High speed and robust illumination invariant face recognition techniques

Lian, Zhichao

2013

Lian, Z. (2013). High speed and robust illumination invariant face recognition techniques..
Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/54910

https://doi.org/10.32657/10356/54910

High Speed and Robust Illumination Invariant Face Recognition Techniques

Lian Zhichao

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2013

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research done by me and has not been submitted for a higher degree to any other University or Institute.

....

Date Lian Zhichao

To my parents and my wife,
for their encouragement and love.

Summary

The variation in illumination is still a challenging issue in automatic face recognition although considerable progresses have been achieved in controlled environments. It has been proven that differences between various lighting conditions are much greater than differences between individuals. Though numerous approaches have been proposed to address the issue in recent years, their performances are still not satisfying. In this thesis, some high speed approaches for face recognition under varying illuminations are developed.

Firstly, a novel illumination normalization approach with low computation complexity is proposed based on Walsh-Hadamard transform (WHT) characterized by low-complexity. By discarding an appropriate number of low-frequency coefficients in the block-wise WHT domain, the effects caused by illumination variations are removed. The proposed method is validated on the Yale B and the Extended Yale B databases. In addition, both analytical proof and experimental results demonstrate that Principal Component Analysis (PCA) and Null-space-based Linear Discriminant Analysis (NLDA) can be directly implemented in the WHT domain without the inverse WHT to reduce computational burden further.

Secondly, a novel illumination normalization approach is proposed for face recognition under varying illuminations. In the proposed approach, illumination is estimated in two steps. First of all, low frequency Discrete Cosine Transform (DCT) coefficients in the logarithm domain obtained in local areas are used to estimate illumination coarsely rather than estimating illumination in a global way. After that, a refining estimation step with mean operator is applied to estimate the illumination of every point more precisely. Experimental results demonstrate that the method is superior to other existing methods. Furthermore, a simplified version of the method

is also proposed. Both theoretical analysis and experimental results demonstrate the validity and high computational efficiency of the simplified version. Performances of the proposed methods under different values of parameters are also discussed.

In addition to the aforementioned illumination normalization methods, an illumination invariant facial feature local relation map (LRM) is explored according to local properties of human faces. A face model under varying illuminations with an additive term as noise is investigated besides the common multiplicative illumination term. High frequency coefficients of the DCT are zeroed to remove the noise. Experimental results validate the proposed face model and the assumption on noise.

Different from the common assumption, the illumination and the reflectance cannot be well approximated by only low and high frequency components respectively. In this thesis, an adaptive illumination normalization approach is proposed based on a data-driven soft-thresholding denoising technique. The proposed method models each DCT coefficient except the DC component as Generalized Gaussian distribution (GGD). More information of the reflectance in the low-frequency band is preserved while illumination variations in the high-frequency band are removed. Moreover, the key parameters are adaptively determined without any prior information.

Finally, a novel robust face descriptor named Local Line Derivative Pattern (LLDP) is presented for face recognition to deal with not only illumination variations but also expression and aging variations. High-order derivative images in two directions are obtained by convolving original images with Sobel Masks. In the LLDP, an improved binary coding function and three standards on arranging the weights are proposed, and a novel distance measuring both pixel-level and global-level information is also introduced. Promising experimental results are obtained from various face recognition databases.

Acknowledgements

I would like to take this opportunity to express my sincere appreciation to many people who have helped me during my doctoral study at Nanyang Technological University.

In the first place I would like to express my most sincere gratitude to my supervisor, Professor Er Meng Joo, for all these years of conscientious supervision, sound advices, invaluable guidance, constant encouragement and support. I would also like to express my special thanks to Dr Chen Weilong and Liu Fan for their valuable advices on my research work.

I am deeply grateful to my family for their support and unfailing blessings, My dad, Mr. Lian Zhongxiang, my mum, Ms. Wang Suqin, and my wife, Ms. Han Lu. Their endless love, support, encouragement and care are a strong motivation for me.

Last but not least, I would like to thank all my colleagues and friends for their friendship and countless help rendered during my graduate study.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Face Recognition by a robot or machine has been one of the challenging research topics in recent years. It is an active research area which crosscuts several disciplines such as image processing, computer vision, pattern recognition, neural networks and robotics. Compared with other biometric recognition technologies, such as fingerprint, voice, retina, iris and palm geometry etc., face recognition technique possesses more convenience because it is non-intrusive and requires little cooperation of participants. 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 for face recognition under uncontrolled conditions shown as follows.

- Illumination Variations

  It has been proven that differences between different lighting conditions are much greater than differences between individuals [1]. Various methods have

been developed to address the issue. Generally, they can be classified into three categories: (1) Face and illumination modelling, which constructs face models to render face images under new illumination conditions; (2) Illumination invariant feature extraction, which generates facial features robust against illumination variations; (3) Illumination normalization, which eliminates various unwanted effects caused by different illumination conditions. The most difficult problem to deal with illumination variations is non-uniform illumination where light may come from different sources and can generate various regions of shadowing on human faces.

- Pose Variations

  In some cases, a face recognition system must be able to recognize faces with different poses. The performances of face recognition systems drop significantly when large pose variations exist. Pose variations can be divided into two cases: (1) In-plane rotation which can be adjusted by a normalization process; (2) Out-of-plane rotation or rotation in depth which is totally different and not easily solved to certain extents. Various methods based on two dimensional (2D) images have been proposed to tackle the rotation issue, which can be generally divided into three classes: (1) Multiview image methods, which create a dataset with different poses for each subject; (2) Hybrid methods, where only one image per person is used as gallery but multiple images with different poses are employed for training; (3) Single-image-based methods which do not involve any training step. Employing three dimensional (3D) images is another solution for different poses because they are robust against pose variations.

- Expression Variations

In real life, human face is always full of various expressions. Facial expression variations also result in difficulties for precise face recognition. One common solution in recognition systems is to contain more training samples with different expressions to improve the performance. Morphable face model on non-rigid face matching and texture facial features are other efficient ways to deal with the problem for the case with limited training samples.

- Aging Variations

  Another challenging issue in automatic face recognition is to achieve temporal invariance. Human faces undergo complex and considerable variations with aging which affect not only the shapes of faces but the textures. Both changes will cause serious degradation of automatic face recognition systems. However, aging variations have not received as much attention as other variations such as illumination, pose and expression. Considering that the aging is a 3D process, 3D modelling is a promising way to extract the aging patterns. Local descriptors are also suited to address the issue because they are robust against texture changes.

- Real-time performance

  In addition to accuracy, performance of face recognition system is usually assessed by computational burden. Although some existing approaches can achieve good performances for face recognition under different variations, their computational cost is too high for real-time applications such as surveillance systems. Therefore, it is also an important issue for a face recognition system whether its real-time performance can be suited in practical applications.

In practical applications, illumination variation is quite common in uncontrolled environment and it is also a challenging issue in face recognition. Although there

are various approaches proposed for this issue, their performances are still not satisfactory especially in the cases where significant illumination changes exist.

1.2 Objectives

My research mainly focuses on the issue of illumination variation. The main objective of this thesis is to develop some face recognition approaches which are robust against illumination variations as well as to achieve computational efficiency. Since illumination variations are usually coupled with other variations in complex environments, approaches robust against both illumination variations and other variations should also be developed in order to improve the performance of a face recognition system in practical applications.

1.3 Contributions

The major works and contributions are summarized as follows:

- A novel illumination normalization approach based on the block-wise Walsh-Hardmard Transform (WHT) with low computation complexity for face recognition is developed. An appropriate number of low-frequency WHT coefficients are zeroed to compensate for illumination variations. Furthermore, it is shown that Principal Component Analysis (PCA) and Null-space-based Linear Discriminant Analysis (NLDA) can be directly implemented in the WHT domain to further reduce computational burden.

- A novel illumination normalization approach for face recognition under varying lighting conditions is proposed. In the proposed approach, illumination is

estimated in two steps. First of all, low frequency Discrete Cosine Transform (DCT) coefficients in the logarithm domain obtained in local areas are used to estimate illuminations coarsely rather than in a global way. After that, a refining estimation step with mean operator is applied to estimate the illumination of each point more precisely. Furthermore, a simplified version of the method is also proposed. Both theoretical analysis and experimental results demonstrate the validity and high computational efficiency of the simplified version.

- In addition to the above illumination normalization methods, an illumination invariant facial feature local relation map (LRM) is explored according to local properties of human faces. The proposed feature is not only effective but also simple which has been proven by experimental results and analysis. Different from most of existing methods, a face model under varying illuminations is investigated where an additive term is modelled as noise in addition to the common multiplicative illumination term. High frequency coefficients of DCT are used to estimate the noise. Experimental results validate the proposed face model and the assumption on noise.

- An adaptive illumination normalization approach based on denoising technique is proposed. The proposed method models each DCT coefficient except the DC component as Generalized Gaussian distribution (GGD). A data-driven soft-thresholding denoising technique is employed to remove illuminations from facial images. Based on the information provided by gallery images, the proposed denoising technique can adaptively choose appropriate thresholds for different DCT coefficients. The proposed method preserves more information of the reflectance in the low-frequency band instead of discarding them, and removes the effects caused by illumination variations in the high-frequency

band. Moreover, the denoising method does not need any prior information and none of the parameters is determined by using an empirical way.

- To deal with complex real environments, a good face recognition approach should be not only illumination invariant, but also robust enough against other variations such as expression and aging variations. In this thesis, a novel robust face descriptor named Local Line Derivative Pattern (LLDP) is proposed for face recognition. High-order derivative images in two directions are obtained by convolving original images with Sobel Masks. In the LLDP, the binary coding function is improved and three standards on arranging the weights are proposed. A novel distance measuring both pixel-level information and global-level information is also proposed to improve the performances of the LBP and its variants under varying illuminations. Experimental results demonstrate that the LLDP has good robustness against expression, illumination and aging variations.

1.4 Organization of this Thesis

This thesis consists of eight chapters. In Chapter 2, broad literatures on face recognition under illumination variations are reviewed. In Chapter 3, a WHT-based illumination normalization algorithm is introduced. In Chapter 4, another novel illumination normalization based on the DCT is presented where illumination is estimated in two steps. Chapter 5 presents a novel facial feature that are robust against illumination variations. A face model under varying illuminations with an additive term as noise is also investigated in this chapter. In Chapter 6, an adaptive illumination normalization approach based on denoising technique is developed. The proposed method preserves more information of the reflectance in the low-frequency

band instead of discarding them, and removes the effects caused by illumination variations in the high-frequency band. In Chapter 7, a novel robust face descriptor for face recognition named LLDP is introduced. High-order local information are extracted to achieve good robustness against expression, illumination and aging variations. In Chapter 8, conclusion and recommendations for future work are presented.

Chapter 2

Literature Review

One of the challenging issues in automatic face recognition is to achieve illumination invariance. It has been proven that differences caused by different lighting conditions are much greater than differences between individuals [1]. Various methods have been developed to address the issue. Generally, they can be classified into three categories: (1) Face and illumination modelling, which constructs face models to render face images under new illumination conditions; (2) Illumination invariant feature extraction, which generates facial features robust against illumination variations; (3) Illumination normalization, which eliminates various unwanted effects caused by different illumination conditions. 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 changes such as expression, pose and aging variations, a good feature representation of human face should not only be illumination invariant, but also robust enough against other 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 also carried out, as well as its various combinations with other techniques to handle illumination invariant face recognition under a complex environment. Most of the work has been presented in [2].

2.1 Face and Illumination Modelling

In this category of methods, researchers have attempted to construct images under different illumination conditions according to a statistical model or a physical model. The principal component analysis (PCA) [3–5] and linear discriminant analysis (LDA) [6, 7] are two of the most representative statistical modelling methods where no assumptions concerning the surface are needed. Modelling based on a physical model needs to make some assumptions on certain surface reflectance properties such as Lambertian surface [8]. The famous Illumination Cone [9–11], 9D linear subspace [12] and nine point lights [13] all belong to the illumination modelling based on a physical model.

Belhumeur and Kriegman [9] proposed an illumination model illumination cone where 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, is proven to construct a convex polyhedral cone in IR^n named as illumination cone [9]. If there are k point light sources, the image X of the illuminated object can be modelled as

$$X = \sum_{i=1}^{k} \max(Bs_i, 0) \tag{2.1}$$

where s_i is a single light 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. We can use a convex combination of images to model an image

in the illumination cone as

$$X_{ij} = \max(Bs_{ij}, 0) \tag{2.2}$$

where $s_{ij} = b_i \times b_j$ are rows of B with $i \neq j$. The dimension of illumination cone has been proven to equal the number of distinct surface normals by [9]. Furthermore, the illumination cone has been proven in [9] that it can be modelled from as few as three images. In addition, the images of an object with a more general reflectance function instead of Lambertian reflectance function, illuminated by all possible lighting 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. [11] present a generative appearance-based face recognition method to deal with illumination and viewpoint variations. Their method uses a small number of training images of each subject illuminated by different lighting conditions to reconstruct the shape and albedo of the face. For each pose, a corresponding illumination cone is constructed by a low-dimensional linear subspace using the proposed model.

In [12], Basri and Jacobs prove that the set of images of a subject with Lambertian surface under distant, isotropic lighting can be represented by a nine dimensional (9D) linear subspace. Moreover, the 9D linear space is proven that it can be directly calculated as low-degree polynomial functions of scaled surface normals. Spherical harmonics is used to represent lighting and the effects of Lambertian materials are considered as a convolution. The results can be used to develop linear-methods-based approaches for object recognition as well as algorithms that use convex optimization to solve non-negative lighting functions.

In [13], Lee et al. show that the subspace calculated by k images of an object taken under several point light sources is an effective representation for recognition under varying illuminations, where k is typically selected from 5 to 9. 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 construct the subspace of images under complex light conditions.

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

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 [14–17]. The representative method is the 3D morphable model proposed in [17] which represents not only face shape but also facial texture under multiple variations such as poses and illuminations. The model is learned based on couples of textured 3D scans of heads and all the parameters are estimated by maximizing a posteriori estimator. In this approach, 3D shapes and textures of faces are represented by the parameters in the constructed model. High computational load is one of the disadvantages for this kind of methods.

2.2 Illumination Invariant Feature Extraction

In these approaches, the main purpose is to obtain robust facial features against varying illuminations. The common representations include edge map, Gabor-like filtering image and image intensity derivatives [1]. However, the recognition experiment in [8] indicated that none of these representations is robust enough against illumination direction variations.

Recently, quotient-image-based methods [18–27] are reported to be a simple and

efficient solution to illumination variances and have become one active research direction. Quotient Image (QI) [18] is defined as image ratio between a input image and linear combinations of three images under unknown independent illuminations. The quotient image is free of illumination and only related to the surface texture information. However, the performance of QI depends on the bootstrap database.

Shan et al. [28] proposed a novel approach naming Quotient Illumination Relighting (QIR) according to an assumption that faces are an ideal class of objects. But, the method requires the lighting condition of the image to be known as a priori.

Without the bootstrap database and known lighting conditions, Wang et al. [21, 22] proposed a novel faical feature Self-Quotient Image (SQI) for solving the problem of illumination variations. 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

$$Q = \frac{I}{\widehat{I}} = \frac{I}{F * I} \tag{2.3}$$

where I is the input image, F is the convolution kernel, $\widehat{I}$ is the convolution result, and the division is defined a point-wise division. The authors present a detailed analysis to prove that the self-quotient image is free of illumination variation under: 1) the regions with small surface normal variation and without shadow and 2) the shadow regions. However, the image is still illumination dependent in the regions without shadows but with large surface normal changes. Another advantage of the SQI is that although the analysis is carried out according to Lambertian model, it is also effective for other illumination models such as a linear combination of several illumination sources. Furthermore, the SQI does not need an alignment procedure compared to other quotient methods.

In Morphological Quotient Image (MQI) proposed by He et al. [25], morphology

operation is applied to smooth up images to estimate illumination. 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 [25]. Considering the effect of singularity noise, the author also proposed a method based on local maximum or minimum to remove the singularity noise before the closing operation. The method is modelled as

$$Denoise(I(i,j)) = \begin{cases} \max_{(m,n)}(I(m,n)), & \forall(m,n) \in neighbor(i,j), I(i,j) > I(m,n) \\ \min_{(m,n)}(I(m,n)), & \forall(m,n) \in neighbor(i,j), I(i,j) < I(m,n) \\ I(i,j), & otherwise \end{cases} \tag{2.4}$$

where neighbour(i,j) is the 8-neighbour 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 has a detailed discussion in this parameter and presented a method named dynamic morphological quotient image (DMQI), based on the following formula

$$Dclose(I(i,j)) = \begin{cases} Close^l(i,j), & Close^l(i,j) > \alpha Close^s(i,j) \\ Close^m(i,j), & \alpha Close^s(i,j) \geq Close^l(i,j) \geq \beta Close^s(i,j) \\ Close^s(i,j), & Close^s(i,j) > Close^l(i,j) \end{cases} \tag{2.5}$$

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 [25]. The proposed methods obtain superior performances compared to the SQI but close to the LTV. In fact, the DMQI transfers the problem of selection of the size to the selection of the threshold. It is still a problem to determine the

appropriate values. Furthermore, the DQMI will increase the computational burden because the search of appropriate size.

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. As one of the best local texture descriptors, the LBP operator is robust against not only pose and expression variations, but 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.

2.3 Preprocessing and Normalization

The goal of preprocessing and normalization approaches is to preprocess face images taken under illumination variations so that images under normal lighting can be obtained. Further recognition will be performed based on the normalized images. Histogram Equalization [29]is the most commonly used method.After histogram equalization, pixel intensities in the resulting images are better distributed on the histogram and the contrast of images are increased. Therefore improved performances can be achieved after histogram equalization. Adaptive histogram equalization (AHE) [30], region-based histogram equalization (RHE) [28] and block-based histogram equalization (BHE) [31] are three 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 illumina-

tion variations and have become one active research direction [32–39]. In Chen et al. [32], the image gray value level f(x,y) is modelled as

$$f(x,y) = r(x,y) \cdot e(x,y) \tag{2.6}$$

where $e(x,y)$ is the illumination, and $r(x,y)$ is the reflectance. The purpose is to obtain the reflectance $r(x,y)$ which is a stable facial feature free of illumination, from faces $f(x,y)$ taken under varying illuminations. After logarithm transform, we have

$$\log f(x,y) = \log r(x,y) \cdot \log e(x,y) \tag{2.7}$$

In addition to some casting shadows and specularities on faces, the changes of illumination are usually slow compared to the reflectance [32]. Therefore, illumination variations can be assumed to depend on the low-frequency band. Therefore, we can remove the low frequency part to reduce illumination variation. In [32], illumination variations are eliminated by zeroing the low frequency DCT coefficients. Normally, we define the 2D DCT as follows:

$$A(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \alpha(u)\alpha(v) f(x,y) \cos\left[\pi(2x+1)u/2M\right] \cos\left[\pi(2y+1)v/2N\right] \tag{2.8}$$

and the inverse transform is defined as

$$f(x,y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \alpha(u)\alpha(v) A(u,v) \cos\left[\pi(2x+1)u/2M\right] \cos\left[\pi(2y+1)v/2N\right] \tag{2.9}$$

where

$$\alpha(u) = \begin{cases} 1/\sqrt{M}, & u = 0 \\ 2/\sqrt{M}, & u = 1, 2, \ldots, M-1 \end{cases} \tag{2.10}$$

$$\alpha(v) = \begin{cases} 1/\sqrt{N}, & v = 0 \\ 2/\sqrt{N}, & v = 1, 2, \ldots, N-1 \end{cases} \tag{2.11}$$

Following (2.9), after setting the DCT coefficients to zero, we have

$$f^{'}(x,y) = \sum_{u=0}^{M-1}\sum_{v=0}^{N-1} E(u,v) - \sum_{i=1}^{n} E(u_i, v_i) \tag{2.12}$$

where

$$E(u,v) = \alpha(u)\alpha(v)A(u,v)\cos\left[\pi(2x+1)u/2M\right]\cos\left[\pi(2y+1)v/2N\right] \tag{2.13}$$

and n is the number of discarded DCT coefficients. Because illumination variations depend on the low-frequency band, the term E(u,v) can be approximately considered as the illumination $\log e(x,y)$. By comparing (2.7) and (2.12), we can see that $f^{'}(x,y)$ obtained by (2.12) is the expected reflectance part in (2.7).

Therefore, zeroing the low frequency coefficients in the logarithm domain can eliminate the effects caused by 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 CMU PIE database and Yale Face database B are obtained.

Based on Chen's idea, Vishwakarma et al. [33] proposed to rescale low-frequency DCT coefficients (except the first) to lower values instead of zeroing them in Chen's. In [33], 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 by 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 [34] also proposed a similar method which applied Genetic Algorithms to search appropriate weights to rescale low-frequency DCT coefficients. The author tests two different strategies to select the weights. 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 last section, the local binary patterns (LBP) operator is one of the best local texture descriptors which is robust against not only pose and expression variations, but also monotonic gray-level changess resutled from varying illuminations. Because of these advantages of the LBP, Heydi et al. [35] presented a new approach combining the DCT and the LBP for illumination normalization. The method divided images into blocks and applied Chen's method [32] in each block. After that, the LBP is used to represent facial features. The experimental results demonstrate the proposed method can achieve good performances when several training samples per person are used. However, in the case that 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 where larger illumination variations exist.

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 the 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 [36–39]. Similar to the idea in [32], a method on discarding low frequency coefficients of the DWT instead of the DCT are proposed by Nie et al. [37]. The 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 [37], Han et al. [39] proposed that the coefficients in low-high, high-low, and high-high subband images are 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 for separating 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 CMU PIE database and Yale Face database B.

