-DCT)	n/a	n/a	n/a	12.9	12.4	15.2	0
Histogram Equalization	9.2	54.2	41.1	62.3	78.7	89.9	47.8
DCT	0	0.18	1.71	16.4	14.5	16.2	0.36
LocalDCT+LBP	10.12	15.33	17.29	n/a	n/a	n/a	n/a
DWT	0	0	0.53	n/a	n/a	n/a	n/a

Table 1. Performance comparisons of different methods under only illumination variances

In addition to the above drawbacks of each category, there are some other issues. Firstly, the experiments of most of the methods are based on aligned images whose important points are manually marked. The sensitivity of the methods to misalignment is seldom studied except some SQI-based methods. Secondly, the common experimental databases are not very promising because of the small size and limited illumination variations. For instance, the Yale B database only contains 10 subjects and the CMU PIE database contains limited illumination variations. When the database is larger and contains more illumination variations, the outstanding performances of existing methods may not be sure. The LTV achieves excellent performance in (Chen et al., 2006) when the Yale B database is used. While the Extended Yale B database containing more subjects (28) is used as the test database in (Xie et al., 2008), the performance drops significantly as shown in Table 1.

To evaluate the performances of methods under a complex environment where illumination variances coupled with other variances, FERET is the most popular database which contains 1196 subjects with expression, lighting and aging variations. In addition to the gallery set fa, there are four probe sets, fb (1195 images with expression variations), fc (194 images with illumination variations), dup I (722 images with aging variations) and dup II (234 images with larger aging variations). The performances of common methods, the LBP and several extensions and combinations of the LBP are shown in Table 2. The results are directly referred from their papers since they are based on the same database.

Methods	Error Rate (%)			
	FERET			
	fb	fc	Dup I	Dup II
Fisherface	6	27	45	69
PCA, MahCosine	15	35	56	78
Bayesian, MAP	18	63	48	68
LBP, non-weighted	7	49	39	50
LBP, weighted	3	21	34	36
Local Directional Pattern	3	18	28	31
3-order Local Derivative Pattern on gray-level images	10	13	38	40
3-order Local Derivative Pattern on Gabor feature images	2	1	20	20
LGBPHS non-weighted	6	3	32	47
LGBPHS weighted	2	3	26	29
LGXP	2	0	18	17
Fusing LGBP_mag+LGXP	1	1	6	7

Table 2. performance comparisons of different method under multiple variances

By comparing experimental results on table 2, we can easily find that the performance of most LBP-based method can outperform other methods under expression, illumination and aging variations. However, considering the performances of LBP-based methods in the probe set fc with illumination variation, we can easily find that quotient-image-based methods outperform most of LBP-based methods. But it still can be an effective and promising research direction because of its robustness against other variations such as aging and expression variations. There is seldom the case in practical applications that only illumination variations exist. Illumination variations are always coupled with other variations in practical environments. Furthermore, after the combination with Gabor features such as the weighted LGBPHS, LGXP and fusing method of LGBP and LGXP, the performances of LBP-based methods obtain significant improvements and even achieve comparable level with quotient-image-based methods under illumination variations. Hence, the LBP can be employed to carry out face recognition in a complex practical environment, combined with other recognition techniques such as Gabor wavelets.

7. Conclusions

In summary, the modeling approach is the fundamental way to handle illumination variations, but it always takes heavy computational burden and high requirement for the number of training samples. For illumination invariant feature, the quotient-image-based method is a promising direction. The LBP is also an attractive area which can tackle illumination variation coupled with other variations such as pose and expression. For normalization methods, the methods on discarding low-frequency coefficients are simple but effective way to solve the illumination variation problem. However, a more accurate model needs to be studied instead of simply discarding low-frequency coefficients. In a real complex condition, the LBP combined with other techniques such as Gabor wavelets is an easier and more promising way to deal with illumination variances coupled with other variances.

8. References

Adini, Y.; Moses, Y. & Ullman, S. (1997). Face recognition: the problem of compensating for changes in illumination direction, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, 721-732, ISSN: 0162-8828.

Ahonen, T.; Hadid, A. & Pietikainen M. (2004). Face recognition with local binary patterns, *Proceedings of 8th European Conf. Computer Vision*, pp. 469-481, ISBN: 978-3-540-21984-2, May 2004, Prague, Czech Republic.

Ahonen, T.; Hadid, A. & Pietikainen M. (2006). Face description with local binary patterns: application to face recognition, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 12, 2037-2040, ISSN: 0162-8828.

Amnon, S. & Tammy R.R. (2001). The quotient image: class-based re-rendering and recognition with varying illuminations, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 23, No. 2, 129-139, ISSN: 0162-8828.

Basri, R. & Jacobs, D.W. (2003). Lambertian reflectance and linear subspaces, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 25, No. 2, 218-233, ISSN: 0162-8828.