Different from the above methods, Gong et al. [38] modified a face model into

$$F^{'}(m,n) = G(m,n) \cdot F(m,n) + B(m,n) \tag{2.14}$$

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 [38], 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 recognition under illumination variations. One of the novelties 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 (2.6), which may be instructive in modeling the face under illumination variations in future.

Chen et al. [23, 24] 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 [24]. The TV-L decomposes the input image I into two parts, large-scale features u and small-scale features v, as the following formulae

$$u = \arg\min_{u} \int_{\Omega} |\nabla u| + \lambda \left|\log I - u\right| \mathrm{d}x \tag{2.15}$$

$$v = \log I - u \tag{2.16}$$

Because the logarithm of the input image I can be modelled as

$$\log I(x, y) = \log \rho(x, y) + \log S \tag{2.17}$$

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

$$u \approx \log s \tag{2.18}$$

$$v \approx \log \rho \tag{2.19}$$

As $\log \rho$ 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 can elminate the noise from input images more and hence improves the performance. The model has an ability of preserving edge better 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.

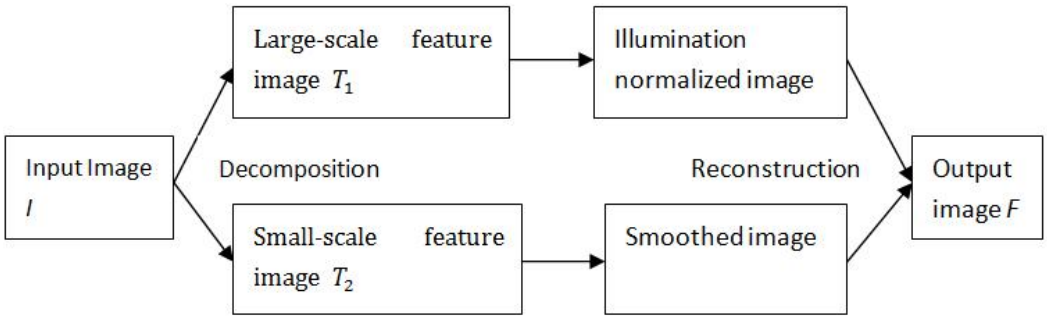


Figure 2.1: Framework of the proposed method

In Xie et al. [26, 27], an illumination compensation frame similar to the LTV is proposed, as shown in Figure 2.1. 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 centre point in convolution region is greater than an empirical threshold. Please note here that the smoothing may not improve the performance but increase the quality of resulting images. For the normalization on large-scale image, two different methods, the DCT [32] and the non-point light quotient image (NPL-QI), are separately applied to large-scale image. The experimental results on the CMU PIE and the Yale B as well as the Extended Yale B database show that the proposed framework outperforms existing methods.

2.4 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 recognition [40–45] and authentication [46], as well as face detection [47,48] and facial expression analysis [49–51]. 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 [52]. The main idea in LBP is to compare the gray value of central point with the gray values of other points in the neighbourhood, and set a binary value to each point based on the comparison [40]. After that, a decimal label can be obtained from the binary string as

$$LBP_{P,R}(x,y) = \sum_{i=0}^{P-1} s(g_i - g_c)2^i \tag{2.20}$$

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

$$s(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0. \end{cases} \tag{2.21}$$

After decimal labels are obtained, a histogram of the labels can be calculated as

$$H_i = \sum_{x,y} I\{LBP_{P,R}(x,y) = i\}, i = 0, ..., n-1 \tag{2.22}$$

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

$$I(A) = \begin{cases} 1, & A\, is\, true \\ 0, & A\, is\, false. \end{cases} \tag{2.23}$$

Later, a circular neighbourhood is proposed to extend the operator. The values of pixels were bilinearly interpolated when the point is not in the centre of a pixel. Furthermore, Ojala et al. [53] define a local binary pattern as uniform if it contains at most two bitwise transitions from 0 to 1 or vice verse when the binary string is considered circular. 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 [40, 47], LBP based methods were presented to deal with face detection and recognition for the first time, and achieved good performance. Images were firstly divided into non-overlapped blocks. After that, a histogram of each block was calculated based on the LBP operator. Finally, we concatenate all the histograms to one histogram as facial features for further recognition task. The salient features of the LBP are that the description can provide three levels information: (1) a pixel-level information based on the labels; (2) a regional level information from the histogram calculated over a small region; (3) a global information from the entire

histogram concatenated by regional histograms [40]. Based on the features, the LBP is insensitive to monotonic gray-level variations resulted from illumination variations in addition to its robustness against variations in pose and expression. It is evident that different regions on the face have different discriminative capacities. Therefore, different regions are assigned different weights in similarity measurement. In [47], 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 are extracted from the whole image as a coarse feature representation. Then, a finer histogram, extracted from smaller but overlapped regions, can be used to carry out face detection and recognition further.

In [48], 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

$$LBP_{P,R}(x,y) = \sum_{i=0}^{P-1} s(g_i - m)2^i + s(g_c - m)2^P \tag{2.24}$$

where

$$m = \frac{1}{P+1}(\sum_{i=0}^{P-1} g_i + g_c) \tag{2.25}$$

and other variables are the same as defined in (2.20). Although the proposed extension obtains good result, the uniform pattern argument cannot apply in the extension. In [42, 54], 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 as the following equation instead of (2.21)

$$s(i) = \begin{cases} 1, & i \geq t \\ 0, & |i| < t \\ -1, & i \leq -t. \end{cases} \tag{2.26}$$

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 [42]. The resulted halves can be treated as two separated LBPs and used for further recognition task.

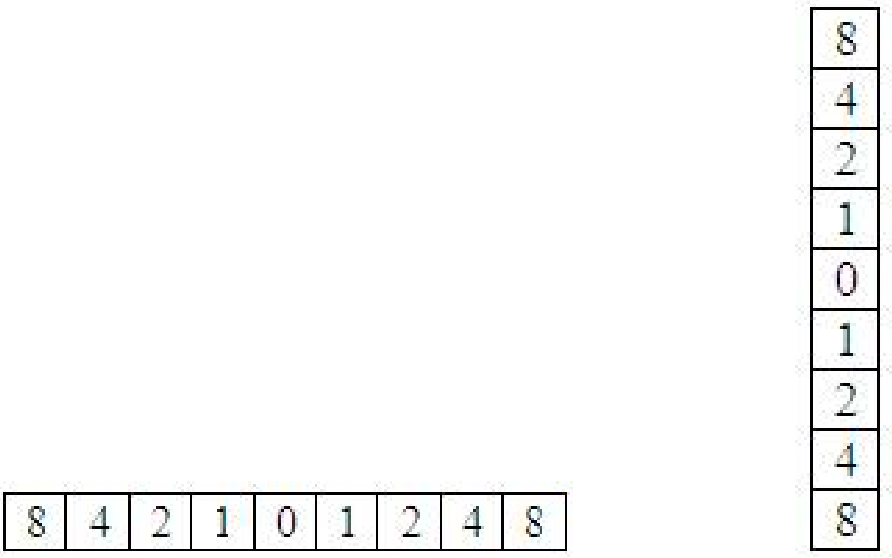
In [43], a Local Line Binary Pattern (LLBP) is proposed.The main idea of LLBP is similar to the LBP but there are some salient features as follows: 1) its neighbourhood shape is a straight line instead of a square shape in the LBP; 2) The binary weights are different from the weights in the LBP as shown in Figure 2.2. From the figure, we can see that pixels close to the centre pixel have higher weights than those away from the centre pixel, and the weights in right side are mirror to the weights in the left side. The LLBP firstly extracts the binary labels along with horizontal and vertical direction respectively. After that, a magnitude is calculated based on both horizontal and vertical labels, which represents the change in image intensity such as edges and corners. The labels and magnitude can be calculated as

$$LLBP_h(N,c) = \sum_{n=1}^{c-1} s(h_n - h_c) \cdot 2^{c-n-1} + \sum_{n=c+1}^{N} s(h_n - h_c) \cdot 2^{(n-c-1)} \tag{2.27}$$

$$LLBP_v(N,c) = \sum_{n=1}^{c-1} s(v_n - v_c) \cdot 2^{c-n-1} + \sum_{n=c+1}^{N} s(v_n - v_c) \cdot 2^{(n-c-1)} \tag{2.28}$$

$$LLBP_m = \sqrt{LLBP_h^2 + LLBP_v^2} \tag{2.29}$$

where LLBPh, LLBPv, LLBPm are LLBP on horizontal direction, vertical direction, and its magnitude respectively, N is the length of the line in pixel, $c = \frac{N}{2}$ is the



(a) (b)

Figure 2.2: Horizontal and vertical line weights of LLBP operator with 9 pixels

position of the centre pixel, h_c and v_c are the pixel on the horizontal line and the pixel on the vertical line respectively, h_n is the pixel along with the horizontal line and v_n is the pixel along with the vertical line, and s function is same as in (2.21).

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 can not describe the velocity of local change. In order to address the issue, Huang et al. [55] proposed to apply the LBP to Sobel gradient-filtered image instead of original image. Jabid et al. [56, 57] proposed local directional pattern which is 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 named local derivative pattern (LDP), is proposed by Zhang et al. [44,58]. The LDP describes high-order local derivative direction variations instead of only

the first derivation information in the LBP. The third-order LDP extracts more discriminative information and obtains better experimental results compared to the second-order LDP and the LBP. The details of the LDP can be referred to [44, 58].

The experimental results in [41] prove that the LBP outperforms other texture descriptors and several existing method 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 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 [59], a local Gabor binary pattern histogram sequence (LGBPHS) method is proposed in which Gabor wavelet filters are used as a preprocessing stage for LBP feature extraction. The LBP is applied in different Gabor wavelets filtered image instead of the original images and only Gabor magnitude pictures are used because Gabor phase information are considered sensitive to position variations. To overcome the problem, Xie et al. [60] proposed a novel framework to fuse LBP features of Gabor magnitude and phase images. A local Gabor XOR patterns (LGXP) is developed whose basic idea is that two phases are considered to reflect similar local feature if two phases belong to the same interval. Furthermore, the paper presents 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 are simply concatenated into one histogram and the resulting histogram is used for measuring similarity. In the score-level, two different kinds of histograms are used to compute similarities respectively and then two similarity scores are fused together based on a weighted sum rule.

2.5 Comparison and Summary

The performances of several representative approaches for each category are shown in Table 2.1. It can be seen that several methods achieved satisfactory performances. 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 variations. The LTV and TVQI obtain the best performance among all the mentioned methods. However, they are very time consuming because they needs to find an optimal solution to decomposition step by step. The methods on discarding low frequency coefficients in various transformed domains are simple and efficient solutions to tackle illumination variations. But in fact, facial features and illumination variations cannot be perfectly separated based on frequency bands, 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 performances of illumination normalization methods generally depend on the choices of parameters, most of which are determined only by experience and cannot be suitable for different cases.

There are some other issues in addition to the above disadvantages of each group. 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

Table 2.1: Recognition error rates of different methods in different databases

Methods	Yale B			Extended Yale B			CMU
	Subset			Subset			
	3	4	5	3	4	5	
Without any methods [32]	10.8	51.4	77.4	n/a	n/a	n/a	43.0
Histogram Equalization [32]	9.2	54.2	41.1	62.3	78.7	89.9	47.8
Eigenface [11]	25.8	75.7	n/a	n/a	n/a	n/a	n/a
Eigenface w/o 1st 3 [11]	19.2	66.4	n/a	n/a	n/a	n/a	n/a
Linear Space [11]	0	15.0	n/a	n/a	n/a	n/a	n/a
Cones-attached [11]	0	8.6	n/a	n/a	n/a	n/a	n/a
Cones-cast [11]	0	0	n/a	n/a	n/a	n/a	n/a
9PL [13]	0	2.8	n/a	n/a	n/a	n/a	1.9
QI [24]	38.1	65.9	76.7	n/a	n/a	n/a	n/a
QIR [24]	0	9.4	21.2	n/a	n/a	n/a	n/a
SQI [24]	0	3.6	2.1	n/a	n/a	n/a	n/a
MQI [25]	0	0	2.1	n/a	n/a	n/a	n/a
DCT [26, 32]	0	0.18	1.71	16.4	14.5	16.2	0.36
LocalDCT+LBP [35]	10.12	15.33	17.29	n/a	n/a	n/a	n/a
LTV [24, 26]	0	0	0	20.6	23.9	21.7	0
TVQI [24]	0	0	0	n/a	n/a	n/a	0
RLS(LOG-DCT) [26]	n/a	n/a	n/a	12.9	12.4	15.2	0

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 [24] when the Yale B database is used. While the Extended Yale B database containing more subjects (28) is used as the test database in [26], the performance drops significantly as shown in Table 2.1.

To evaluate the performances of methods under a complex environment where

Table 2.2: Performance comparisons of different methods under multiple variances

Methods	FERET Error Rates (%)			
	fb	fc	dup I	dup II
Fisherface [41]	6	27	45	69
PCA, MahCosine [41]	15	35	56	78
Bayesian, MAP [41]	18	63	48	68
LBP, non-weighted [41]	7	49	39	50
LBP, weighted [41]	3	21	34	36
Local Directional Pattern [56]	3	18	28	31
3-order Local Derivative Pattern on gray-level images [44]	10	13	38	40
3-order Local Derivative Pattern on Gabor feature images [44]	2	1	20	20
LGBPHS non-weighted [59]	6	3	32	47
LGBPHS weighted [59]	2	3	26	29
LGXP [60]	2	0	18	17
Fusing $LGBP_{mag}$+LGXP [60]	1	1	6	7

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 2.2.

By comparing experimental results on Table 2.2, 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, with other recognition techniques such as Gabor wavelets.

In summary, the modelling 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 QI-based method is a promising direction. The LBP is also an attractive area which can tackle illumination variations as well as pose and expression variations. For normalization methods, the methods on discarding low-frequency coefficients are simple but effective way to address the issue of illumination variations. 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.

Chapter 3

Illumination Normalization for Face Recognition in a Transformed Domain

3.1 Introduction

Walsh-Hadamard transform (WHT) is a generalized class of Fourier transforms [29]. It is very fast because it does not require any multiplication in the transformation calculation as it only contains ± 1 values. Because of its advantage of low-complexity, it has been widely used in signal processing [61–63], data compression algorithms [64–67] and watermarking [68, 69], such as HD Photo and MPEG-4AVC. However, in face recognition, it only has limited applications [70–73]. In [70], a face image was decomposed into a set of binary images, and 2D truncated WHT was applied to the binary images to reduce data dimensionality and to produce translation-invariant features. Guo et al. [71] proposed a novel face feature representation based on multi-scale 2D WHT in face detection. In [72], Hassan et al. proposed a facial feature

extraction approach based on the WHT and compared the proposed method with the DCT and the principal component analysis (PCA). In [73], Macros et al. also proposed a low-complexity face recognition system based on the WHT, where only the WHT coefficients of low-frequency parts were used for recognition. The method achieved similar results to the DCT with reduced computational cost. Only the low-frequency WHT coefficients are used for recognition.

In this chapter, the research mainly focus on the study of illumination normalization in the WHT domain because of its low computational requirement. The main difference between the proposed method and other mentioned WHT-based methods in [70–73] is that the low-frequency parts of WHT coefficients are discarded to eliminate illumination variations in our method. It is the first time that the WHT is applied in illumination invariant face recognition. Furthermore, the principal component analysis (PCA) and the null-space based linear discriminant analysis (NLDA), an important variant of linear discriminant analysis (LDA), are proven that they can be directly implemented in the WHT domain without the inverse transform to further reduce the computational burden. A brief version of the proposed method has been introduced in [74] with some preliminary results and detailed results have been presented in [75].

3.2 Illumination Normalization in the Logarithm Block-wise WHT Domain

3.2.1 Face and Illumination Model

In a simple situation, the image gray level f(x,y) can be regarded to be proportional to the product of the reflectance r(x,y) and the illumination e(x,y), i.e.,

$$f(x, y) = r(x, y) \cdot e(x, y) \tag{3.1}$$

As shown in [2], illumination compensation can be implemented in the logarithm domain. After logarithm transform, the model changes to

$$\log f(x, y) = \log r(x, y) + \log e(x, y) \tag{3.2}$$

The first part is a stable characteristic of facial features and the second part is related to illumination. The purpose of illumination normalization is to obtain $\log r(x, y)$ from $\log f(x, y)$.

In a face image, illumination usually changes slowly compared with the reflectance. Therefore, in [32], low-frequency DCT coefficients in the logarithm domain are discarded to eliminate effects of illumination variations, because illumination variations mainly lie in the low-frequency band. From another point of view, the method can also be considered to estimate the illumination of each point $\log e(x, y)$ based on the low-frequency part of DCT coefficients obtained in a global way.

In this chapter, the proposed method uses low-frequency WHT coefficients obtained from local blocks to estimate the illumination instead of using the DCT coefficients obtained from the entire image.

3.2.2 Walsh Hadamard Transform

The Walsh-Hadamard transform can be regarded as a generalized class of Discrete Fourier transforms and is in fact equivalent to a multidimensional DFT of size $2 \times 2 \times \cdots \times 2 \times 2$. A WHT matrix H_n is a $N \times N$ matrix of ± 1 values where $N = 2^n$.

Recursively, 1×1 Hadamard transform matrix H_0 is defined as the identity $H_0 = 1$ and then H_n for $n > 0$ is defined as follows:

$$H_n = \frac{1}{\sqrt{2}} \begin{pmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{pmatrix} \tag{3.3}$$

For example, for $n = 3$ the Hadamard transform matrix is given by

$$H_3 = \frac{1}{\sqrt{8}} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix} \tag{3.4}$$

The WHT for an image F of $2^n \times 2^n$ pixels is obtained by

$$T = H_n F H_n \tag{3.5}$$

and the inverse transform is

$$F = H_n T H_n \tag{3.6}$$

In this work, images are divided into 8×8 blocks and then the WHT is implemented on these blocks.

3.2.3 Illumination Normalization

As mentioned before, the WHT is a generalized class of Discrete Fourier transforms. Therefore, it can be used to transform an image from the spatial domain to the frequency domain. Because the transform matrix only contains ± 1 values, it does not require any multiplication in the transform. As a consequence, the transformation has a higher efficiency compared to the DCT. Since illumination variation mainly lies in the low-frequency band, the illumination $\log e(x, y)$ can be estimated using low-frequency DCT coefficients. From (3.2), subtracting $\log e(x, y)$ from original images in the logarithm domain will result in illumination normalized images. Therefore, low-frequency coefficients can be set to zero, which is equivalent to subtracting $\log e(x, y)$ from the original image in the logarithm domain with the purpose of eliminating the effects of illumination variations. The details are as follows:

1) Transform an image into the logarithm domain.

2) Divide the image into 8×8 blocks.

3) Apply the WHT to each block.

4) Set the values of k dimension of low-frequency WHT coefficients to zeros.

5) Apply the inverse WHT to each block.

The obtained images can be used in the succeeding recognition task. Different from [32, 73], here images are divided into blocks and then the WHT is applied to blocks instead of the entire images directly. The reasons are: 1) The WHT matrix must be $2^n \times 2^n$. Hence, if the WHT is applied directly to the entire images, the size of the images must be the same. Otherwise, the images must be rescaled. The process may influence the succeeding recognition performances, as well as taking extra computational costs. Dividing images into blocks will not take extra computational

Table 3.1: Computational complexity comparison

Transforms	Computational complexity
DCT	$N^2 \log_2 N$ multiplications
WHT	$N^2 \log_2 N$ additions
block-wise WHT (block Size=$M << N$)	$N^2 \log_2 M$ additions

cost (the computation complexity is even lower as shown in Table 3.1). Furthermore, the requirement of size is lower which only require that the sizes of width and height are multiplies of the block size. 2) Considering illumination variations affect less in localities than those in the entire image, we propose to apply the WHT to blocks. As a comparison, the WHT is also implemented in entire images. The comparison results are discussed in the following section.

3.3 PCA and NLDA in WHT domain

Here, the PCA and the NLDA, an important variant of linear discriminant analysis (LDA), are proven that they can be implemented directly in the WHT domain. Therefore, the inverse WHT can be skipped and the computational burden can be reduced further. First of all, the result of NLDA is proven to be invariant for the orthonormal transformed data. Furthermore, it will be proved that the 2D WHT is an orthonormal transformation as well as the 2D block-wise WHT. Combining the above proofs with the conclusion in [76] that the projection results of the PCA are invariant for the orthonormally transformed data, we can conclude that the PCA and the NLDA can be implemented directly in the 2D WHT domain and the block-wise 2D WHT domain.

Firstly, we prove that the result of the NLDA is invariant for the orthonormal transformed data. The key idea of the null-space based linear discriminant analysis

methods [77–79] is that the null space of the within-class scatter matrix contains most discriminative information. In [79], Liu et al. proposed the NLDA which is simpler than all other null space approaches. To save the computational cost, the NLDA removes the null space of the total scatter matrix firstly and then extracts the null space of the within-class scatter in the resulting subspace. The brief steps of the NLDA in [79] as follows:

Step 1. Calculate the total scatter S_t and the within-class scatter S_w.

$$S_t = \sum_{i=1}^{N} (X_i - \overline{X})(X_i - \overline{X})^T \tag{3.7}$$

$$S_w = \sum_{i=1}^{C} \sum_{X_j \in C_i} (X_j - \overline{X_i})(X_j - \overline{X_i})^T \tag{3.8}$$

where N is the number of all samples, c is the number of classes, and n_i is the number of samples in class C_i, and $\overline{X_i} = (1/n_i) \sum_{k=1}^{n_i} X_k$ is the mean of class C_i, and $\overline{X} = (1/c) \sum_{k=1}^{c} \overline{X_k}$ is the mean of all samples.

Step 2. Perform eigen-analysis on S_t to obtain the projection matrix P, whose columns are the eigenvectors of S_t corresponding to the $N - 1$ largest eigenvalues. Then the within-class scatter in the resulting subspace is obtained as

$$S_w^{'} = P^T S_w P \tag{3.9}$$

Step 3. Perform eigen-analysis on $S_w^{'}$. Discard those eigenvectors with eigenvalues far from 0, and keep $c - 1$ eigenvectors of $S_w^{'}$. Let Q be the null space of $S_w^{'}$ which contains the $c - 1$ eigenvectors of $S_w^{'}$. We have

$$Q^T S_w^{'} Q = 0 \tag{3.10}$$

Step 4.The final NLDA projection matrix is

$$W = PQ \tag{3.11}$$

and the projection result is

$$Y = W^T X \tag{3.12}$$

Theorem 1. The NLDA subspace projection result remains the same if the transformation is orthonormal.

Proof: Assume that the original data X is transformed by an orthonormal matrix H, i.e., $Z = H^T X$. From (3.7) and (3.8), the total scatter and the within-class scatter are obtained as

$$\widetilde{S_t} = H^T S_t H \tag{3.13}$$

$$\widetilde{S_w} = H^T S_w H \tag{3.14}$$

Following the steps of the NLDA, firstly eigen-analysis is performed on $\widetilde{S_t}$, i.e.

$$\widetilde{S_t}\widetilde{e_i} = \lambda_i \widetilde{e_i}, i = 1, 2, \cdots m. \tag{3.15}$$

Substituting (3.13) into (3.15), we have

$$H^T S_t H \widetilde{e_i} = \lambda_i \widetilde{e_i} \Longrightarrow S_t H \widetilde{e_i} = \lambda_i H \widetilde{e_i} \tag{3.16}$$

Compared to the eigenvectors e_i of S_t, it can be seen that

$$e_i = H\widetilde{e_i} \tag{3.17}$$

Hence, the projection $\widetilde{P}$ satisfy

$$\widetilde{P} = H^T P \tag{3.18}$$

Based on (3.9), we have

$$\widetilde{S_w'} = \widetilde{P}^T \widetilde{S_w} \widetilde{P} \tag{3.19}$$

Substituting (3.14) and (3.18) into (3.19), we have

$$\widetilde{S_w'} = (H^T P)^T H^T S_w H (H^T P) = P^T S_w P = S_w' \tag{3.20}$$

Therefore, the null space $\widetilde{Q}$ of $\widetilde{S_w'}$ is equivalent to the null space Q of S_w'. Combining with (3.18), the final projection matrix is obtained as

$$\widetilde{W} = \widetilde{P}\widetilde{Q} = H^T PQ = H^T W. \tag{3.21}$$

Hence, the projection result of Z is

$$\widetilde{Y} = \widetilde{W}^T Z = (H^T W)^T (H^T X) = W^T X = Y \tag{3.22}$$

From (3.22), we can conclude that the projection result of NLDA is invariant for the orthonormally transformed data.

Referring to the definition, when n=1, the WHT matrix

$$H_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{3.23}$$

Obviously, the matrix is symmetric and orthonormal. When $n > 1$, suppose H_n is a symmetric and orthonormal matrix, from (3.3), H_{n+1} is also symmetric and

$$\begin{aligned} H_{n+1} * H_{n+1} &= \frac{1}{\sqrt{2}} \begin{pmatrix} H_n & H_n \\ H_n & -H_n \end{pmatrix} * \begin{pmatrix} H_n & H_n \\ H_n & -H_n \end{pmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} H_n * H_n + H_n * H_n & H_n * H_n - H_n * H_n \\ H_n * H_n - H_n * H_n & H_n * H_n + H_n * H_n \end{pmatrix} \end{aligned} \tag{3.24}$$

As H_n is a symmetric and orthonormal matrix, $H_n * H_n = I_{2^n}$, and we have

$$H_{n+1} * H_{n+1} = \frac{1}{2} \begin{pmatrix} 2I_{2^n} & 0 \\ 0 & 2I_{2^n} \end{pmatrix} = I_{2^{n+1}} \tag{3.25}$$

Hence, H_{n+1} is orthonormal. Based on the above proof, it is clear that the WHT matrix is symmetric and orthonormal, which is similar to the DCT matrix. In [76], the 2D DCT and block-wise 2D DCT are proved to be orthonormal transformations

because the DCT matrix is orthogonal. Therefore, we can conclude that the 2D WHT and block-wise 2D WHT are both orthonormal transformations.

Combining the above proof that the result of NLDA is invariant for the orthonormal transformed data, it can be concluded that the NLDA can be directly implemented in the 2D WHT and block-wise 2D WHT domain. Furthermore, as the projection results of the PCA are invariant for the orthonormal transformed data [76], the PCA can also be directly implemented in the 2D WHT and block-wise 2D WHT domain.

3.4 Experimental Results and Discussions

3.4.1 Database

The test databases are Yale Face database B and Extended Yale Face database B. These two databases contain face images with illumination variations. In the Yale Face database B, there are 10 persons with 64 different illumination conditions for nine poses per person [11]. In the Extended Yale Face database B [13], there are 16128 images of 28 persons with the same conditions as Yale B. Because the main concern in this work is illumination variations, only 64 frontal face images per person under different illumination conditions are chosen. The Extended Yale B database is combined with the Original Yale B database and named as Completed Yale B. There are 38 subjects in total in the Completed Yale B. Because there are 18 corrupted images of 7 subjects in the Extended Yale B, all these corrupted images are discarded. The images are divided into 5 subsets based on the angle between the light direction and the camera axis as other methods shown in Table 3.2. Because of lack of coordinates of the eyes in the Extended Yale B database, we directly use

Table 3.2: Subsets divided based on light source direction

	Subset 1	Subset 2	Subset 3	Subset 4	Subset 5
Light Angle	$0 \sim 12$	$13 \sim 25$	$26 \sim 50$	$51 \sim 77$	> 77
Number of Images in Original Yale B	70	120	120	140	190
Number of Images in Extended Yale B	193	336	335	386	524
Number of Images in Completed Yale B	263	456	455	526	714

the cropped and aligned images 192×168 provided by the database [13].

3.4.2 Experimental Results

In the experiments, there are two different test cases. In Case 1, Subset 1 is used as training samples (7 pictures per person) as in [32], and in Case 2, only one frontal image per person is applied as a training sample, which increases the difficulty of recognition.

For Case 1, because there are some corrupted images in Subset 1 of the Extended Yale B database, only Original Yale B database is used for the test. For Case 2, Original Yale B and Completed Yale B are applied as test database respectively. Recognition is performed with the nearest neighbour classifier measured with the Euclidean distance. For comparison, the proposed method of [32] is also implemented and named as the DCT. As in [32], all the face images are normalized with zero mean and unit variance. The results are shown in Table 3.3.

From the table, it is clear that the block-wise WHT provides a similar performance as the DCT, and it outperforms the DCT for the cases with larger illumination variations such as Subset 4 and 5. Meanwhile, the performances in the Case 2

Table 3.3: Recognition error rates of different methods

Methods	Yale B Case 1			Yale B Case 2			Completed Yale B Case 2		
	Subset			Subset			Subset		
	3	4	5	3	4	5	3	4	5
DCT	0	1.43	0	0	2.86	3.68	10.55	10.84	12.61
Block-wise WHT	0	0.71	0	1.67	0.71	1.58	10.99	9.32	12.61

test of Completed Yale B are still not satisfying because the database contains more subjects and the test only uses one picture per person as a training sample. Please note that the results in Table 3.3 are slight different from the results mentioned in Table 2.2 because the size of images is 192*168 different from that applied in the Table 2.2.

3.4.3 Performance Comparison for Different Parameters

As the DCT method of [32], one important factor of the performance is how to select low-frequency coefficients to be discarded. Zeroing the low-frequency coefficients can be implemented by multiplying the image with a mask. In this chapter, the mask is defined as follows:

$$Z(x,y) = \begin{cases} 1, & |x-a| + |y-b| \leq k \\ 0, & |x-a| + |y-b| > k \end{cases} \tag{3.26}$$

where (x, y) denote the coordinate of the pixel, (a, b) is the original point (normally it is (0, 0)) and k is the size of the mask.

For comparison, the performances of the above mask with different sizes k are tested. The results are shown in Figures 3.1 and 3.2. From the figures, it can be seen that the method always achieves the best average performance when $1 \leq k \leq 4$. The performances of different shapes of the mask will be studied in future work.

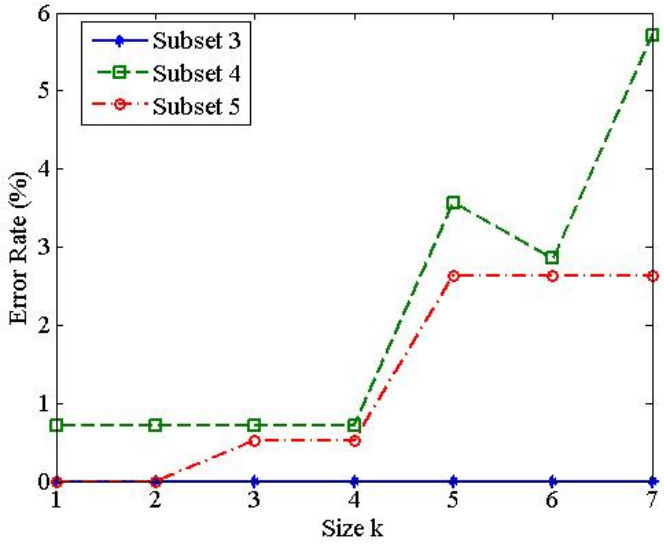


Figure 3.1: Performance on Case 1 of Yale B with different sizes

3.4.4 Performance Comparison between WHT in Entire Images and Blocks

For comparison, the WHT is implemented in the entire images instead of blocks. Because the WHT matrix must be $2^n \times 2^n$, the size of the images must be the same. Therefore, the images are rescaled to 128×128. The Yale B database is used for testing. The results are shown in Table 3.4. From the table, the block-wise WHT achieves much better performances than the WHT implemented in the entire images, both in Case 1 and Case 2.

3.4.5 Performance Comparison between PCA and NLDA in Different Domains

Furthermore, the PCA and NLDA implemented in the spatial domain are compared with these methods in the block-wise WHT domain. The Yale B database is used for the Test Case 1. For the PCA, the number of eigenvectors is set as 50. For

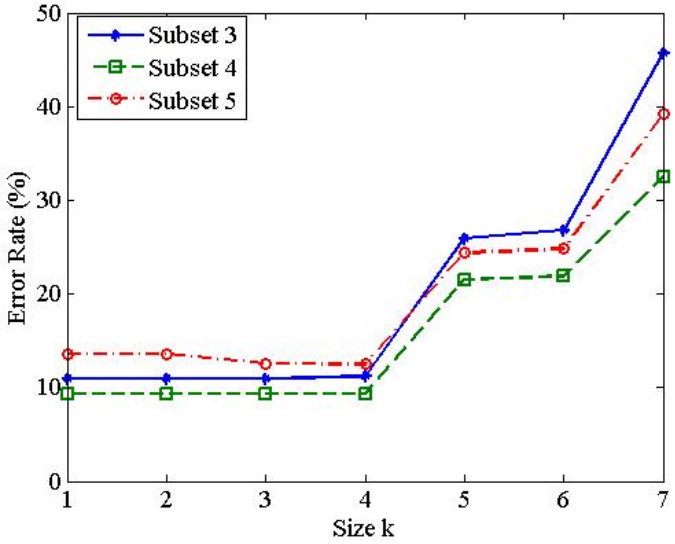


Figure 3.2: Performance on Case 2 of Completed Yale B with different sizes

Table 3.4: Performance comparisons of WHT in entire image and blocks

Method	Recognition Error Rate (%)					
	Yale B Case 1			Yale B Case 2		
	Subset3	Subset4	Subset5	Subset3	Subset4	Subset5
WHT in entire images	8.3	55.7	36.8	15	48.6	46.3
WHT in blocks	0	2.1	1.1	5	7.1	4.2

the NLDA, the number of final projection dimensions is 9. From Table 3.5, it is clear that these methods carried out in the spatial domain obtain the exact same performances as the methods directly implemented in the block-wise WHT domain. Therefore, the PCA and NLDA can be directly implemented in the block-wise WHT domain without the inverse WHT.

Table 3.5: Performance comparisons of PCA and NLDA in different domains

Size k	Recognition Error Number			
	PCA in block-wise WHT domain	PCA in spatial domain	NLDA in block-wise WHT domain	NLDA in spatial domain
1	2	2	9	9
2	2	2	9	9
3	2	2	6	6
4	2	2	6	6
5	11	11	11	11
6	11	11	14	14
7	14	14	94	94

3.5 Conclusions

In this chapter, a low computation complexity face recognition method is proposed to address the problem of illumination variations based on block-wise WHT. Low-frequency WHT coefficients are zeroed to compensate illumination variations. The method makes a good compromise between computational burden and recognition accuracy. Experiments on the CMU PIE, Yale B and Extended Yale B database show that the method can achieve performances comparable to the DCT method with a higher efficiency. For some cases with larger illumination variations, the proposed method even outperforms the DCT. Both analytical proof and experimental results demonstrate that the PCA and NLDA can be directly implemented in the WHT domain to further reduce the computational burden. However, the performances are still not satisfactory for some cases with larger illumination variations. Further research will be carried out to improve the performance.

Chapter 4

A Novel Efficient Local Illumination Compensation Method Based on DCT in Logarithm Domain

4.1 Introduction

One of the challenging issues in automatic face recognition is to achieve illumination invariance. It has been proven that in face recognition, differences between different lighting conditions are much greater than differences between individuals [1]. Various methods have been developed to address the issue. Generally, they can be classified into three categories:

a) Face and Illumination Modelling: There are two modelling classes, namely face modelling and illumination modelling. Regarding face modelling, because il-

lumination 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. The representative method is the 3D morphable model proposed in [17] which represents not only face shape but also facial texture under multiple variations such as poses and illuminations. High computational load is one of the disadvantages for this kind of methods. For illumination variation modelling, researchers have attempted to construct images under different illumination conditions according to a statistical model or a physical model. The principal component analysis (PCA) [3–5] and linear discriminant analysis (LDA) [6, 7] are two of the most representative statistical modelling methods where no assumptions concerning the surface are needed. Modelling based on a physical model needs to make some assumptions on certain surface reflectance properties such as Lambertian surface [8]. The famous Illumination Cone [9–11], 9D linear subspace [12] and nine point lights [13] all belong to the illumination modelling based on a physical model. One of the main drawbacks is that several images are required for modelling.

b) Illumination Invariant Feature Extraction: In these approaches, the main purpose is to obtain robust facial features against varying illuminations. The common representations include edge map, Gabor-like filtering image and image intensity derivatives [1]. 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) [18] is defined as image ratio between a test image and linear combinations of three unknown independent illumination images. The quotient image is free of illumination and only related to the surface texture information. However, the performance of QI depends on the bootstrap database. Shan et al. in [28] proposed Quotient Illumination Relighting (QIR) based on the assump-

tion that faces are an ideal class of objects. But, the method requires the lighting condition of the image to be known a priori. Without the bootstrap database and known lighting conditions, Wang et al. [21, 22] 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. In Morphological Quotient Image (MQI) proposed by He et al. [25], morphology operation is applied to smooth up images to estimate illumination.

c) Pre-processing and Normalization: In this approach, face images under illumination variations are pre-processed so that images under normal lighting can be obtained. Further recognition will be performed based on the normalized images. Histogram Equalisation developed in [29] is the most commonly used method. Recently, in [32], low-frequency discrete cosine transform (DCT) coefficients are discarded to eliminate effects of illumination variations because illumination variations can be assumed to depend on the low-frequency band. The method does not require multiple images to be trained and good experimental results both in the CMU PIE database and Yale Face database B are obtained. Based on Chens idea, [33] proposed to rescale low-frequency DCT coefficients (except the first) to lower values instead of discarding them in Chens approach. Perez and Castillo [34] also proposed a similar method which applied Genetic Algorithms to search appropriate weights to rescale low-frequency DCT coefficients. In [35], Heydi presented a new approach combining the DCT and the LBP for illumination normalization. The method divided images into blocks and applied Chens method in each block. After that, the LBP is used to represent facial features because the LBP can represent facial structures well when variations in lights are monotonic. In another recent work, Chen T. et al. [24] developed Logarithmic Total Variation (LTV) model, which decomposes the image into two parts, naming one with small-scale features and the other with large-

scale features. Only small-scale features will be used for recognition. The model is inspired by the SQI but has better edge-preserving ability and simpler parameter selection and promising results have been achieved. However, the computational burden is very high because it has to find an optimal solution of decomposition step by step. In Xie et al. [26], a similar illumination compensation frame is proposed. In the frame, the image is also decomposed into two parts: small scale and large scale. Normalization is performed in large-scale features rather than the entire image while small-scale features should be smoothed up. Subsequently, an image 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 this chapter, we mainly focus on the pre-processing and normalization process. Given that illumination variations affect less in localities than those in the entire image, we propose a new illumination compensation method, which uses low-frequency DCT coefficients obtained in a local area to estimate the illumination of each point, followed by a refining estimation step with the mean operator. Based on the analysis, a simplified version of the proposed method is also proposed. Theoretical proof demonstrates the validity and high efficiency of the simplified method. Both methods are tested on the CMU, the Yale Face database B and the Extended Yale Face database B and superior performances are obtained, particularly when only one frontal image per subject is used as a training set in the Yale Face database B and the Extended Yale Face database B. The proposed method has been presented in [80].

4.2 Local Illumination Normalization Based on DCT

4.2.1 Face and Illumination Model

In Lambertian theory, the face image can be modelled as follows:

$$I(x, y) = \rho(x, y)n(x, y)s \tag{4.1}$$

where $I(x, y)$ is the intensity value of a point (x,y) in the image, $0 \leq \rho(x, y) \leq 1$ is the corresponding albedo, $n(x, y)$ is the surface normal vector and s is the light vector whose magnitude is the intensity of the incoming light. The term $H = \rho(x, y)n(x, y)$,which depends on albedo and normal, is the intrinsic feature of an individual.

As shown in [32], illumination compensation can be implemented in the logarithm domain. After logarithm transform, the model changes to

$$\log I(x, y) = \log \rho(x, y) + \log(n(x, y)s) \tag{4.2}$$

The first part is a stable characteristic of facial features and the second part is related to illumination. The purpose of illumination compensation is to obtain $\log \rho(x, y)$ or some functions which are only related to $\log \rho(x, y)$ from $\log I(x, y)$.

In [32], low-frequency DCT coefficients are discarded to eliminate effects of illumination variations because illumination variations mainly lie in the low-frequency band. From another point of view, the method proposed in [32] can also be considered to estimate the illumination of each point $\log(n(x, y)s)$ based on low-frequency parts of DCT coefficients obtained in a global way.

In this chapter, we made two assumptions: 1) Illumination variations can be considered to mainly lie in the low-frequency band. The reason is that illumination

usually changes slowly compared with the reflectance except some casting shadows. Thus the proposed method uses low-frequency DCT coefficients obtained in a local area to estimate $\log(n(x,y)s)$ coarsely. This assumption has been widely made in illumination invariant face recognition techniques [24, 26, 32]; 2)Illumination in a small facet is considered to be constant because the direction of light source s is constant in the facet as well as the intensity of the incoming light, which has been assumed in [91] and obtained good performances based on their experiments. Based on this assumption, a mean operator is applied to a small local area to refine the estimation of illumination.

4.2.2 Discrete Cosine Transform

In face recognition, the DCT-2 is more widely used in [81–86] than other types of discrete cosine transform. In this chapter, the DCT-2 is often simply referred to as "the DCT". The 2D $M \times N$ DCT is defined as follows

$$A(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \alpha(u)\alpha(v) f(x,y) \cos\left[\pi(2x+1)u/2M\right] \cos\left[\pi(2y+1)v/2N\right] \quad (4.3)$$

and the inverse transform is defined as

$$f(x,y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \alpha(u)\alpha(v) A(u,v) \cos\left[\pi(2x+1)u/2M\right] \cos\left[\pi(2y+1)v/2N\right] \quad (4.4)$$

where

$$\alpha(u) = \begin{cases} 1/\sqrt{M}, & u = 0 \\ 2/\sqrt{M}, & u = 1, 2, \ldots, M-1 \end{cases} \quad (4.5)$$

and

$$\alpha(v) = \begin{cases} 1/\sqrt{N}, & v = 0 \\ 2/\sqrt{N}, & v = 1, 2, \ldots, N-1 \end{cases} \quad (4.6)$$

In [32], the DCT is performed on the entire face image to obtain all frequency components of the face image. In the proposed method, the DCT is applied to a local area around each point.

4.2.3 Local Illumination Compensation

As mentioned before, since illumination variations can be assumed to depend on the low-frequency components and illumination variations affect less in localities than those in the entire image, the low-frequency DCT coefficients obtained in a local area are used to estimate the illumination of each point l(x,y) coarsely. Considering the face model (2), in a small neighbourhood around each point (x, y), the surface normal n(x,y) does not have much variation. Furthermore, the light source s can be considered as a constant. Hence, the second part $\log(n^T s)$in (2) can be taken as a constant C. Ideally, there is no error (noise) in illumination estimation l(x, y). Hence, illumination estimation of each point l(x, y) in a small neighbourhood should be equal to the constant C. However, there may be some estimation errors (or noises) at each point. Hence, $l(x, y) = C + e(x, y)$ where e(x,y) is the noise of the point (x, y) in a small neighbourhood. Our objective is to provide an optimal estimation of C from a set of l(x, y). To estimate the illumination more accurately, we use the mean operator to refine the estimation.

Based on the above idea, the steps of the algorithm are as follows:

1) Transform an image into the logarithm domain and obtain the image I.