Belhumeur, P.N.; Hespanha, J.P. & Kriegman, D.J. (1997). Eigenfaces vs. fisherfaces: recognition using class specific linear projection, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 20, No. 7, 711-720, ISSN: 0162-8828.

Belhumeur, P. & Kriegman, D. (1998). What is the set of images of an object under all possible illumination conditions, *International Journal of Computer Vision*, Vol. 28, No. 3, 245-260, ISSN: 1573-1405 (electronic version).

Blanz, V. & Vetter, T. (2003). Face recognition based on fitting a 3D morphable model, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 9, 1063-1073, ISSN: 0162-8828.

Bowyer, K.W.; Chang, K. & Flynn, P. (2006). A survey of approaches and challenges in 3d and multi-modal 3d+2d face recognition. *Computer Vision and Image Understanding*, Vol. 101, No. 1, 1-15, ISSN: 1077-3142.

Chang, K.; Bowyer, K.W. & Flynn, P.J. (2005). An evaluation of multimodal 2d+3d face biometrics. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 4, 619-624, ISSN: 0162-8828.

Chen, W.; Er, M.J. & Wu, S. (2006). Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain, *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics,* Vol. 36, No. 2, 458-466, ISSN: 1083-4419.

Chen, T.; Yin, W.; Zhou, X.S.; Comaniciu, D. & Huang, T.S. (2005). Illumination normalization for face recognition and uneven background correction using total variation based image models, *Proceedings of the IEEE Conference on CVPR*, pp. 532-539, ISBN: 0-7695-2372-2, June 2005, San Diego, USA.

Chen, T.; Yin, W.; Zhou, X.S.; Comaniciu, D. & Huang, T.S. (2006). Total variation models for variable lighting face recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 28, No. 9, 1519-1524, ISSN: 0162-8828.

Etemad, K. & Chellappa, R. (1997). Discriminant analysis for recognition of human face images, *Journal of the Optical Society of America A*, Vol. 14, No. 8, 1724-1733, ISSN: 1520-8532 (online).

Georghiades, A.S.; Belhumeur, P.N. & Kriegman, D.J. (2001). From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 23, No. 6, 643-660, ISSN: 0162-8828.

Gonzales, R.C. & Woods, R.E. (1992). *Digital Image Processing, second ed.*, Prentice Hall, ISBN-10: 0201180758, Upper Saddle River.

Gong, W.G.; Yang, L.P.; Gu, X.H. & Li, W.H. (2008). Illumination compensation based on multi-level wavelet decomposition for face recognition, *Optics and Precision Engineering*, Vol. 16, No. 8, 1459-1464, ISSN: 1004-924X.

Hadid, A.; Pietikainen, M. & Ahonen T. (2004). A discriminative feature space for detecting and recognizing faces, *Proceedings of the IEEE Conference on CVPR*, pp. 797-804, ISBN: 0-7695-2158-4, June 2004, Washington DC, USA.

Han, H.; Shan, S.G.; Chen, X.L. & Gao, W. (2008). Illumination transfer using homomorphic wavelet filtering and its application to light-insensitive face recognition, *Proceedings of the International Conf. on Automatic Face and Gesture Recognition*, pp. 1-6, ISBN: 978-1-4244-2153-4, Sep. 2008, Amsterdam, Nethelands.

He, X.G.; Tian, J.; Wu, L.F.; Zhang, Y.Y. & Yang, X. (2007). Illumination normalization with morphological quotient image, *Journal of Software*, Vol. 18, No. 9, 2318-2325, ISSN: 1796-217X.

Heydi, M.V; Edel, G.R. & Yadira, C.M. (2008). A new combination of local appearance based methods for face recognition under varying lighting conditions, *Proceedings of the 13th Iberoamerican congress on Pattern Recognition: Progress in Pattern Recognition, Image Analysis and Applications*, pp. 535-542, ISBN: 978-3-540-85919-2, Sep. 2008, Havana, Cuba.

Huang, X.; Li, S.Z. & Wang, Y. (2004). Shape Localization based on statistical method using extended local binary pattern, *Proceedings of International Conference on Image and Graphics*, pp. 184-187, ISBN: 0-7695-2244-0, Dec. 2004, Hong Kong, China.

Jabid, T.; Kabir, M.H. & Chae, O. (2010). Local Directional Pattern (LDP) for face recognition, *Proceedings of 2010 Digest of Technical Papers International Conference on Consumer Electronics*, pp. 329-330, ISBN: 978-1-4244-4314-7, Jan. 2010, Las Vegas, NV, USA.

Jin, H.; Liu, Q.; Lu, H. & Tong, X. (2004). Face detection using improved LBP under bayesian framework, *Proceedings of International Conference on Image and Graphics*, pp. 306-309, ISBN: 0-7695-2244-0, Dec. 2004, Hong Kong, China.