2) For each point of the image with the coordinate (m, n) in the image, carry out the following steps:

a) Determine the shape and the size of a local area ϕ around the point. In this

chapter, a conventional square shape is chosen and it is defined as follows:

$$|x - m| \leq k \; and \; |y - n| \leq k \tag{4.7}$$

where (x,y) is the coordinate of a point in the local area ϕ and k is the radius (size) of the local area ϕ.

b) Apply the DCT to the local area ϕ and obtain the DCT coefficient matrix A as defined in (3).

c) Replace the values of all DCT coefficients with zeros except the low-frequency part in matrix A and obtain the reconstructed matrix B.

d) Apply inverse DCT to matrix B to estimate the luminance of each point

$$l(x, y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \alpha(u)\alpha(v)B(u, v) \cos[\pi(2x + 1)u/2M] \cos[\pi(2y + 1)v/2N] \tag{4.8}$$

where $\alpha(u)$ and $\alpha(v)$ are defined in (5) and (6) respectively.

e) Choose the shape and the size of another local area ω. In this chapter, we will consider three kinds of local area shapes, namely square, diamond and circle. For convenience, we take the size as the same as that in step (a). A square area is defined as the same as that in step (a). A diamond area is defined as

$$|x - m| + |y - n| \leq k \tag{4.9}$$

while a circle area is defined as

$$(x - m)^2 + (y - n)^2 \leq k^2 \tag{4.10}$$

where (x,y) denotes coordinates of points in area ω and k is the radius of the local area ω.

f) Using the mean operator to refine the illumination estimation of this point

in local area ω, we have

$$\begin{aligned} r(m,n) &= mean(l(x,y)), \forall (x,y) \in local\,area\,\omega\,around\,point(m,n) \\ &= \sum_{\forall (x,y) \in local\,area\,\omega\,around\,point(m,n)} l(x,y)/the\,number\,of\,points\,in\,local\,area\,\omega \end{aligned} \tag{4.11}$$

g) Illumination compensation is given by

$$f(m,n) = I(m,n) - r(m,n) \tag{4.12}$$

4.2.4 Simplified Method

When we choose the shape of the local block ω as a square area, the same as the shape of ϕ, the algorithm can be simplified and a much more efficient version can be obtained. We will discuss in details and provide an analytical proof in this section.

Noting that the low-frequency coefficients in matrix B have the same values as those in matrix A in step (c), regardless of how many low-frequency coefficients are used for a square local area ϕ with $M \times M$, we have

$$A^{'}(0,0) = A(0,0) \Rightarrow \sum_{x=0}^{M-1}\sum_{y=0}^{M-1} l(x,y)/\sqrt{M \times M} = \sum_{x=0}^{M-1}\sum_{y=0}^{M-1} I(x,y)/\sqrt{M \times M} \tag{4.13}$$

When a square shape is chosen in step (e) as that in step (a), because both area ϕ and ω have the same size (supposed to be $M \times M$), we have

$$\begin{aligned} r(m,n) &= mean(l(m,n)) \\ &= \sum_{x=0}^{M-1}\sum_{y=0}^{M-1} l(x,y)/M \times M \\ &= \sum_{x=0}^{M-1}\sum_{y=0}^{M-1} I(x,y)/M \times M \\ &= mean(I(m,n)) \end{aligned} \tag{4.14}$$

In view of this, regardless of how many low-frequency coefficients are used, if the ω is chosen as a square area, the refined estimation r(m, n) is equal to the mean of I(m, n) in the same local area. To reduce the computational burden, the value of pixels after illumination compensation can be directly calculated as follows:

$$f(m,n) = I(m,n) - mean(I(m,n)) \tag{4.15}$$

where

$$mean(I(m,n)) = \sum_{\forall (x,y) \in local\ area\ \omega\ around\ point\ (m,n)} I(x,y)/the\ number\ of\ points\ in\ local\ area\ \omega \tag{4.16}$$

Therefore, the steps of the simplified method are as follows:

1) Transform an image into the logarithm domain and obtain the image I.

2) Determine the size of the local area ω.

3) For each point of the image with the coordinate (m,n) in the image, calculate the value of pixels f (m,n) using (4.15).

As mentioned in the beginning of Section 4.2.3, the second part $\log(n^T s)$ in (4.2) can be taken as a constant C. Hence, following the face model (4.2) and the above simplified algorithm, we have

$$\begin{aligned} f(m,n) &= I(m,n) - mean(I(m,n)) \\ &= \log C + \log\left(\rho(m,n)\right) - mean(\log C + \log\left(\rho(m,n)\right)) \\ &= \log\left(\rho(m,n)\right) - mean(\log\left(\rho(m,n)\right)) \end{aligned} \tag{4.17}$$

After compensation, the value of pixels is only related to intrinsic features of individuals and does not depend on illumination.

For an image of $N \times N$ and a square local area of $M \times M$, computational complexities of the simplified algorithm and the original proposed method as well

Table 4.1: Computational complexities of different methods

Methods	Computational Complexity	Real Computational Time (ms)
DCT	$O(N^2 \log_2^N)$	503
Proposed Method	$O(N^2M^2 \log_2^M)$	14292
Simplified Method	$O(N^2M^2)$	910

as the DCT method mentioned in [32] are shown in Table 4.1. Real computational times of the above methods for an image of 192×168 in a personal computer with a 2.66GHz CPU are also listed in the table. For comparison, the LTV takes 17163.9 ms per image in the same condition. From the table, we can see that computational complexity of the simplified method decreases obviously and it is suitable to practical applications. Please note that if we choose a different shape of the local area ω in step (e), meaning not square, or choose a different operator, not mean operator, the above conclusion is not right. We cannot deduce the algorithm in the above way in these cases. The performances of different shapes of local area and different operators will be discussed in the sequel.

4.3 Experimental Results and Discussions

4.3.1 Face Database

In this chapter, in order to validate the proposed method, we test the method on three popular databases: CMU PIE [87], Yale Face database B [11] and Extended Yale Face database B [13]. For the CMU PIE, because we are concerned with the illumination variation problem, only 21 frontal face images per subject under different illumination conditions are chosen. All images are aligned, cropped, and re-sized to 120×105 by [32]. In the Yale Face database B, there are 64 different illumination

Table 4.2: Subsets divided based on the light source direction

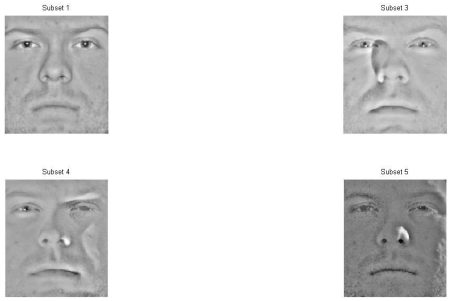
	Subset1	Subset2	Subset3	Subset4	Subset5
Light angle	$0 \sim 12$	$13 \sim 25$	$26 \sim 50$	$51 \sim 77$	> 77
Number of images in Original Yale B	70	120	120	140	190
Number of images in Completed Yale B	263	456	455	526	714

conditions for nine poses per subject. The Extended Yale Face database B has the exact conditions as the Yale Face database B but with 28 different subjects. For both databases, only 64 frontal face images per subject under different illumination conditions are used in the experiments. In the experiments, we combine the Yale Face database B with Extended Yale Face database B and name the combined database as Completed Yale B. To avoid ambiguity, we name the Yale Face database B as Original Yale B. All images are aligned, cropped, and then re-sized to 192×168 images by [13]. We divide the images into 5 subsets based on the angle between the lighting direction and the camera axis as other methods shown in Table 4.2. Figure 4.1 (a) shows the images of one subject with different illumination conditions and Figure 4.1 (b) shows the normalized images processed by the proposed method where k is set as 12 and the number of discarded DCT coefficients is 6. From the figure, we can see that the differences between images caused by illumination variations are reduced by the proposed method and more face features can be seen in the dark part of original images.

4.3.2 Experimental Results

In the first experiment, we test our proposed methods on the CMU and compare them with other existing methods. In the experiment, one frontally lighted face

(a) Original Images



(b) Normalized Images

Figure 4.1: Samples of Subject 01 with different illumination conditions in the Yale B

image per subject is selected as the gallery and the other 20 images per subject under different illuminations are used as a probe set. Recognition is performed by the nearest neighbor classifier with normalized correlation as the distance measurement, i.e. cosine of the angle between two image vectors. For the purpose of comparison, we implement the methods in [24, 32], and name them as the DCT and LTV respectively. Some results of other existing methods are directly referred to other papers because the databases and other conditions are the same. The performance results on the CMU are shown in Table 4.3. From Table 4.3, we can see that the proposed methods including the simplified version outperform other existing methods

Table 4.3: Comparison of recognition performances for different methods in CMU PIE

Methods	Error Rate (%)
No Illumination Compensation	43.0
Histogram Equalization	47.8
QI	24.34
QIR	11.35
SQI	2.09
DCT	0.37
LTV	1.03
Our Method	0.22
Simplified Method	0.22

though the improvement of our both proposed methods compared to the DCT looks very slight. One of the probable reasons is that illumination variations in the CMU database are limited and it is easier to deal with variations in the database.

The proposed methods are also tested on the Original Yale B which contains more variations. To make our method more convincing, we employ the Colorado State University (CSU) Face Identification Evaluation System in [88] to evaluate our methods combined with other recognition approaches besides the nearest neighbour classifier. In the experiment, the images from Subset 1 are used as the gallery and other images from Subsets 2-5 are used as probe sets. In the CSU face recognition system, a standard PCA and a combination PCA and LDA algorithm are tested as baseline recognition approaches respectively. In the experiment, we only compare our proposed methods with the DCT and LTV combined with the mentioned two recognition methods. To generate the cumulative match curve, the Euclidean distance measurement is used. The performance results of different illumination compensation methods combined with the PCA on Subsets 3-5 are shown in Figures 4.2-4.4. From the figures, we can see that all the four illumination compensation

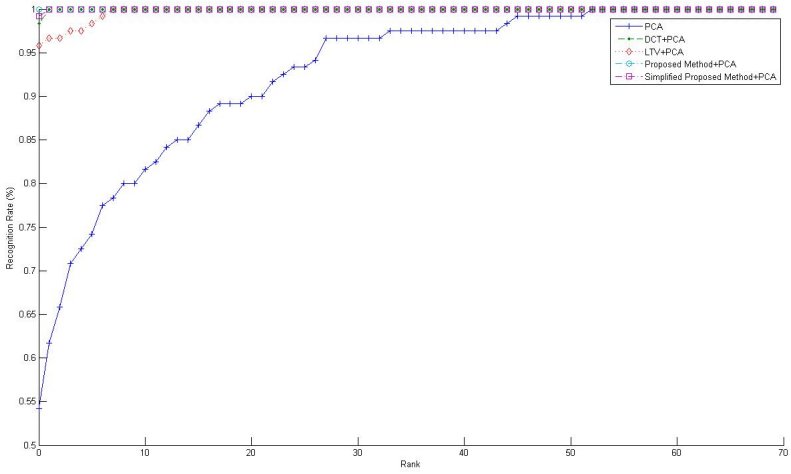


Figure 4.2: Performance comparison for the PCA combined with different illumination compensation methods in Subset 3

methods improve the performances of the standard PCA under illumination variations greatly, especially under larger illumination variations such as the cases in Subsets 4 and 5. Similar to the first experiment, our proposed methods still achieve the best performances in all the cases. For the combination PCA and LDA algorithm, the results are shown in Figures 4.5-4.7. Compared to the standard PCA, a combination PCA and LDA algorithm obtains better performances in all cases. It is also clear that our proposed methods improve the performance of a combination PCA and LDA algorithm significantly and outperform other existing illumination compensation methods. Please note that there is no obvious difference between the performances of our both proposed methods in the Original Yale B.

Compared to the Original Yale B database, the Completed Yale B database is later and has a lot more subjects so that a comparison in this database is more promising. In this experiment, only one image per subject from Subset 1 is chosen as the gallery, and the rest are used as probe sets. The case that only one image per subject is involved in gallery is more difficult in practical applications. The nearest

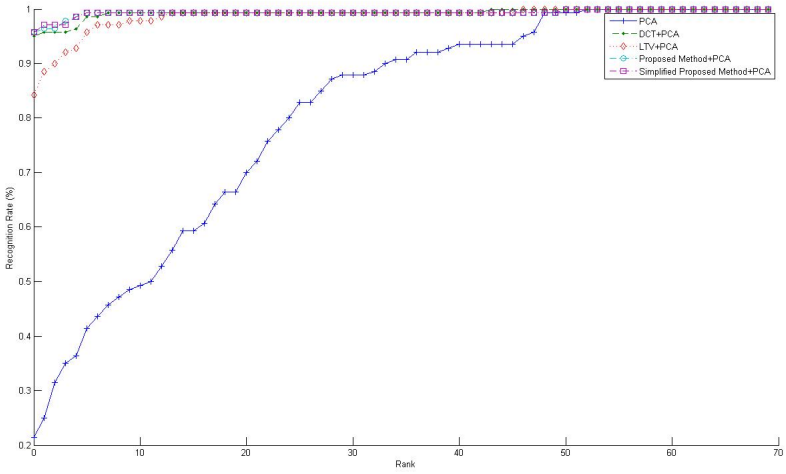


Figure 4.3: Performance comparison for the PCA combined with different illumination compensation methods in Subset 4

neighbour classifier is employed in this experiment with normalized correlation as the distance measurement. As shown in Table 4.4, our methods still obtain better performances than other methods. Moreover, we can see the differences between the proposed original method and the simplified method. For the cases with fewer illumination variations as Subset 3, the simplified method can provide the best performance. However, for the cases with larger illumination variations as Subset 4 and 5, the original proposed method is a preferred choice.

In addition to the nearest neighbour classifier, we implement a well-known face recognition system based on the LBP proposed in [41] and test the performances of different illumination compensation techniques combined with the face recognition system on the Completed Yale B. The results are shown in Table 4.5. The LBP is insensitive to monotonic gray-level variations resulted from illumination variations in addition to its robustness against variations in pose and expression. From the table, we can see that the performance of the LBP in the cases with small illumination changes as Subsets 3 is very good. However, the LBP is still sensitive to

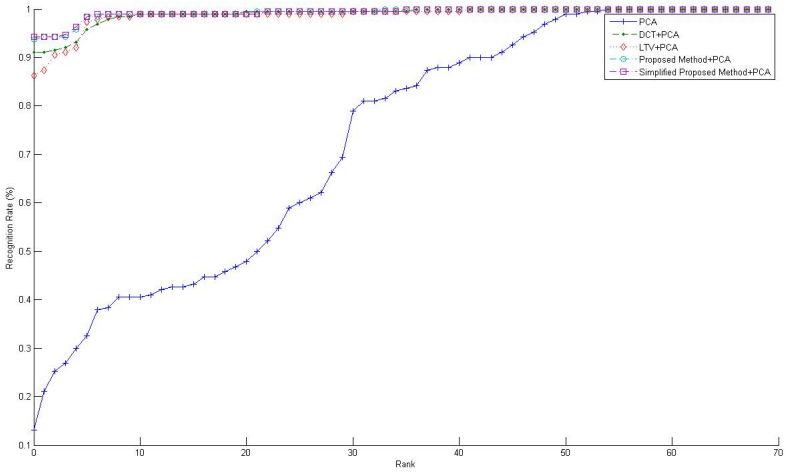


Figure 4.4: Performance comparison for the PCA combined with different illumination compensation methods in Subset 5

larger illumination variations such as Subsets 4 and 5. Combined with illumination compensation techniques such as the DCT method and our proposed methods, the LBP improves the performances under larger illumination variations significantly. Among all these techniques, our proposed methods obtain the best performances in all the cases.

Based on the above comparisons, we can conclude that our methods outperform other methods for all the conventional test databases. To make our claim more convincing, we carry out a statistics test based on the idea in [89]. Following its idea, we obtain the ranks of all the classifiers based on their performances on different databases, as shown in Table 4.6. Firstly, the Friedman test is carried out to check the null-hypothesis which assumes that all the methods are equivalent. The statistic F_F derived from the Friedman statistic χ_F^2 is calculated by

$$F_F = \frac{(N-1)\chi_F^2}{N(k-1)-\chi_F^2} \tag{4.18}$$

$$\chi_F^2 = \frac{12N}{k(k+1)}[\sum_{j=1}^{k} R_j^2 - \frac{k(k+1)^2}{4}] \tag{4.19}$$

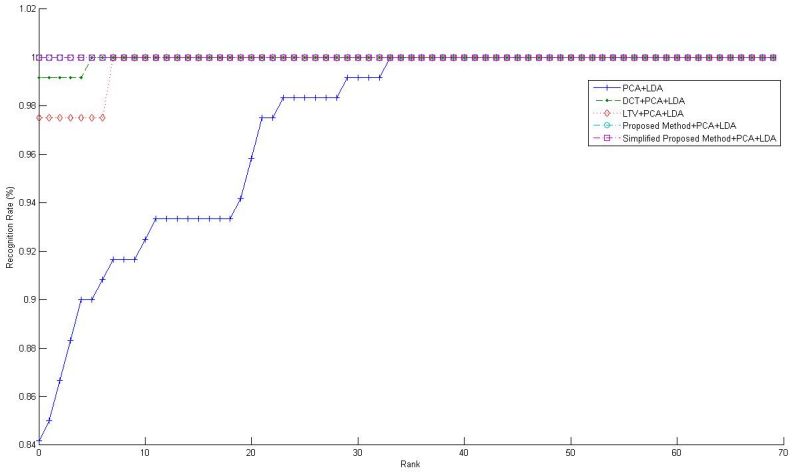


Figure 4.5: Performance comparison for the PCA+LDA combined with different illumination compensation methods in Subset 3

where N is the number of data set, k is the number of comparison methods, and R_j is the average rank of each method. Under the hypothesis, the statistic F_F is distributed according to the F-distribution with k-1 and (k-1)(N-1) degrees of freedom. In our case, with k=5 algorithms and N=13 data set, F_F is distributed according to the F distribution with 5-1=4 and (5-1)*(13-1)=48 degrees of freedom. The critical value of F(4,48) for $\alpha = 0.05$ is 2.56, but in our case $F_F = 75.76$ which is much higher than the critical value. Therefore we reject the null-hypothesis and the test proves that there are significant differences among all these methods.

Furthermore, we will test whether the proposed methods are better than the DCT and the LTV based on the Bonferroni-Dunn test. The details of Bonferroni-Dunn test can be referred to [90]. In the test, the critical difference (CD) is defined as

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}} \tag{4.20}$$

where q_α is set as 2.241 at $\alpha = 0.10$ for 5 algorithms. In our case, the CD with Bonferroni-Dunn test (at$\alpha = 0.10$) is 1.3898. In the test, the differences between

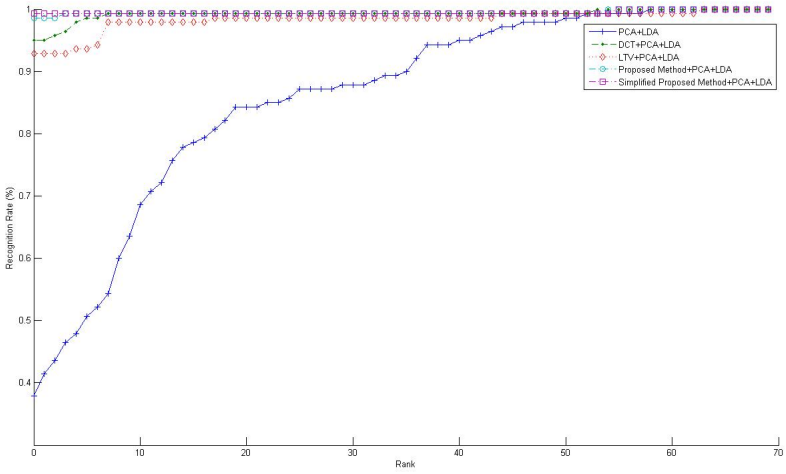


Figure 4.6: Performance comparison for the PCA+LDA combined with different illumination compensation methods in Subset 4

our proposed methods and the LTV (4-1.54=2.46 and 4-1.58=2.42) are greater than the CD 1.3898. Compared to the DCT, the differences are 3.04-1.54=1.5 and 3.04-1.58=1.46 respectively. Therefore, we can claim that there are significant improvements of our proposed methods compared to the LTV and DCT.

Our idea is similar to the RLS (LOG_DCT) to a certain extent. The images constructed by low-frequency part coefficients can be considered as large-scale features and the reminder can be taken as small-scale features. Our refining estimation focusing on the low-frequency part is similar to the idea in the RLS [26] that the normalization should be performed on a large-scale feature image instead of the entire image. One of the differences between our method and the RLS is that we just use low-frequency DCT coefficients to coarsely construct a large-scale image instead of an optimum solution of image decomposition. Hence, our method has a lower computational burden.

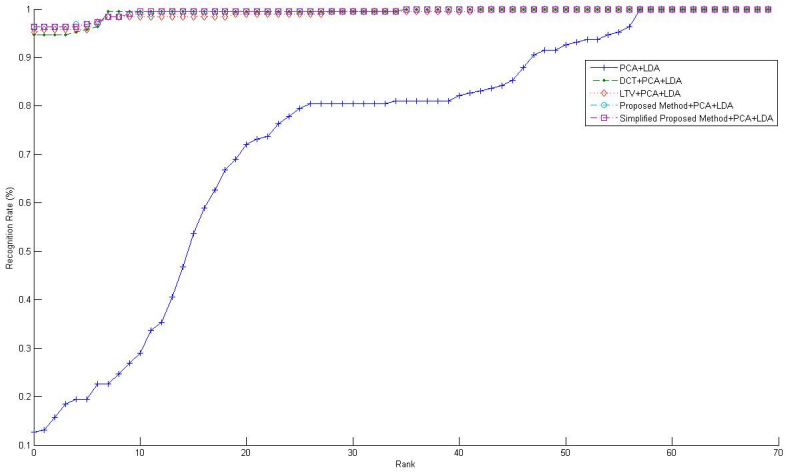


Figure 4.7: Performance comparison for the PCA+LDA combined with different illumination compensation methods in Subset 5

4.3.3 Analysis of Different Parameter Values

As foreshadowed, there are two important parameters which affect the performances of the proposed methods. One is the number of discarded DCT coefficients and the other is the local area shape and size in step (e). We will discuss the influences of these two parameters for our methods in this section. Because the Completed Yale B database contains more subjects and has more illumination variations, we use the Completed Yale B as the test database.

Because the number of discarded DCT coefficients does not influence the performance of the simplified method which uses the square local area, we only test the performances of diamond shape and circle shape in the comparison. Figure 4.8 and Figure 4.9 show the results of the proposed method with different number of discarded DCT coefficients for the diamond shape and the circle shape respectively. All the results are the average performance with the radius $11 \leq k \leq 14$. From the figures, we can see that the best performance can be obtained when $D_{dis} = 3$ for Subset 4 and Subset 5, regardless of whether the shape is diamond or circle. When

Table 4.4: Comparison of recognition performances for different methods in Completed Yale B databases

Methods	Error Rate (%)		
	Subset 3	Subset 4	Subset 5
No Illumination Compensation	49.23	88.02	96.50
Histogram Equalization	62.3	78.7	89.9
DCT	10.55	10.84	12.61
LTV	14.07	14.07	14.57
RLS(LOG_DCT)	12.9	12.4	15.2
Our Method	10.76	7.98	10.36
Simplified Method	10.55	8.17	10.92

Table 4.5: Comparison of recognition performances for different methods combined with the LBP in Completed Yale B database

Methods	Error Rate (%)		
	Subset 3	Subset 4	Subset 5
No Illumination Compensation+LBP	1.98	22.43	64.43
DCT+LBP	2.64	11.79	45.38
LTV+LBP	4.40	15.21	55.60
Our Proposed Method+LBP	1.54	5.70	35.99
Our Simplified Method+LBP	1.32	5.89	38.52

$D_{dis} > 3$, the performance only has a very slight decrease and is still better than that of the DCT. However, for the case of Subset 3 with fewer illumination variations, the best performances for the diamond shape and the circle shape are achieved with $D_{dis} = 2$. When D_{dis} is larger than 2, the performance also decreases slightly. 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. Nevertheless, our experimental results show that high performance can still be achieved without these features. Hence, for some

Table 4.6: Performance rank comparison for different methods

Methods	Rank												Average Rank
	CMU	Original Yale B						Completed Yale B					
		PCA			PCA+LDA			NN			LBP		
		3	4	5	3	4	5	3	4	5	3	4	5	
No compensation	5	5	5	5	5	5	5	5	5	5	3	5	5	4.85
DCT	3	3	3	3	3	3	4	1.5	3	3	4	3	3	3.04
LTV	4	4	4	4	4	4	3	4	4	4	5	4	4	4
Proposed method	1.5	1	2	1.5	1.5	2	1.5	3	1	1	2	1	1	1.54
Simplified method	1.5	2	1	1.5	1.5	1	1.5	1.5	2	2	1	2	2	1.58

cases with large illumination variances, low-frequency features are less effective for recognition. For some cases as Subset 3 where illumination variance is not very large, fewer low-frequency coefficients need to be discarded.

Besides the number of discarded DCT coefficients, the other important parameter is the local shape and size in step (e). We fix the shape in step (a) as square and choose the same radius k as that in step (e) for convenience. The study of different choices in step (a) will be carried out in further research. The following analysis focuses on the shape and size of the local area in step (e).

Figures 4.10-4.12 show the results of the proposed method with different shapes and sizes of the local area for Subset 3, Subset 4 and Subset5 respectively. Based on the figures, the performance of any shape will provide a better performance in Subsets 4 and 5 when a radius of k in the range $11 \leq k \leq 14$ is chosen. For Subset 3, a bigger radius $14 \leq k \leq 16$ is recommended. For any radius out of the recommended range, the performance decreases obviously. For the same radius, the square shape always obtains the best performance for any size in Subset 3, followed by the circle and the diamond shape. For Subsets 4 and 5, the best performances are obtained by the circle shape when the size is chosen in the recommended range.

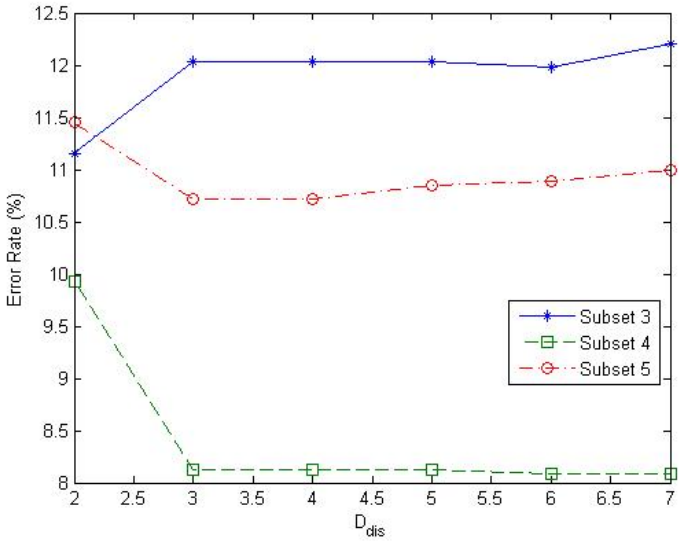


Figure 4.8: Performance comparison for different discarded low-frequency DCT coefficients with diamond local area

When $k \leq 11$, the square shape always outperforms the diamond shape. However, when $k > 12$, the diamond shape provides a better performance. Hence, for the cases with fewer illumination variations, the square shape can provide the best performance. However, for the cases with larger illumination variations, the circle shape is a better choice. It is also clear that a reasonable radius is the key point to obtain the best performance. Actually, when the local area is too small, the mean operator cannot provide a reliable statistical estimation. From the point of view of the face property, the illumination estimation of each point will ignore necessary information surrounding the point if the local area is too small. In such case, the performances of different shapes are quite different from Figures 4.10-4.12. One of the probable reasons is that the diamond shape contains much less points compared to the circular and square shapes and thus more information around the point are lost. However, when the area is too large, the assumption that the surface normal does not have much variation is not convincing. Furthermore, the light source can also be considered to be changing in such an area. The study on how to choose the

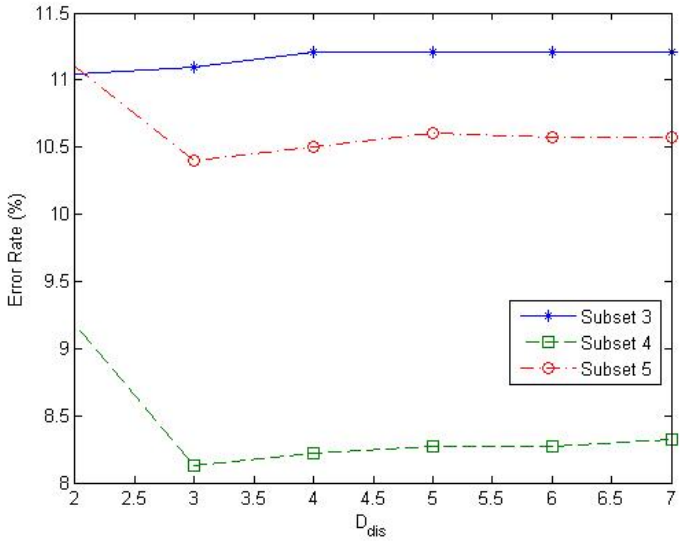


Figure 4.9: Performance comparison for different discarded low-frequency DCT coefficients with circle local area

value of radius automatically for different illumination variations and image sizes will be carried out in future research.

4.3.4 DCT Versus Local DCT Versus Proposed Method

As mentioned before, the DCT method in [32] can be considered as estimating the illumination of each point in a global way. In our method, the illumination estimation is firstly carried out coarsely in a local area, and then followed by a refining step with the mean operator. To compare global estimation with local estimation, we also implement our proposed method without the mean operator, named as local DCT. Besides, to study the effect of the mean operator, we also compare the local DCT with our proposed method. The Completed Yale B database is used for comparison. Performance comparisons are shown in Table 4.7. It is clear that better performance is achieved while DCT coefficients obtained in a local area are used for estimation. This is in contrast with the case where those DCT

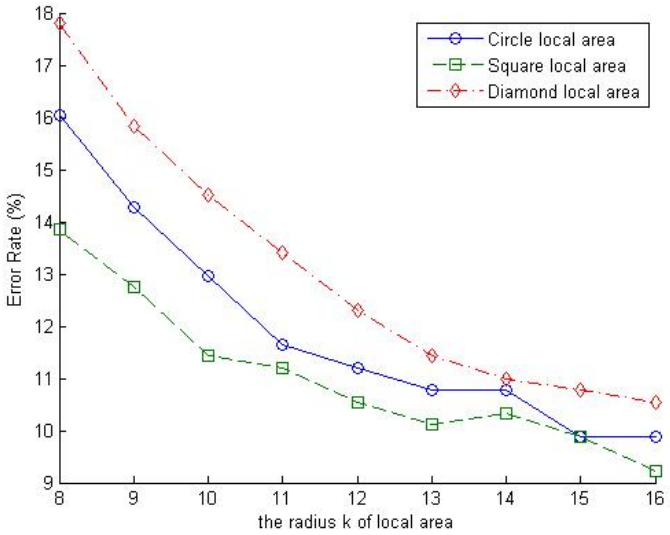


Figure 4.10: Performance comparison for different shapes and sizes of local area in Subset 3

coefficients obtained in an entire image are used for estimation, except for that in Subset 3. It is also evident that using additional mean operator to refine the estimation can obtain a better performance compared to the Local DCT.

4.3.5 Gaussian Operator Versus Mean Operator

In Section 4.2.4, we mentioned that different operators may be applied to refine luminance estimation. Besides the mean operator, the Gaussian operator is another popular operator as in [21] and the idea is based on another assumption that the luminance is Gaussian distributed in a local area. The mean operator can be considered as a mask with equal weights while the Gaussian operator can be taken as a mask with different weights. The points near the central point will have larger weights and the points away from the central point will have smaller weights. For the Gaussian operator, Eq. (4.10) used in the step (f) of the algorithm is changed

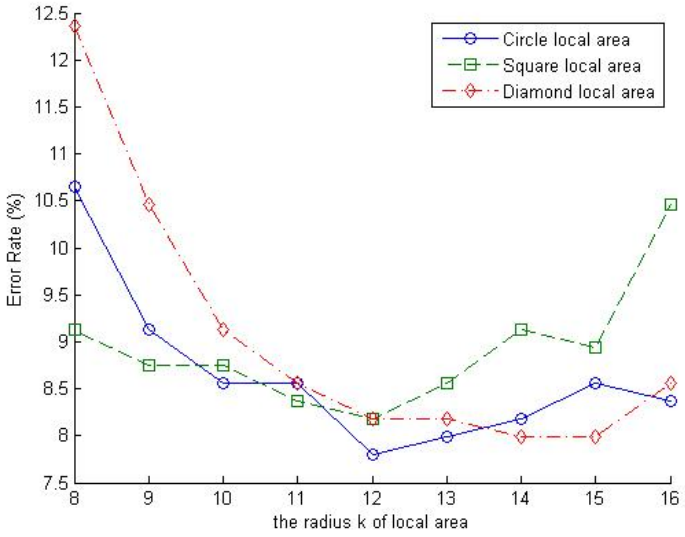


Figure 4.11: Performance comparison for different shapes and sizes of local area in Subset 4

to

$$r(m,n) = \sum_{s=-k}^{k}\sum_{t=-k}^{k} w(s,t)l(m+s,n+t) / \sum_{s=-k}^{k}\sum_{t=-k}^{k} w(s,t) \tag{4.21}$$

where

$$w(s,t) = \exp\left(-(w^2+t^2)/2\delta^2\right)/2\pi\delta^2 \tag{4.22}$$

and k is the radius of local area.

A comparison between the mean operator and the Gaussian operator is shown in Table 4.8. As in previous sections, the Completed Yale B database is used as the test database. From the table, we can see that the mean operator obviously obtains better performances. Although most of the previous works in other fields demonstrate that varying weights usually obtain a better performance than equal weights do, the result is not so in our illumination compensation. Based on the results, we find that the assumption that the illumination in an appropriate small local area is constant is more convincing than the assumption that the illumination in a local area is Gaussian distributed.

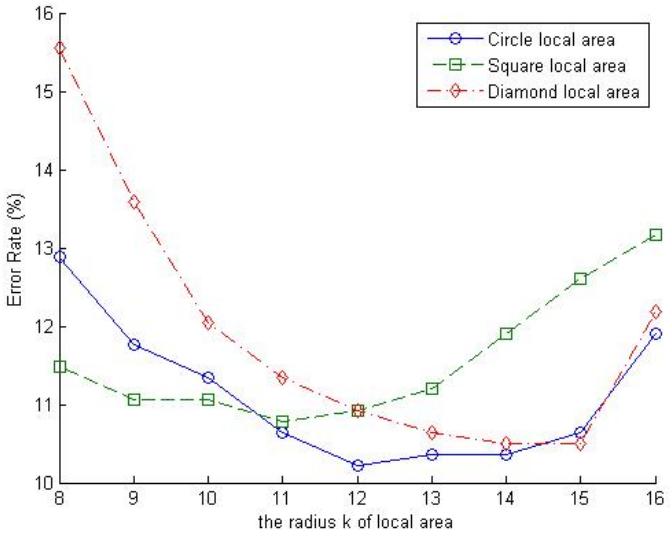


Figure 4.12: Performance comparison for different shapes and sizes of local area in Subset 5

4.4 Conclusions

A novel illumination compensation method for face recognition is developed in this chapter. The proposed method uses low-frequency DCT coefficients obtained in a local area to estimate illumination of each point in the logarithm domain of images instead of illumination compensation in a global way. A mean operator is applied to refine estimation of illumination at each point. Besides, a simplified version of the method is also proposed. Both theoretical analysis and experimental results prove the validity and high computational efficiency of the simplified version. In summary, the salient features of our methods are: 1) No modelling step and prior information are required; 2) Illumination compensation approach regarding DCT coefficients obtained in a local area is seldom reported and no similar research on refining the estimation obtained by DCT coefficients exists; 3) Our methods outperform the existing approaches and obtain a good computational efficiency. Nevertheless, an appropriate shape and size of the local area are still an important factor for good

Table 4.7: Performance comparison for DCT, Simply Local DCT and the proposed method

Methods	Error Rate (%)		
	Subset 3	Subset 4	Subset 5
DCT (Discarded Dimension=12)	10.55	10.84	12.61
Local DCT (Discarded Dimension=3)	11.65	9.88	12.04
Our Method (Discarded Dimension=3, Circle area radius=12)	10.76	7.98	10.36

Table 4.8: Performance comparison for mean operator and Gaussian operator

Methods	Error Rate (%)		
	Subset 3	Subset 4	Subset 5
Gaussian Operator (Discarded Dimension=3 Square area radius=8 Delta=3.0)	13.63	8.37	11.48
Our Method (Discarded Dimension=3, Circle area radius=12)	10.76	7.98	10.36

performance. Our future work will focus on how to choose the parameter values adaptively for different image sizes and illumination. Furthermore, more different estimation refining operators will be explored besides the mean operator and the Gaussian operator in future.

Chapter 5

Local Relation Map: A Novel Illumination Invariant Face Recognition Approach

5.1 Introduction

As one of three categories of illumination invariant face recognition approaches, illumination invariant feature extraction methods attempt to extract facial features that are robust against illumination variations. The common representations include edge map, Gabor-like filtering image and image intensity derivatives [1]. Recently, local binary pattern (LBP) [41] was proposed as another effective illumination invariant feature and gained much attention. The LBP operator is one of the best local texture descriptors. In addition to 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 main idea in LBP is to assign a texture pattern to each pixel based on the relationship between the gray value of

central pixel and those of around pixels, and then calculate the distribution of the patterns [40]. Based on the LBP, Ahonen et al. [41] propose an improved method to tackle the illumination variation. In [41], the face image is firstly divided into several regions and the LBP is applied in each region to calculate the histogram respectively. The resulting histograms are concatenated into an enhanced feature. Different weights are set for different regions because it is assumed that some of the facial regions contribute more than others in terms of extra personal variance such as the region containing eyes [41]. The experimental results prove that the weighted LBP outperforms the non-weighted LBP and other texture descriptor for face recognition under illumination variations. However, the LBP is still not robust enough against illumination variations especially larger change of illumination direction exists.

From another point of view, most of existing approaches explore facial features using (5.1). In this model, only reflectance and illumination are considered. The DCT method proposed in [32] is one of the most representative methods. The original image is processed by logarithm transformation followed by the DCT. After that, low frequency coefficients are zeroed to eliminate the effects caused by illumination variations since illumination variations are considered to mainly lie in the low frequency band [32]. Chen et al. propose another representative method in [24] based on a logarithmic total variation (LTV) model which decomposes an input image into large-scale output u and small-scale output v. The small-scale output is regarded as illumination invariant features which can be used in further recognition task. The idea behind the model is similar to the above mentioned DCT method but the LTV has a better edge-preserving ability. However, the performance of the LTV is still not satisfying especially in the cases where larger illumination variations exist, as shown in the following experiments. Furthermore, the LTV is an iterative

method and it takes a very high computational burden.

Different from most of previous works, the LN proposed in [91] involves the noise in the illumination model as an additive term besides the multiplicative term illumination. In the LN, the noise in a local area is considered as a constant. Based on local properties of facial images, the authors in [91] extend the assumption that illumination is related to low frequency band to that illumination can be considered as a constant (only related to the DC component) in a small local area. The proposed method achieves promising experimental results on the Yale, Yale B, AR and CMU PIE databases.

In this chapter, we propose a novel illumination invariant feature local relation map (LRM). The main idea of the LRM is based on local properties of facial images and it has a much lower computational load compared to the similar method LN [91]. Furthermore, to reduce the effects caused by noise, an additive term as noise is considered in the illumination model. Different previous researchers in [91], we propose to apply high frequency DCT coefficients obtained in the entire image to estimate the noise. After that, a logarithm transform is taken to change the model into an additive model. With the simplified model, the LRM is extracted as facial features for further recognition task. A brief version of the proposed method has been introduced in [92] with some preliminary results and detailed results have been presented in [93].

5.2 Illumination Invariant Face Recognition Technique

5.2.1 Face and Illumination Model

In most of existing approaches, the image under different illuminations is simply modeled as

$$f(x, y) = r(x, y) \cdot e(x, y) \tag{5.1}$$

where f(x,y) is the image gray level, r(x,y) is the reflectance and e(x, y) is the illumination. Based on the model, illumination variations are proposed to mainly lie in the low frequency band [32]. Therefore, low frequency DCT coefficients in the logarithm domain are discarded in [32] to compensate illumination variations.

In this chapter, the face model under varying illuminations is modeled as (5.2)

$$f(x, y) = r(x, y) \cdot e(x, y) + n(x, y) \tag{5.2}$$

In this model, an additive term n(x,y) as noise is involved besides the reflectance r(x,y) and illumination term e(x,y). In [91], the noise n(x, y) in a local area is modeled as a constant based on the local properties of human face. In this chapter, observing the noise in an entire image, we propose to apply the high frequency DCT coefficients obtained in the entire image to estimate the noise. To remove the effects of noise, the DCT is firstly applied in face images. The values of k dimensions of high frequency DCT coefficients are set to zeros in a zigzag mode. Experimental results and more discussions in Section 5.3 will be presented to demonstrate the validity of the model and our assumption.

After the denoising described in the above section, the model is simplified into

$$f^{'}(x, y) = r(x, y) \cdot e(x, y) \tag{5.3}$$

Taking logarithm transform on Eq.(5.3), we have

$$\log f^{'}(x,y) = logr(x,y) + \log e(x,y) \tag{5.4}$$

From the numerically view, the logarithm transform can compress light pixel values and expand the dark ones [1]. Moreover, the logarithmic mapping from an input $\log f^{'}(x,y)$ to the perceived sensation $\log f^{'}(x,y)$ is justified by Weber's law, proposed by Ernst Weber in 1834. From another view, the logarithm transform can convert the product into a sum which makes easier the normalization task mentioned before.

5.2.2 Local Relation Map

A human face can be treated as a combination of lots of small and flat facets [91]. In such a small facet W, the illumination e(x,y) can be considered as a constant as in [91]. Compared with the model (5.4), the $\log e(x,y)$ can be taken as a constant A. Therefore, for a special illumination condition, a small facet W can be modeled as

$$I(x,y) = \rho(x,y) + A(x,y) \in W \tag{5.5}$$

where

$$I(x,y) = \log f^{'}(x,y) \tag{5.6}$$

and

$$\rho(x,y) = \log r(x,y) \tag{5.7}$$

Based on Eq.(5.5), we propose a simple illumination invariant feature local relation map, which eliminates the effect of A by comparing the relation between the gray level of target point with those of the points in the boundary of W. The details of the LRM are described as following:

1) Given a point (x,y), determine the local facet W. In this chapter, we mainly focus on a square local facet because it is easier to implement.

2) Determine the boundary points U in the facet. For a square facet with size of n, there is only 4(n-1) boundary points.

3) Compare the gray level of point (x, y) with those of points in U as

$$I^{'}(x,y) = I(x,y) - \sum_{(a,b)\in U} I(a,b)/4(n-1) \tag{5.8}$$

4) After all the points on a given image are processed, the illumination invariant feature, named local relation map, is obtained.