Lee, K.C.; Ho, J. & Kriegman, D. (2005). Acquiring Linear Subspaces for Face Recognition under Variable Lighting, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 27, No. 5, 684-698, ISSN: 0162-8828.

Nie, X.F.; Tan, Z.F. & Guo, J. (2008). Face illumination compensation based on wavelet transform, *Optics and Precision Engineering*, Vol. 16, No. 1, 150-155, ISSN: 1004-924X.

Ojala, T.; Pietikainen, M. & Harwood, D. (1996). A comparative study of texture measures with classification based on feature distributions, *Pattern Recognition,* Vol. 29, No. 1, 51-59, ISSN: 0031-3203.

Ojala, T.; Pietikainen, M. & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. 24, No. 7, 971-987, ISSN: 0162-8828.

Perez, C.A. & Castillo, L.E. (2008). Genetic improvements in illumination compensation by the discrete cosine transform and local normalization for face recognition, *Proceedings of SPIE - The International Society for Optical Engineering, International Symposium on Optomechatronic Technologies*, pp. 1-8, ISSN:0277-786X (print), Nov. 2008, San Diego, CA, USA.

Pizer, S.M. & Amburn, E.P. (1987). Adaptive histogram equalization and its variations, *Comput. Vis. Graph. Image Process*, Vol. 39, No. 3, 355–368, ISSN: 0734-189X.

Shan, S.; Gao, W.; Cao, B. & Zhao, D. (2003). Illumination normalization for robust face recognition against varying lighting conditions, *Proceedings of the IEEE workshop on Analysis and Modelling of Faces and Gestures*, pp. 157-164, ISBN: 0-7695-2010-3, Oct. 2003, Nice.

Sim, T.; Baker S. & Bsat M. (2003). The CMU pose, illumination, and expression database, *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. 25, No. 12, 1615-1618, ISSN: 0162-8828.

Tan, X. & Triggs, B. (2010). Enhanced local texture feature sets for face recognition under difficult lighting conditions, *IEEE Trans. on Image Processing*, Vol. 19, No. 6, 1635-1650, ISSN: 1057-7149.

Turk, M. & Pentland, A. (1991). Eigenfaces for recognition, *Jounal of Cognitive Neuroscience*, Vol. 3, No. 1, 71-86, ISSN: 1530-8898 (online).

Vishwakarma, V.P.; Pandey, S. & Gupta, M.N. (2007). A novel approach for face recognition using DCT coefficients re-scaling for illumination normalization, *Proceedings of International Conference on Advanced Computing and Communications*, pp. 535-539, ISBN: 0-7695-3059-1, Dec. 2007, Guwahati, Assam.

Wang, H.; Li, S.Z.; Wang, Y. & Zhang, J. (2004a). Self quotient image for face recognition, *Proceedings of International Conference on Image Processing*, pp. 1397-1400, ISBN: 0-7803-8554-3, Oct. 2004, Singapore.

Wang, H.; Li, S.Z. & Wang, Y. (2004b). Generalized quotient image, *Proceedings of the IEEE Conference on CVPR*, pp. 498-505, ISBN: 0-7695-2158-4, June 2004, Washington DC, USA.

Xie, X. & Lam, K.-M. (2005). Face recognition under varying illumination based on a 2D face shape model, *Pattern Recognition*, Vol. 38, No. 2, 221-230, ISSN: 0031-3203.

Xie, S.; Shan, S.; Chen, X. & Chen, J. (2010). Fusing local patterns of gabor magnitude and phase for face recognition, *IEEE Trans. on Image Processing*, Vol. 19, No. 5, 1349-1361, ISSN: 1057-7149.

Xie, X.; Zheng, W.; Lai, J. & Yuen, P. (2008). Face illumination normalization on large and small scale features, *Proceedings of 26th IEEE Conference on CVPR*, pp. 1-8, ISBN: 978-1-4244-2242-5, June 2008, Anchorage, AK.

Zhang, B.; Gao, Y.; Zhao, S. & Liu, J. (2010). Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor, *IEEE Trans. on Image Processing*, Vol. 19, No. 2, 533-544, ISSN: 1057-7149.

Zhang, W.; Shan, S.; Gao, W.; Chen, X. & Zhang, H. (2005). Local gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition, *Proceedings of IEEE Int'l Conf. Computer Vision*, pp. 786-791, ISBN: 0-7695-2334-X, Oct. 2005, Beijing, China.

Zou, X.; Kittler J. & Messer K. (2007). Illumination invariant face recognition: a survey, *Proceedings of IEEE Conference on Biometrics: Theory, Applications and Systems*, pp. 1-8, ISBN: 978-1-4244-1596-0, Sept. 2007, Crystal City, VA.