In addition to the square facet employed by this work, a circular or diamond facet is an appropriate choice. Mathematically, any point within the small facet W can be taken as U. In this chapter, we only consider boundary points in the facet as U. One of the reasons is that it reduces the computational complexity to $O(n)$, compared to $O(n^2)$ when all the points in the square facet are involved in the process. The effect on real computational time will be discussed in the experimental section. Another reason is that boundary points can summarize local properties of the facet, similar to the idea in local binary pattern. Our ongoing study suggests boundary points can be used in adaptive size selection of the facet.

Next,we will demonstrate that the LRM is illumination invariant through theoretical analysis. Given two images I_1 and I_2 of the same person taken under different illumination conditions, for the same point (x, y), we have

$$I_1(x,y) = \rho(x,y) + A_1 \tag{5.9}$$

$$I_2(x,y) = \rho(x,y) + A_2 \tag{5.10}$$

After the process of the LRM, we have

$$\begin{aligned} I_1^{'}(x,y) &= I_1(x,y) - \sum_{(a,b)\in U} I_1(a,b)/4(n-1) \\ &= \rho(x,y) - \sum_{(a,b)\in U} \rho(a,b)/4(n-1) \end{aligned} \tag{5.11}$$

$$\begin{aligned} I_2^{'}(x,y) &= I_2(x,y) - \sum_{(a,b)\in U} I_2(a,b)/4(n-1) \\ &= \rho(x,y) - \sum_{(a,b)\in U} \rho(a,b)/4(n-1) \end{aligned} \tag{5.12}$$

By comparing (5.11) and (5.12), we have

$$I_1^{'}(x,y) = I_2^{'}(x,y) \tag{5.13}$$

This means that the LRM is unrelated to illumination conditions. Therefore, we can use the LRM for further face recognition.

5.3 Experimental Results and Discussions

In this section, we present the evaluation of our algorithm and other existing methods on the Yale Face database B [11], the Extended Yale Face database B [13] and CMU PIE [87]. In [24], the LTV has been proven to achieve the best performances in the Yale B and CMU PIE among several representative methods. However, the authors in [24] have not provided the comparison between the LTV and the methods proposed in [32, 91]. To make our evaluation convincing, we implement the approaches in [24, 32, 41, 91] besides our proposed method and name them as the LBP, DCT, LN, and LTV respectively for comparison. In the following experiments, the nearest neighbourhood classifier is used with the Euclidean distance.

5.3.1 Experiments on CMU PIE Database

There are altogether 68 subjects with pose, illumination and expression variations in the CMU PIE [87]. Because we are concerned with the illumination variation problem, only frontal face images (Pose 27) under different illumination variations (without expression variations) of CMU PIE are used, and there are 1428 images (21 images per person) in total. Only the images with the frontal flash (Flash 08) are chosen as gallery set (one image per subject) and the remaining images are used as probe set (20 images per subject). All the images are manually cropped, aligned and resized by [32]. To compare the illumination model (5.2) with noise with the model (5.1) without noise, we simply apply the LRM in the logarithm domain of the images without the step of denoising, and name the method as L̈RM without denoising. All the results are shown in Table 5.1.

From the table, it is easy to see that all the methods obtains good performances and our methods obtains the best performance as same as the LN. Compared to the performances in the second experiment on the Yale B and Extended Yale B, the performances of all the methods in the CMU are better. One of the probable reasons is that the CMU database contains limited illumination variations, which are similar to Subsets 1, 2 and 3 in the Yale B and Extended Yale B. Hence, it is easier to deal with illumination variations in the CMU database. Furthermore, there is no difference between the performances of our proposed two methods.

5.3.2 Experiments on Yale B and Extended Yale B Database

In the experiments, we use the Yale Face database B and Extended Yale Face database B as the test database. In the Yale Face database B, there are 10 persons

Table 5.1: Performance comparisons of different methods on CMU

Method	Error rate (%)
DCT	0.37
LTV	1.03
LN	0.22
LRM	0.22
LRM without denoising	0.22

Table 5.2: Subsets divided based on light source direction

	subset 1	subset 2	subset 3	subset 4	subset 5
Light angle	$0 \sim 12$	$13 \sim 25$	$26 \sim 50$	$51 \sim 77$	> 77
Number of images	263	456	455	526	714

with 64 different illumination conditions for nine poses per person [11]. In the Extended Yale Face database B, there are 16128 images of 28 persons with the same conditions as Yale B [13]. Because the main concern in this chapter is on illumination variation, only 64 frontal face images per person under different illumination conditions are chosen. After combining the Extended Yale B with the Yale B except 18 corrupted images, there are 2414 images of 38 subjects named as the Completed Yale B. The images are divided into 5 subsets based on the angle between the light direction and the camera axis as other methods shown in Table 5.2. All the images are manually cropped, aligned and resized by [13].

In the experiments, only one frontal image per person with normal illumination (0°light angle) is applied as gallery, and other images are used as probe sets. All the results are shown in Table 5.3. From the table, it is clear that the proposed method achieves the best total performance compared with other methods. The results demonstrate the validity of our assumption that the noise can be modeled based on high frequency components. For small illumination variations such as Subset 3, the DCT and LRM without denoising (they both only involve the illumination

Table 5.3: Performance comparisons of different methods on Complete Yale B

Method	Recognition Error Rate (%)			
	subset 3	subset 4	subset 5	total
DCT	10.5	10.8	12.6	8.1
LBP	2.0	22.4	64.4	24.3
LTV	14.1	14.1	14.6	10.0
LN	12.3	6.3	8.4	6.2
LRM	11.2	7.6	7.6	6.0
LRM without denoising	10.5	8.2	10.9	7.0

in the face model as Eq. (5.1)) obtain better performances. For large illumination variations such as Subset 4 and 5, the LN and LRM (they consider the face model as Eq. (5.2)) outperform other two methods. The comparison demonstrates that the noise does not need to consider when only small illumination variation exists, and the noise needs to be modeled as an additive term when larger illumination variation happens. Please note that although the LBP does not obtain good performances under large illumination variations, its performance under small illumination variations is the best. The reasons for the sensitivity of the LBP to different illumination conditions are: 1) when large illumination variations exist, facial images will have more shadows. Thus, an additive term should be involved in the facial model for the shadow while the LBP does not consider the additive term; 2) In the LBP, the intersection distance or chi-square distance is used for measuring the differences among histograms, which only considers the block-level,and global-level differences but ignore the differences in pixel-level. In addition, the LBP as a texture descriptor is robust to multiple variations such as pose and expression variations, which can not be handled by the above illumination normalization techniques.

5.3.3 Noise Modelling

As discussed in the last section, noise can be ignored when illumination variation is small and noise needs to be considered more seriously when larger illumination variation exists. In [91], the noise is modelled as a constant in a small neighbourhood. In this chapter, the noise is proposed to be related to high frequency components obtained in an entire image. In this section, we will compare these two different assumptions through experiments in Subsets 4 and 5 where larger illumination variations exist.

No matter which assumption is more accurate, the purposes of above two assumptions are the same as obtain the face model under varying illuminations as Eq. (5.3) after some process. To test these two assumptions, we define a denoising performance criterion (DPC) based on the following equations

$$DPC = \sum_{i=1}^{test\,person\,number} \sum_{j=1}^{different\,lighting\,conditions} \sum_{(x,y)\in I} VR(x,y,i,j) \tag{5.14}$$

$$VR(x,y,i,j) = var(PDR(x,y,i,j,a,b))(a,b) \in W(x,y) \tag{5.15}$$

$$PDR(x,y,i,j,a,b) = (I_{i,j}(a,b) - I_{i,j}(x,y))/(I_{i,1}(a,b) - I_{i,1}(x,y)) \tag{5.16}$$

Where $I_{i,j}$ is the image of the individual i under the lighting condition j (when $j = 1$, it is the normal illumination condition, so $I_{i,1}$ is the gallery image of the individual i in the database), $I_{i,j}(x,y)$ is the gray value of point (x, y) in the image $I_{i,j}$, W(x, y) is a small neighborhood of point (x, y) and $VR(x,y,i,j)$ is the variance of $PDR(x,y,i,j,a,b)$ within W.

Based on (5.3) and (5.14), if the illumination e(x,y) in a facet is a constant, we have

$$PDR(x,y,i,j,a,b) = e_{i,j}(a,b)\cdot(r_{i,j}(a,b) - r_{i,j}(x,y))/e_{i,1}(a,b)\cdot(r_{i,1}(a,b) - r_{i,1}(x,y)) \tag{5.17}$$

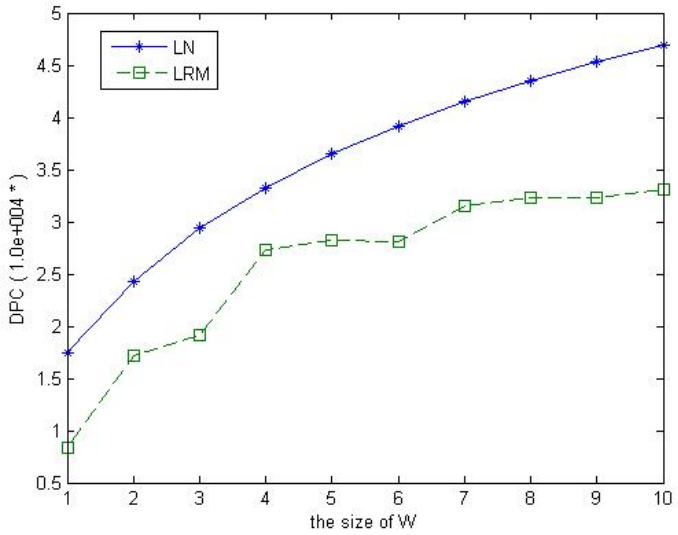


Figure 5.1: DPCs of different methods in Subset 4.

where e is the illumination term unrelated to the person i, and r is the reflectance unrelated to illumination condition j. Therefore, we have

$$\begin{aligned} PDR(x,y,i,j,a,b) &= e_j(a,b)\cdot(r_i(a,b)-r_i(x,y))/e_1(a,b)\cdot(r_i(a,b)-r_i(x,y)) \\ &= e_j(a,b)/e_1(a,b) \end{aligned} \tag{5.18}$$

Because the illumination is considered to be a constant in W, PDR(x,y,i,j,a,b) should be a constant for each (a, b) in a small neighbourhood W near the point (x, y). Based on the definitions of (5.14) and (5.15), the DPC should be zero.

From the definition of the DPC and above analysis, it is easy to find that if the DPC is smaller, the corresponding assumption on noise and the related solution is more convincing. In this section, we calculate the real DPCs of each method in subsets 4 and 5 of Extended Yale B as shown in Figures 5.1 and 5.2.

From the figures, we can see that for the same size neighbourhood, each DPC of the LRM in each subset is smaller than the corresponding DPC of the LN except that in subset 5 when the W size equals to 10. Therefore our assumption that

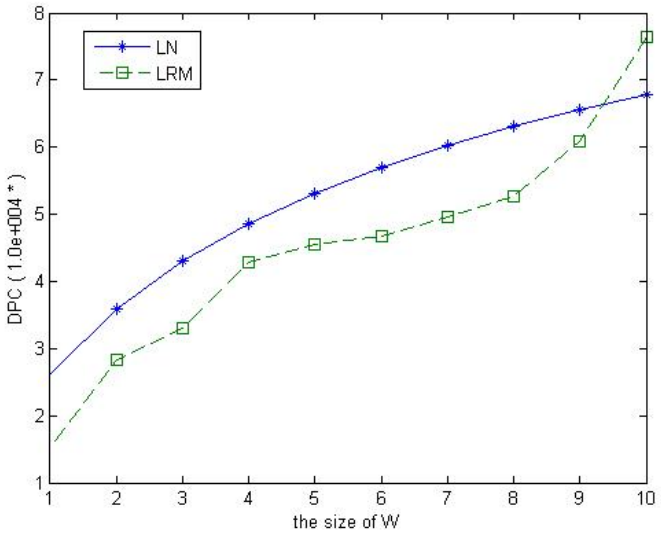


Figure 5.2: DPCs of different methods in Subset 5.

the noise should be taken as high frequency component in the entire image is more convincing.

5.3.4 Computational Complexity

Furthermore, we compare computational time of the LN and that of our proposed method because they achieve a better total recognition performance and apply similar local properties of human face. Suppose that the image size is $m \times m$ and the size of local area is $n \times n$. The real computational time is calculated with the Matlab in a personal computer with a 2.66GHz CPU. The comparison is shown in Table 5.4. For comparison, the LTV in [24] takes 17163.9 ms per image in the same condition. From the table, we can see that our method significantly reduces computational burden and speed up 64%.

Table 5.4: Comparison of computational complexities

Methods	Computational complexity	Real computational time (per image)
The LN	$O(n^2m^2)$	2.51s
The LRM	$O((\log m + n)m^2)$	0.91s

5.4 Conclusions

In this chapter, a low computation complexity face recognition approach is proposed to address the problem of illumination variations. Different from most of existing methods, an additive term is involved in the face model under varying illuminations as noise. As an assumption, we propose to zero an appropriate number of high frequency DCT coefficients to eliminate the effect caused by the noise. Experimental results demonstrate the validity of the proposed face model and assumption on the noise. Based on local characteristic of human face, a simple but effective illumination invariant feature LRM is proposed. Theoretical analysis and experimental results shows that the LRM is robust against illumination variations and achieves superior performances on the CMU PIE, Yale B and Extended Yale B. Furthermore, the proposed method shows a good computational efficiency which is important in real time applications. Further research on adaptive size selection of local area will be carried out in future.

Chapter 6

Adaptive Illumination Normalization Approach based on Denoising Technique for Face Recognition

6.1 Introduction

As a common assumption in illumination invariant face recognition, illumination usually changes slowly while the reflectance can change abruptly. As a result, illumination variations mainly lie in the low-frequency band [32]. Based on the assumption, Chen et al. [32] proposed a representative illumination normalization method using Discrete Cosine Transform (DCT). In [32], low-frequency DCT coefficients in the logarithm domain are zeroed to compensate illumination variations. However, the illumination cannot be well approximated by only low frequency components

while the reflectance does not depends on high frequency components only. Therefore, the method has trouble in keeping sharp edges in the low-frequency field and eliminating the effects in the high-frequency band caused by illumination variations. Ekenel and Stiefelhagen [94] proposed a Block-DCT based approach as an improvement of the method in [32]. The proposed method selects most reliable DCT coefficients on their classification output as features for further classification task. However, the method proposed in [94] only selects more discriminative coefficients and still cannot separate reflectance from all frequency coefficients. In [24], a logarithmic total variation (LTV) model is proposed which decomposes an input image into large-scale output u and small-scale output v. The small-scale output is regarded as illumination invariant features which can be used in further recognition task. The idea behind the model is similar to the above method but the LTV has a better edge-preserving ability. The authors in [24] compared the LTV with other representative methods and concluded that the LTV obtained the best performances in the Yale B and CMU PIE. However, the performance of the LTV is still not satisfying especially in the cases where larger illumination variations exist, as shown in the following experiments. Furthermore, the LTV is an iterative method and it takes a very high computational burden. All the above mentioned approaches need to determine the values of some important parameters by experiences, which can greatly influence their performances. In real applications, it is difficult to choose the values in different cases such as different illumination conditions and different image sizes.

In this chapter, we propose an illumination normalization approach based on denoising technique to deal with illumination variations. In the proposed method, we first extend the method in [32] from applying the DCT in a globe way to the block-DCT method. After that, we model each corresponding coefficient of blocks obtained

from the block-DCT as one Generalized Gaussian distribution (GGD) variable. For each block-DCT coefficient except the DC component, we take the illumination as main signal and take the reflectance as "noise". A soft-thresholding technique is applied to each block-DCT coefficient to remove the noise. After denoising, the illumination can be estimated based on denoised coefficients and then the reflectance is obtained indirectly. In the proposed soft-thresholding technique, the process of parameter estimation is improved so that the value of the key parameter is adaptive and data-driven. Hence the proposed method does not need to determine the value of the key parameter by experiences or experiments. The proposed method preserves more information of the reflectance in the low-frequency band instead of discarding all the information, and also eliminates the effects caused by illumination variations in the high-frequency band. Experimental results on the Yale B, Extended Yale B and CMU PIE databases show that our method obtains superior performances compared to other existing methods.

6.2 Illumination Normalization based on Denoising Technique

6.2.1 Illumination Model

In general, a face image I under different illumination conditions can be modeled as

$$I(x, y) = R(x, y)L(x, y) \tag{6.1}$$

where L(x, y) is the illumination and R(x, y) is the reflectance at each point (x, y). Because R can be regarded as illumination invariant features, the main purpose of illumination invariant face recognition is to separate R from I or obtain some

function unrelated to L.

First of all, we apply logarithm transform to Eq. (6.1) and have

$$I^{'} = R^{'} + L^{'} \tag{6.2}$$

where $I^{'} = \log(I)$,$R^{'} = \log(R)$ and $L^{'} = \log(L)$. The first reason for taking logarithm transform is that we need to convert model (6.1) from a multiplicative model to an additive one as in [32]. Another reason is that logarithm transform can compress the light pixel values and expand the dark ones [1]. Hence it can partially remove the effects caused by illumination variations.

6.2.2 Discrete Cosine Transform and Distributions of its Coefficients

In this chapter, we propose to divide an image into non-overlapping blocks with the size of 8×8. After that, the DCT is performed independently on the blocks in the logarithm domain and we have

$$Y_{ijk} = X_{ijk} + N_{ijk} \tag{6.3}$$

where Y_{ijk} is the block-DCT coefficient at position (i, j) of block k, X_{ijk} is a variable caused by reflectance and N_{ijk} is a variable resulted from illumination.

In [32], high frequency components are selected for recognition because they are considered to be insensitive to illumination variations. In another point of view, N_{ijk} of high frequency components are taken as zeroes or constants for the same object. In [94], N_{ijk} of selected frequency components instead of high frequency components are taken as zeroes or constants for the same object. The selected frequency components are the most reliable ones to perform classification. However, the existing methods cannot separate N_{ijk} or X_{ijk} from Y_{ijk} for all the coefficients.

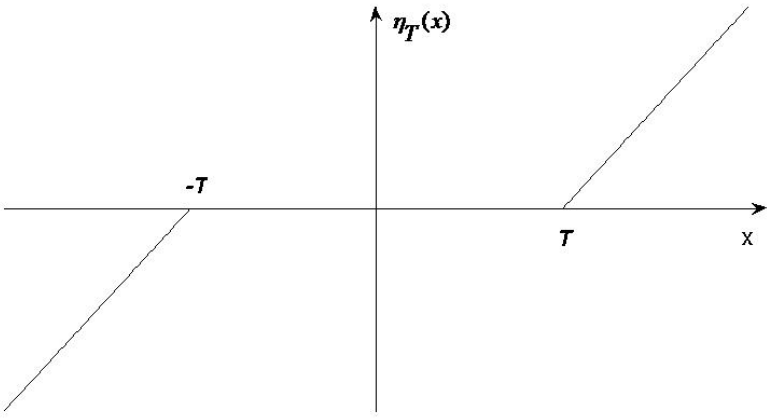


Figure 6.1: Soft-thresholding function

In the proposed method, we investigate the statistical distribution for each corresponding DCT coefficient from the blocks, which means the coefficients at the same position (i, j) but from different blocks k. Generally, the Laplacian distribution is a popular choice because of its simplicity and fidelity to the empirical data [95–98]. The Generalized Gaussian distribution (GGD) can result in an equal or better fit to the empirical data than Laplacian, yet at the expense of more complex parameters [99–103]. Therefore, we can summarize each corresponding DCT coefficient Y_{ij} of blocks (except the DC coefficient) in face images by different GGDs. Based on the distribution of DCT coefficients, a proposed soft-thresholding denoising technique, which will be presented in details in the next section, is applied to recover the reflectance from Y_{ijk} affected by illumination.

6.2.3 Soft-thresholding Technique

Soft-thresholding Techniques are popular in image denoising. The soft-thresholding function is given by

$$\eta_T^x = sgn(x) \cdot \max(|x| - T, 0) \tag{6.4}$$

where x is the signal and T is the threshold, shown in Figure 6.1. There have been several works [104–110] on determining the threshold in wavelet domain, but seldom in the DCT domain. BayesShrink [110] is one of the most representative works in wavelet domain. In [110], a data-driven estimation of the threshold which is adaptive to different GGDs is given by

$$T = \begin{cases} \max(|Y|), & if\ \sigma_x = 0 \\ \frac{\sigma^2}{\sigma_x}, & otherwise \end{cases} \tag{6.5}$$

$$\sigma_x = \sqrt{\max(\sigma_y^2 - \sigma^2, 0)} \tag{6.6}$$

where Y is the observed signal, σ_y^2 is the variance of Y and σ^2 is the estimation of noise variance. From the above equations, it can be seen that the noise variance is important to determine the threshold T. In [110], the noise variance is estimated from the HH band by the robust median estimator. However, the robust median estimator is not effective in our work. One of probable reasons is that it is difficult to determine which block-DCT components are dominated by the noise.

In the proposed denoising technique, we model illumination as main signal while reflectance is considered to be "noise". There are two main reasons: 1) illumination variations have a greatly influence on the DCT coefficients especially low frequency coefficients; 2) the noise variance is one of the most important parameters. If the illumination is modeled as noise, it is difficult to estimate the variances of illuminations from gallery set which does not contain sufficient varying illuminations normally. While the reflectance is modeled as the noise, the information provided by gallery set is enough to estimate the variances of reflectance.

In our work, the Block-DCT is performed in each image of gallery set. The corresponding DCT coefficients of blocks in each image are collected and the variance of each DCT coefficient is calculated the noise variation σ^2. With the σ^2, we can calculate the threshold T for each DCT coefficient by (6.5) and (6.6). After the threshold T is obtained, we can perform soft-thresholding in each DCT coefficient using (4).

The threshold is data-driven and adaptive to both input signal and noise. When $\sigma/\sigma_x << 1$, which means the signal is much stronger than the noise, the threshold is chosen to be small. As a result, most of signal information remains and only little noise is removed. While $\sigma/\sigma_x >> 1$, meaning that the noise is dominant, a larger threshold is chosen and most of signal information is removed as noise. In the case $\sigma_y^2 \leq \sigma^2$, σ_x is taken to be zero, which means that there is no useful information in the observed signal except noise. Therefore, the threshold T is chosen as the maximum absolute value of the observed signal and all the signals are zeroed by using (6.4).

6.2.4 Summary of our Proposed Method

To be better understood, the proposed method is summarized as follows:

For the gallery set:

1) Take logarithm transform for each image and obtain $I^{'}$.

2) Divide the image into blocks and perform the DCT in each block independently.

3) Collect the corresponding coefficients from the blocks, calculate the variance of each DCT coefficient (except the DC component) as σ^2 and calculate the threshold

T by using (6.5) and (6.6).

4) For each DCT coefficient of each block (except the DC component), perform soft-thresholding by using (6.4). For the DC component, we keep the value of the coefficient unaltered.

5) Apply the inverse DCT in each block and take the result as the illumination $L^{'}$.

6) Subtract $L^{'}$ from $I^{'}$ and obtain $R^{'}$.

For the probe set, the process is as the same as that for the gallery set without step (3).

From the process, it can be seen that there is no parameter whose value needs to be determined by experiences. All the parameters are adaptively determined by the gallery set and the probe set. Therefore, the method is data-driven and adaptive. In fact, the size of blocks is not compulsory as 8×8 as in the JEPG and it can be chosen as any size as if $M/N >> N$ where $M \times M$ is the image size and $N \times N$ is the block size.

In previous works such as [32], low-frequency coefficients are zeroed to remove the effects caused by illumination variations and the number of low-frequency coefficients is determined by experience, which has a great influence on the performance. In our proposed method, we can also choose several DCT AC coefficients as low-frequency components and do not perform soft-thresholding in these DCT coefficients. In other words, the illumination is estimated by using both the low-frequency components without soft-thresholding and the high-frequency components processed by soft-thresholding. However, it is difficult to determine the number of low-frequency coefficients perfectly for different cases such as different illumination variations and different image sizes. Therefore, in our work, we do not determine which components

are low-frequency and which are not. The data-driven soft-thresholding is performed in each DCT coefficient except the DC component because it is not adapted to Laplacian distribution. Therefore, the proposed method is completed data-driven and all the parameters are adaptive to input images under any case. Furthermore, the method can preserve more information of the reflectance in the low-frequency band instead of discarding them all. The experiments shown in the next section prove that the proposed method obtains superior results compared to other existing methods, even without any prior information or experience.

6.3 Experimental Results and Discussions

In this section, we evaluate the proposed method by several different experiments on different databases. In [24], the LTV has been proven to achieve the best performances in the Yale B and CMU PIE among several representative methods. In our experiments, we compare the performances of our proposed method with the methods in [24,32,41,54,94] on the Yale B, Extended Yale B and CMU PIE. Because our method is implemented based on the Block-DCT, we also compare our method with the Block-DCT. In the experiments, the nearest neighborhood classifier is used and the Euclidean distance is used as distance measure.

6.3.1 Experiments on the Yale B and Extended Yale

The Extended Yale Face Database B contains 38 subjects under 64 different illumination conditions for nine poses per subject. Because the main concern in this chapter is illumination variations, only frontal images under 64 illumination conditions are chosen as samples. All images are manually aligned, cropped, and then

Table 6.1: Performance comparison using subset 1 as gallery set

Methods	Recognition rates (%)			
	Subset2	Subset3	Subset4	Subset5
LBP	100	100	72.24	51.54
LTP	100	100	77.76	47.48
DCT	100	100	96.96	95.66
Block DCT	100	100	95.06	94.54
LTV	100	100	95.82	94.12
Anto. Freq. Sel. [94]	100	100	98.7	99.0
Our method	100	100	98.48	98.32

re-sized to 192×168 images by [13]. The images are divided into 5 subsets based on illumination angle, i.e., subset 1(0-12), subset 2(13-25), subset 3(26-50), subset 4(51-77) and subset 5(>77).

In this first experiment, all the images of subset 1 are used as gallery set, and the other images from subsets 2 to 5 are used as probe sets. Please note that there are 7 different illumination conditions in subset 1, and therefore there are 7 images per subject in the gallery set in the experiment. The results are shown in Table 6.1. From the table, we can see that all the methods obtain 100% recognition rates in the cases where small illumination variations exist such as subset 2 and 3. However, recognition performances decrease when larger illumination variations happen such as subsets 4 and 5. As can be seen, the proposed method achieves better performances compared to the other methods but be slightly inferior to the automatic frequency band selection method proposed in [94]. When smaller block size is applied in our method, the performances of our method and the method in [94] are almost the same as shown in Table 6.5. One of the advantages of our method is that there is no parameter value to be determined by experiences in our method, compared to the important parameter of window size mentioned in [94].

Table 6.2: Performance comparison using each image of subset 1 as gallery set

Methods	Average recognition rates (%)			
	Subset2	Subset3	Subset4	Subset5
DCT	100	97.11	89.3	91.06
Block-DCT	100	97.93	90.25	91.32
LTV	100	95.2	86.8	91.08
Our Method	100	98.46	90.87	94.26

For some applications, the number of images per subject in gallery set is limited. Therefore, we carry out the second experiment to compare the performances of the above methods when only one frontal image per subject is used as gallery set. Each image under different illumination conditions from subset 1 is selected as gallery set, and the other images from subsets 2 to 5 are used as probe sets. Because there are 7 different illumination conditions in subset 1, we repeat the experiment 7 times and the average recognition rates are shown in Table 6.2. From the table, we can see that the performances of all the methods drop significantly especially in subsets 4 and 5, compared to the performances in the first experiment. As can be seen, our proposed method still achieves better performances than the other methods, but the performances in subsets 4 and 5 still have large potential to improve.

6.3.2 Experiments on the CMU PIE

The CMU PIE dace database is comprised of 68 subjects with different poses, illuminations and expressions [87]. In the database, there are 21 different lighting conditions for each pose, labeled as Flashes 2 to 22. In our experiment, only frontal face images (Pose 27) under different illumination variations (without expression variations) are used, and there are 1428 images (21 images per person) in total. In this experiment, we choose only the images with the frontal flash (Flash 08) as

Table 6.3: Performance comparison on CMU

Methods	Error Number	Recognition Rate(%)
DCT	15	98.90
Block-DCT	15	98.90
LTV	26	98.09
Our Method	15	98.90

gallery set (one image per subject) and the remaining images are used as probe set (20 images per subject). The recognition rates are listed in Table 6.3, which shows that all the methods obtains good performances and our method obtains the best performance as same as the DCT and the Block-DCT. Compared to the performances in the second experiment on the Completed Yale B, the performances of all the methods in the CMU are better. One of the probable reasons is that the CMU database contains limited illumination variations, which are similar to subsets 1, 2 and 3 in the Completed Yale B. Hence, it is easier to deal with illumination variations in the CMU database.

6.3.3 Computational Time Comparison

In this section, we compare the real computational time of our proposed method with those of the existing approaches on the Extended Yale B. The real computational time is calculated with the Matlab in a personal computer with a 2.66GHz CPU and the size of images is 192×168. The results are shown in Table 6.4. From the table, we can see that the DCT is the fastest and our method is comparable to the Block-DCT, while the LTV is the slowest. However, considering the performance improvement and the simplicity of choosing parameters, the proposed method is worth the additional cost of computational time and it is still applicable in real time applications.

Table 6.4: Average process time of each method

Methods	Average Processing Time per Image (ms)
DCT	269.2
Block-DCT	385.4
LTV	17163.9
Our Method	415.6

6.3.4 Performance Comparison for Different Block Sizes

In this section, we investigate whether our proposed method is insensitive to block size. The performances for different block sizes on the Extended Yale B are shown in Table 6.5. As we mentioned in Section 6.2.4, the block size should satisfy $M/N >> N$ where $M \times M$ is the image size and $N \times N$ is the block size. From the table, we can see that the performances of our method are stable good when block sizes satisfy the condition. While the block size is quite larger ($168/12 \approx 12$), the performance drops but still outperforms the Block DCT and the LTV. However, we find that the performances, where block size is 12×12, can be improved to 98.10% (Subset 4) and 98.60% (Subset 5) if two more DCT coefficients are discarded besides the DC component. Based on the experimental results, we can claim that our method can provide stable good performances under varying illuminations when the block size statisfies the mentioned condition. Even if the block size is determined inappropriate, the performances are still acceptable and can be improved by discarding more DCT coefficients.

6.4 Conclusions

In this chapter, a novel illumination normalization approach based on denoising technique is proposed for face recognition under varying illumination conditions.

Table 6.5: Performance comparison for different block sizes

Block size	Recognition rates (%)			
	Subset2	Subset3	Subset4	Subset5
4×4	100	100	98.86	98.88
6×6	100	100	98.86	98.88
8×8	100	100	98.48	98.32
12×12	100	100	95.44	96.36

The proposed approach can effectively estimate the illumination using an adaptive soft-thresholding technique and thus the reflectance can be obtained indirectly. Compared to the existing methods, our method has the following advantages: 1) the proposed method preserves more information of the reflectance in the low-frequency band instead of discarding them, and removes the effects caused by illumination variations in the high-frequency band, 2) no prior information is need such as light condition assumption or large number of training samples, 3) the key parameter in denoising is data-driven and adaptive, therefore we does not need to determine its value by experiences, 4) the method outperforms most of the existing methods and still has a fast speed which is applicable in a real-time system.

Chapter 7

Robust Face Recognition with Local Line Derivative Pattern

7.1 Introduction

In face recognition, one major issue is to find features of human face which are not only effective in controlled environments, but also insensitive to different variations as much as possible. In recent years, local features have gained many attentions due to their robustness against uncontrolled variations such as expression and illumination variations.

Local Binary Pattern (LBP) proposed by Anonen et al. [40, 41] is one of the most successful local features in face recognition. As a texture feature, the LBP is proven to be robust against pose and expression variations. Furthermore, the LBP is also insensitive to monotonic gray-level variations resulted from varying lighting conditions. Due to these advantages, the LBP has gained many attentions and several important variants of the LBP have been proposed. The authors in [48]

proposed Improved Local Binary Pattern (ILBP) which used the mean value of all neighbourhoods as threshold instead of the gray-level value of central point. Therefore, a binary label string of the ILBP has one more bit compared to the LBP with the same shape and size. Petpon and Srisuk [43] proposed local line binary pattern (LLBP). The LLBP applies a straight line as its neighbourhood shape instead of a square or circle shape in the LBP. Furthermore, the weights at left and right side of the central points are mirror so that the number of patterns is reduced. However, pattern labels will be same even if actual binary strings are different due to the mirror weights. In [42], a Local Ternary Pattern (LTP), another important extension of original LBP, was 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 to make the LTP more discriminant and less sensitive to noise. The value of parameter t has a great influence on the performance of the LTP. When the t is too small, the LTP has no improvement of robustness against noise and can be regarded as the LBP. While the t is too large, the LTP will lose too much local information and the performance will also decrease. In [56], Jabid et al. proposed local directional pattern which captures the k most prominent edge directional information by comparing edge response in all eight directions.

The LBP captures the first-order derivative pattern in all directions. However, high-order descriptor can provide more detailed and more discriminant information. In [44], a novel local feature descriptor is proposed, namely Local Derivative Pattern (LDP). The LDP obtains high-order derivative images along four directions first and compares the derivative value of central point with those of neighbourhoods to generate binary strings. Similar to the LBP, binary strings are encoded into integral labels whose distribution are modelled by histograms to perform recognition

task. Based on the experimental results, the third-order LDP can achieve the best performances.

In this chapter, a novel local pattern descriptor is proposed, namely Local Line Derivative Pattern (LLDP). In the LLDP, Sobel Masks are applied in images to obtain high-order derivative images. By comparing the value of central point with those of neighbourhoods in a line shape, a binary string is obtained. Each bit in the string will have a unique weight in a way different from previous works. After that, the string will be encoded into an integral label. Spatial histogram is used to model the distribution of labels for recognition. A novel distance measuring both pixel-level information and global-level information is also proposed to improve the performance of the LBP as well as its several variants under different lighting conditions. The experimental results on the FERET and Extended Yale Face database B demonstrate that the proposed LLDP achieves superior performances compared to the LBP as well as its several representative variants. The proposed LLDP approach has been introduced in [111] and the proposed improved distance measurement has been presented in [112].

7.2 Local Line Derivative Pattern

It contains three steps to calculate the LLDP label for a given images. In this section, these three steps of the LLDP are described in details, as well as a spatial histogram as features for face recognition.

7.2.1 High-order Derivative Images

In [44], high-order derivative images are claimed to have more detailed and more discriminant information. In the LLDP, we also calculate high-order derivative images first. Different from the LDP, we apply Sobel masks to calculate high-order derivative images. The Sobel masks are defined in the x-direction and the y-direction as shown in Figure 7.1. In the LDDP, high-order derivative images only have two directions, along 0° (y-direction) and 90° (x-direction). The nth-order derivative image along the y-direction and nth-order derivative image along the x-direction can be obtained respectively by

$$I_x^n = M_x * I_x^{n-1} \tag{7.1}$$

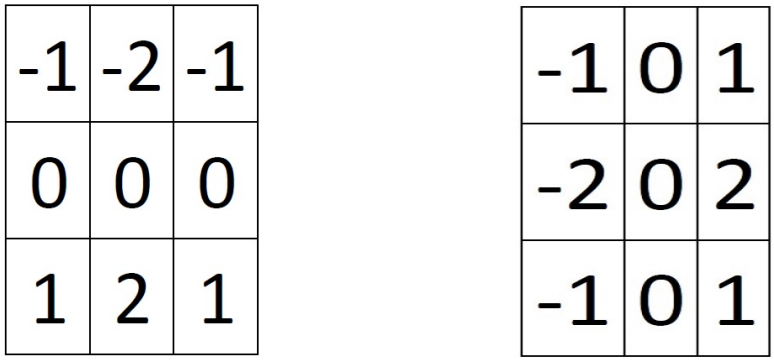
$$I_y^n = M_y * I_y^{n-1} \tag{7.2}$$

where I_x^n is the nth-order derivative image along the x-direction, I_y^n is the nth-order derivative image along the y-direction, M_x is Sobel mask along the x-direction, M_x is Sobel mask along the y-direction, and $*$ is the convolution operator. The original image can be regarded as the 0th-order derivative image along both directions. Please note that only two derivative images are obtained here instead of four derivative images in the LDP. Thus, the number of features can be reduced half compared to the LDP.

7.2.2 Binary Coding Function

In the LDP [44], the binary coding function is defined as

$$s(x,y) = \begin{cases} 1, & if x \cdot y > 0 \\ 0, & otherwise \end{cases} \tag{7.3}$$



(a) Along the x-direction

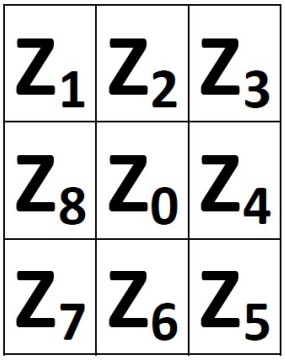
(b) Along the y-directon

Figure 7.1: Sobel masks

where x is the value of central point in the derivative image and y is the value of neighbourhood in the derivative images. From the function, it is clear that the function will encode 1 if there is a change of derivative directions of two points. However, if both points have the same derivative value of zero which means there is no any change of derivative directions in these two points, the function will still encode 1. Therefore, we revise the binary coding function as

$$s(x,y) = \begin{cases} 1, & if(x \cdot y > 0)\ or\ ((x = 0)\,and\,(y = 0)) \\ 0, & otherwise \end{cases} \tag{7.4}$$

After obtaining the nth-order derivative images, the LLDP generates binary strings for each point by comparing the value of central point with those of neighbourhoods in a line shape using (7.4). Similar to Sobel masks, the LLDP also have two directional line neighbourhoods, one of the x-direction as shown in Figure 7.3(a) and the other of the y-direction.



(a) 3×3 Neighbourhood

128	64	32
1	0	16
2	4	8

(b) Weights of each point

Figure 7.2: 3×3 neighbourhood and weights of each point in the neighbourhood

7.2.3 Label Coding and Histogram

All the above mentioned works model the distribution of patterns by spatial histogram as final features for the further recognition task, since spatial histogram is more robust against pose and illumination variations than holistic methods [41]. For spatial histogram, there are two issues to address: one is to code binary strings into decimal integral labels and the other is to select an appropriate number of bins. For the LBP and other extension works, they encode the labels as

$$L(Z_0) = \sum_i V(Z_i)W(Z_i) \tag{7.5}$$

where $L(Z_0)$ is the label of central point Z_0, $Z_i$is the point belonging to the neighbourhood of Z_0, $V(Z_i)$ is a binary value of point Z_i obtained from binary code function, and $W(Z_i)$ is the weights of point Z_i. For a 3×3 neighbourhood illustrated in Figure 7.2(a), we normally define their weights as shown in Figure 7.2(b).

When the number of bins equals to the number of possible pattern labels, there is no question on the arrangement of weights. However, in the case that the number

of bins is smaller than the number of possible patterns, which is quite normal in real applications for reducing the number of features, there are some questions on such weights arrangement. For example, for the LBP with a 3×3 square neighbourhood, there are 256 patterns. When 128 bins are selected for histogram, it will ignore the differences between labels in the bit . In this section, we propose some standards to instruct how to arrange the weights and apply the standards to obtain the weights of line neighbourhood. The standards are defined as follows:

- Uniqueness

  Different binary strings should generate different integral labels. Therefore different points in the neighbourhood should have different weights. Based on this requirement, the weights used in the LLBP are not suitable because the LLBP will obtain the same label even binary strings are different.

- Priority

  The weights of points nearer to the central point should be larger while the weights of points away from the central point should be smaller. The behind idea is that the points near the central point are more important than those away from the central point. The standard may not be important to a square or circle neighbourhood because each neighbour point has the same distance from the central point, but it is important to a line neighbourhood.

- Balance

  Because of the requirement of uniqueness, the weights cannot be mirror so that they cannot be in perfect balance. However, the weights should be arranged in balance as much as possible. The requirement of balance is related to the number of bins. Considering the line neighbourhood as in Figure 7.3(a), if the number of bins is 64, it means the differences in both the bit of weight

Z_1	Z_2	Z_3	Z_4	Z_0	Z_5	Z_6	Z_7	Z_8

(a) A line neighborhood with 9 pixels

1	2	4	8	0	16	32	64	128

(b) Normal weights for the line neighborhood

1	8	16	128	0	64	32	4	2

(c) Proposed weights for the line neighborhood

Figure 7.3: A line neighbourhood of 9 pixels with different weights

equalling to 1 and the bit of weight equalling to 2 will be ignored. If we do not consider the balance, the weights normally are set as Figure 7.3(b). In that case with 64 bins, it is clear that the differences in the right part are more important.

With these three standards, we propose our weights arrangement as shown in Figure 7.3(c). The weights for the line neighbourhood in the y-direction are similar to those in the x-direction.

In this chapter, we choose spatial histogram as feature representation. The LLDP labels in the x-direction and in the y-direction are calculated first based on the image convolved by Sobel Masks. The labelled images are divided into blocks. For each block, two histograms are calculated for the two LLDP labelled images in different directions respectively. All the histograms of different blocks and directions are concatenated into an entire histogram which will be used as face features for

recognition.

7.3 Proposed Distance Measurement

In the classification stage, most of the woks based on the LBP employ spatial histograms as facial features and use Chi-square distance or histogram intersection to measure the distance between the histograms defined as (7.6) and (7.7) respectively.

$$\chi^2(Q,S) = \sum_i \frac{(Q_i - S_i)^2}{Q_i + S_i} \tag{7.6}$$

$$D(Q,S) = \sum_i \min(Q_i, S_i) \tag{7.7}$$

where Q is the concatenated histogram for input image A, and S is the concatenated histogram of image B. Spatial histograms are considered to be more robust against variations in pose or illumination. However, spatial histograms are still sensitive to large illumination variations which can be proven by the experiments shown in the next section. One of the probable reasons is that no matter which distance measurement is used, histograms only contain information on a regional or global level and pattern information on a pixel-level contained in a label is ignored.

In this section, a novel distance function is proposed to measure the differences between images not only in a global level but also in a pixel level. The distance in a global level between two images D_1 is measured by histogram intersection distance as (7.7), while the difference in a pixel-level between images D_2 is measured by the percentage of pixels with different labels in two images, defined as follows.

$$D_2(A,B) = \sum_{i=1}^{M}\sum_{j=1}^{N} Z(label_A(i,j),\, label_B(i,j))/(M \times N) \tag{7.8}$$

$$Z(x,y) = \begin{cases} 0, & if\ the\ hamming\ distance\ between\ x\ and\ y\ is\ smaller\ than\ T \\ 1, & otherwise \end{cases} \tag{7.9}$$

where the size of images is $M \times N$, $label_A(i, j)$ is the decimal label of point (i, j) in the image A, $label_B(i, j)$ is the decimal label of point (i, j) in the image B, and T is a threshold parameter.

The decimal labels of pixels in the LBP are not stable in the cases where some noise exists. To tackle the problem, several variants of the LBP have been proposed to extract stable labels when noise exists [42,56]. From another point of view, a novel distance measurement is proposed to obtain a stable distance between labels which may contain some small differences instead of trying to extract stable labels. In our proposed distance, two decimal labels are taken as the same when the hamming distance between them is not greater than T. It is easy to see that when T is defined as 0, two decimal labels are regarded as the same if and only if they are exact the same.

We define the final distance between two images A and B as

$$D_(A, B) = \alpha D_1(Q, S) + (1 - \alpha) D_2(A, B) \tag{7.10}$$

where Q and S are two concatenated histograms of image A and B respectively, α is the factor ranged from 0 to 1. With the proposed distance, the performances of the LBP and its variants can be improved significantly shown in the following experiments.

7.4 Experiments and Discussions

In this section, we present the evaluation of our algorithm and other existing methods on the FERET [113] and Extended Yale B database [13]. To make the evaluation convincing, we implement the approaches in [41–44, 56] and name them as the LBP, LLBP, LTP, Local Directional Pattern and LDP respectively. The near-

est neighbourhood classifier is employed with Chi-square distance as the distance measurement because all the methods apply spatial histograms as feature representation.

Furthermore, we compare the proposed novel distance with common histogram intersection for face recognition under illumination variations. In the test, the LLDP and the works in [41, 56] are implemented with different distance measurements for comparison on Extended Yale B database.

7.4.1 Experiment Databases

In the experiment, five subsets of the FERET containing expression, illumination and aging variations are used. These subsets are described as follows:

- fa set as gallery set contains 1196 frontal images of 1196 subjects.
- fb set as one of probe sets contains 1195 images with facial expression variations.
- fc set as one of probe sets contains 194 images with illumination variations.
- dupI set as one of probe sets contains 722 images taken later in time between one minute to 1031 days.
- dupII set as one of probe sets contains 234 images which is a subset of dupI taken at least after 18 months.

The other database Extended Yale Face B contains 38 subjects under 64 different illumination conditions for nine poses per subject. In this chapter, only frontal images under 64 illumination conditions are chosen as samples. The images are

Table 7.1: Performance comparison on FERET

Methods	Recognition Rate (%)			
	fb	fc	dupI	dupII
LBP	88.62	55.67	60.53	41.03
LLBP	86.11	56.70	64.13	57.26
LTP	88.20	54.12	61.50	42.31
Local Directional Pattern	76.74	32.47	48.61	26.07
LDP (3rd-order)	82.68	70.62	61.50	58.55
LLDP (3rd-order, bins=256)	93.22	96.39	78.67	78.63

divided into 5 subsets based on illumination angle, i.e., subset 1(0-12), subset 2(13-25), subset 3(26-50), subset 4(51-77) and subset 5(> 77).

All images from the FERET and Extended Yale B are manually aligned, cropped and resized to 96×84 according to the coordinates of two eyes.

7.4.2 Experiment on the FERET

In this experiment, we evaluate the proposed LLDP comparing to other existing methods. As mentioned before, the FERET contains images with expression, illumination and aging variations. Therefore the evaluation on the FERET can provide a convincing benchmark of evaluated approaches for their robustness for against expression, illumination and aging variations. The performances of all the methods are listed in Table 7.1. From the table, it is clear to see that the proposed LLDP outperforms other existing methods in each probe set. The performances prove that the LLDP has a good robustness against expression and illumination variations, but it is a little sensitive to aging variations.

Table 7.2: Performance comparison on Extended Yale B

Methods	Recognition Rate (%)			
	Subset2	Subset3	Subset4	Subset5
LBP	100	100	72.24	51.54
LLBP	100	100	65.97	48.60
LTP	100	100	77.76	47.48
Local Directional Pattern	98.03	98.46	61.03	40.76
LDP (3rd-order)	100	100	67.87	19.33
LLDP (3rd-order, bins=256)	100	100	89.54	91.32

7.4.3 Experiment on the Extended Yale B

In this experiment, we focus on the evaluation of the robustness of all the approaches against illumination variations. All the images of subset 1 are used as gallery set, and the other images from subset 2 to 5 are used as probe sets respectively. The results are shown in Table II. From the table, we can see that all the methods except Local Directional Pattern obtain 100% recognition rates in the cases where small illumination variations exist such as subset 2 and 3. It proves that these patterns based on local properties of human face are insensitive to small illumination variations. However, when larger illumination variations happen such as subset 4 and 5, recognition performances of all the methods decrease significantly. Although the LLDP obtains much better performances in Subsets 4 and 5 compared to other existing methods, the performances of the LLDP still have some potential to improve. Actually these results are natural because local properties of human face do not exist when shadows happen. Therefore the methods based on local properties cannot obtain perfect performances in such cases.

7.4.4 Comparison between Different Weights

Furthermore, we compare the performances of the LLDP with the proposed weights with those with normal weights. The FERET is used as the test database and the results are shown in Table 7.3. Normal weights are set as Figure 7.3 (b). It is easy to see that both weights obtain the exact same performances when the number of bins equals to the maximum value of possible patterns. However, it is quite common in real applications that the number of bins of each direction is much smaller than the maximum value of possible patterns, because the dimension of features can be greatly reduced in the case. Given an image with the size of 96×84 and the block size is set as 8×8, the dimensions of LLDP features for different bin numbers are shown in Table 7.4. Therefore, it is better to set the number of bins as 64 instead of the maximum value of possible patterns in practical applications considering computational burden. In this case, we can see that the proposed weights outperform normal weights in each subset except dupI from Table 7.3.

7.4.5 Performance Comparison for Different Distance Measurements

The proposed distance measurement is compared with common histogram intersection in this section. The LLDP and other works proposed in [41,56] are implemented with different distance measurements. The DCT in [32], a successful illumination invariant face recognition method, is also implemented for comparison. In the experiment, only one frontal image per person with the most neutral illumination (0°light angle) is used as gallery set, which increases the difficulty for face recognition under varying illuminations. All the results are shown in Table 7.5.

Table 7.3: Performance comparison for different weights

Methods	Recognition Rate (%)			
	Subset2	Subset3	Subset4	Subset5
LLDP Proposed weights (3rd-order, bins=256)	93.22	96.39	78.67	78.63
LLDP Proposed weights (3rd-order, bins=64)	91.63	93.30	67.17	65.38
LLDP Normal Weights (3rd-order, bins=256)	93.22	96.39	78.67	78.63
LLDP Normal Weights (3rd-order, bins=64)	89.79	86.60	69.11	63.68

Table 7.4: LLDP features dimensions for different bin numbers

Different Bin Numbers	LLDP Features Dimensions
256	64512
128	32256
64	16128
32	8064

From the table, we can see that the LLDP and the LBP are not robust enough against larger illumination variation such as Subsets 4 and 5, although they are insensitive to small illumination variations as Subsets 2 and 3. The Local Directional Pattern is also sensitive to larger illumination variations and its performances are even worse than the performances of the LBP in Subsets 3 and 4. The reason is that histogram intersection distance only measures the differences in a global level while the differences in a pixel level are ignored. From the results, it is shown that common distances based on histograms cannot extract sufficient discriminant information in the cases with larger illumination variations although they could measure

Table 7.5: Performance comparison of different distances

Methods	Recognition rate (%)			
	Subset2	Subset3	Subset4	Subset5
LLDP with histogram intersection	100	96.3	75.3	72.3
LBP with histogram intersection	99.6	91.0	35.7	9.5
Local Directional Pattern with histogram intersection	99.8	76.9	26.6	16.4
LLDP with our distance measurement	100	98.5	85.0	88.1
LBP with our distance measurement	99.8	96.3	82.9	49.3
Local Directional Pattern with our distance measurement	100	95.8	79.5	56.9
The DCT	100	89.5	89.2	87.4

facial features well in the cases with small illumination variations. Different from common histogram distances, our proposed distance measures not only the differences in a global level but also in a pixel level. It is clear that the performances of the LLDP and LBP are improved greatly with the proposed distance measurement. In Subset 3, our proposed distance improves both the performances of the LLDP and the LBP which are even better than the performance of the DCT. In the cases with larger illumination variations such as Subsets 4 and 5, the LLDP with proposed distance measurement obtains comparable performances to the DCT while the proposed distance measurement also improves the performances of the LBP and Local Directional Pattern obviously. Therefore, it can be seen that compared to the histogram intersection distance, our distance measurement is more suited to the LBP and its variants for face recognition under varying lighting conditions, especially in the cases with larger illumination variations.

7.5 Conclusion

This chapter proposes a novel local face feature descriptor, named Local Line Derivative Pattern (LLDP), for robust face recognition against expression, illumination and aging variations. The advantages of the LLDP are as follows: 1) it applies high-order derivative information; 2) it improves the binary coding function for derivative images; 3) it proposes three standards for weights arrangement which will help future works select appropriate weights for label coding; 4) the proposed distance can improve the performance of LBP-based face recognition approaches under varying illuminations. The results on the FERET and Extended Yale B prove that the LLDP outperforms other existing methods. However, it still has some potential to improve for the cases with aging and larger illumination variations.

Chapter 8

Conclusions and Recommendations for Future Work

8.1 Conclusions

In this thesis, a few high speed approaches for illumination invariant face recognition have been proposed and discussed. The methods on discarding low frequency coefficients in different domains have been proven to be a promising way to deal with the problem of varying illuminations. According to this idea, a low computation complexity face recognition method has been proposed to address the problem of illumination variations based on block-wise WHT. Low-frequency WHT coefficients are zeroed to compensate illumination variations. The method makes a good compromise between computational burden and recognition accuracy. Experimental results on the CMU PIE, Yale B and Extended Yale B database show that the method can achieve performances comparable to the representative DCT method

with a higher efficiency. For some cases with larger illumination variations, the proposed method even outperforms the DCT. It has been proven that the PCA and NLDA can be directly implemented in the WHT domain to further reduce the computational burden.

Similar to the first approach, another novel illumination compensation method applying the DCT instead of the WHT has been developed in this thesis. Different from existing approaches, the proposed method estimates illumination in two steps. Firstly, low-frequency DCT coefficients obtained in a local area are used to estimate illumination of each point in the logarithm domain of images instead of illumination compensation in a global way. After that, a mean operator is applied to refine estimation of illumination at each point. Besides, a simplified version of the method is also proposed. Both theoretical analysis and experimental results demonstrate the validity and high computational efficiency of the simplified version. The results also demonstrate that the assumption that the illumination in an appropriate small local area is constant is more convincing

Furthermore, a simple but effective illumination invariant feature LRM has been proposed according to local properties of human faces. It has been shown that the proposed feature LRM is insensitive to illumination variations and achieves superior performances on the CMU PIE database, Yale Face databased B and Extended Yale Face database B with a good computational efficiency. Moreover, we investigate the face model under varying illuminations and model noise as an additional additive term in the face model. Based on an proposed assumption, an appropriate number of DCT coefficients in the high-frequency band are discarded to remove the effects resulted from noise. Experimental results validate the proposed face model as well as proposed assumption on the noise.

Although the methods on discarding low frequency coefficients in different do-

mains have achieved good performances under varying illuminations, their performances under larger illumination variations are still not satisfactory. One of main reasons is that the methods have trouble in keeping sharp edges in the low-frequency field and eliminating the effects in the high-frequency band caused by illumination variations. In the thesis, a novel illumination normalization approach based on denoising technique has been proposed. The proposed approach can effectively estimate the illumination using an adaptive soft-thresholding technique and thus the reflectance can be obtained indirectly. Compared to existing approaches, our method has advantages as following : (1) the proposed method preserves more information of the reflectance in the low-frequency band instead of discarding them, and removes the effects caused by illumination variations in the high-frequency band, (2) no prior information is need such as light condition assumption or large number of training samples, (3) the key parameter in denoising is data-driven and adaptive, therefore we does not need to determine its value by experiences, 4) the method outperforms most of the existing methods and still has a fast speed which is applicable in a real-time system.

Since illumination variations are usually coupled with other variations in complex environments, approaches robust against both illumination variations and other variations are also investigated in the thesis so as to achieve good recognition performance under multiple variations. A novel local face feature descriptor named Local Line Derivative Pattern (LLDP) has been proposed where extracts high-order derivative information. A binary coding function for derivative images is improved in the LLDP and three standards for weights arrangement are proposed to help future works select appropriate weights for label coding. A novel distance measuring both pixel-level information and global-level information is also proposed to improve the performance of LBP-based face recognition approaches under varying

illuminations. The results on the FERET and Extended Yale B show that the LLDP outperforms other existing methods and is robust against expression, illumination and aging variations.

All the proposed methods have the advantages of low computational burden and easy implementation, especially the WHT-based approach and the LRM, which are important for the implementation of face recognition systems in practice. Furthermore, the proposed denoising approach can adaptively determine the values of key parameters, which can further reduce the difficulty of implementation in real face recognition systems. Moreover, no training step is involved in all these approaches except the denoising method and they do not need multiple images per subject in gallery. Therefore, they are more adaptable to different environments where the number of images per subject in gallery may be limited. However, it should be noted that because of no training step, all these approaches except the proposed denoising approach could be lack of learning ability from the gallery in some extent, meaning that given a larger gallery the proposed methods cannot learn more discriminative information from the gallery. Further research may be carried out to balance the learning ability and algorithm adaptability.

8.2 Recommendations for Further Research

As seen in the experimental results, there is still some potential to improve recognition performances for existing approaches in the cases with larger illumination variations. One of probable reasons for unsatisfactory performances is that it is still difficult to obtain the information in shadows. A better understanding of shadows will be of great help to model images under varying illuminations. A face model based on sparse representation will be a good choice. Some existing works on spares

representation [114–116] have been proposed and have obtained good results on benchmark face databases. However, there is seldom work on sparse representation focusing on illumination variations. Due to the nature of shadow, it is reasonable to assume that the face with large shadow is a sparse matrix and consider the reflectance of human face as a low-rank matrix. Therefore, it will be a key point to investigate state-of-the-art sparse representation-based classification techniques in the application of illumination invariant face recognition and develop more accurate and efficient sparse representation-based face recognition approaches under varying illuminations.

The DCT based illumination normalization approach proposed in Chapter 4 seems very promising, but there remains much to be explored especially on the properties of DCT coefficients. Without training process, the proposed method can be implemented easily and does not require multiple images per subject in gallery. However, from another point of view, due to the lack of training process, the proposed approaches cannot learn more discriminative informations when multiple images per subjects are involved in the gallery set. Chapter 6 has proposed an interesting insight on modelling distributions of DCT coefficients and denosing. A thorough study on distributions of DCT coefficients for images under different illumination conditions, as well as other variations, can be carried out. The study can generate a more accurate statistical model for DCT coefficients and motivate to investigate more reliable parameter and threshold estimation approaches. A Bayes estimator can be applied in the estimation process to maximize the posterior expectation of a utility function defined based on some face recognition performance measurements.

To deal with complex environments where illumination variations are always coupled with other variations, an ideal feature representation of human face should

be not only illumination invariant, but also robust enough against aging, pose and expression variations. Another future research direction will be the development of local features. As shown in Chapter 5 and Chapter 7, local features achieve good robustness against illumination variations. Extracting high-order derivative local information will be a promising direction for face recognition under not only illumination variations, but also expression, aging and pose variations. Both Local Derivative Pattern (LDP) [44] and Local Line Derivative Pattern (LLDP) proposed in Chapter 7 have achieved good performances but still have large potentials. Any improvement on binary coding function, label coding or distance measurement may make illumination invariant face recognition approaches more accurate. Taking distance measurement as an example, popular distance measurements on global information are not as ideal as expected in face recognition. The proposed distance measurement in Chapter 7 has gained good performances but more distance measurements based on local information should be investigated instead of conventional distance measurement. For example, distance between the corresponding part of faces out of the shadow could be measured with a higher weight. This kind of measurement is more robust against some variations in illumination. With the local similarity measurement, better recognition performance under variable circumstances is expected to be achieved.

Author's Publications

Chapter

1. Z. C. Lian and M. J. Er, Face recognition under varying illumination. *New Trends in Technologies: Control, Management, Computational Intelligence and Network Systems*. Sciyo, ISBN 978-953-7619-X-X, 2010, pp. 209-226.

Journal Papers

1. Z. C. Lian and M. J. Er, Illumination normalisation for face recognition in transformed domain. *Electronics Letters*, 2010, Volume 46, Issue 15, pp. 1060-1061.

2. Z. C. Lian, M. J. Er and Y. C. Liang, A novel efficient local illumination compensation method based on DCT in logarithm domain. *Pattern Recognition Letters*, 2012, 33, pp. 1725-1733.

3. Z. C. Lian and M. J. Er, Local Relation Map: A Novel Illumination Invariant Face Recognition Approach. *International Journal of Advanced Robotic Systems*, 2012, Volume 9, art. no. 130.

Conference Papers

1. Z. C. Lian, M. J. Er, and Y. Cong, Local Line Derivative Pattern For Face Recognition. *The International Conference on Image Processing (ICIP 2012)*, art. no. 6467143, pp. 1449-1452, Orlando, USA, 2012.

2. Z. C. Lian, M. J. Er, and J. K. Li, A Novel Illumination Normalization Method based on Local Relation Map. *The 7th IEEE Conference on Industrial Electronics and Applications (ICIEA 2012)*, art. no. 6360731, pp. 250-253, Singapore, 2012.

3. Z. C. Lian, M. J. Er, and J. K. Li, A Novel Face Recognition Approach under Illumination Variations based on Local Binary Pattern. *CAIP 2011*, Part II, LNCS 6855, pp. 89-96, Seville, Spain, 2011.

4. Z. C. Lian, M. J. Er, and J. K. Li, A Novel Local Illumination Normalization Approach for Face Recognition. *the Eighth International Symposium on Neural Networks (ISNN 2011)*, 6676 LNCS (PART 2), pp. 350-355, Guilin, China, 2011.

5. Z. C. Lian and M. J. Er, An efficient illumination normalization method in a transformed domain. *11th International Conference on Control, Automation, Robotics and Vision, ICARCV 2010*, art. no. 5707285, pp. 884-889, Singapore, 2010.

Bibliography

[1] Moses Y Ullman S Adini Y. Face recognition: The problem of compensating for changes in illumination direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):721–732, 1997.

[2] Z. C. Lian and M. J. Er. *Face recognition under varying illumination. New Trends in Technologies: Control, Management, Computational Intelligence and Network Systems.* Sciyo, 2010.

[3] Pentland A Turk M. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[4] Moghaddam Baback Starner Thad Pentland Alex. View-based and modular eigenspaces for face recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 84–91, 1994.

[5] Zhang D Frangi A F Yang J.-Y. Yang J. Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):131–137, 2004.

[6] Chellappa R Etemad K. Discriminant analysis for recognition of human face images. *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, 14(8):1724–1733, 1997.

[7] Hespanha J P Kriegman D J Belhumeur P.N. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[8] Kittler J Messer K Zou X. Illumination invariant face recognition: A survey. In *IEEE Conference on Biometrics: Theory, Applications and Systems, BTAS'07*, 2007.

[9] Kriegman D J Belhumeur P.N. What Is the Set of Images of an Object under All Possible Illumination Conditions? *International Journal of Computer Vision*, 28(3):245–260, 1998.

[10] Belhumeur P N Kriegman D.J. What shadows reveal about object structure. *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, 18(8):1804–1813, 2001.

[11] Belhumeur P N Kriegman D J Georghiades A.S. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001.

[12] Jacobs D W Basri R. Lambertian reflectance and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):218–233, 2003.

[13] Ho J Kriegman D J Lee K.-C. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):684–698, 2005.

[14] Chang K Flynn P Bowyer K.W. A survey of approaches to three-dimensional face recognition. In *Proceedings - International Conference on Pattern Recognition*, volume 1, pages 358–361, 2004.

[15] Chang K Flynn P Bowyer K.W. A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition. *Computer Vision and Image Understanding*, 101(1):1–15, 2006.

[16] Bowyer K W Flynn P J Chang K.I. An evaluation of multimodal 2D+3D face biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):619–624, 2005.

[17] Vetter T Blanz V. Face recognition based on fitting a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, 2003.

[18] Riklin-Raviv T Shashua A. The quotient image: class-based re-rendering and recognition with varying illuminations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):129–139, 2001.

[19] Chen C.-S. Chen C.-P. Lighting normalization with generic intrinsic illumination subspace for face recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, volume II, pages 1089–1096, 2005.

[20] Brajovic V Gross R. An Image Preprocessing Algorithm for Illumination Invariant Face Recognition. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2688:10–18, 2003.

[21] Li S Z Wang Y Zhang J Wang H. Self quotient image for face recognition. In *Proceedings - International Conference on Image Processing, ICIP*, volume 5, pages 1397–1400, 2004.

[22] Li S Z Wang Y Wang H. Generalized quotient image. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages II498–II505, 2004.

[23] Yin W Zhou X S Comaniciu D Huang T S Chen T. Illumination normalization for face recognition and uneven background correction using total variation based image models. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 532–539, 2005.

[24] Yin W Zhou X S Comaniciu D Huang T S Chen T. Total variation models for variable lighting face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1519–1524, 2006.

[25] Tian J Wu L.-F. Zhang Y.-Y. Yang X He X.-G. Illumination normalization with morphological quotient image. *Ruan Jian Xue Bao/Journal of Software*, 18(9):2318–2325, 2007.

[26] Zheng W.-S. Lai J Yuen P C Xie X. Face illumination normalization on large and small scale features. In *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2008.

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

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

[29] Woods R E Gonzales R.C. *Digital Image Processing,2nd edition.* Prentice Hall, Upper Saddle River, 1992.

[30] Amburn E.Philip Austin John D Cromartie Robert Geselowitz Ari Greer Trey ter Haar Romeny Bart Zimmerman John B Zuiderveld Karel Pizer Stephen

M. ADAPTIVE HISTOGRAM EQUALIZATION AND ITS VARIATIONS. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987.

[31] Lam K.-M. Xie X. Face recognition under varying illumination based on a 2D face shape model. *Pattern Recognition*, 38(2):221–230, 2005.

[32] Er M J Wu S Chen W. Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 36(2):458–466, 2006.

[33] Pandey S Gupta M N Vishwakarma V.P. A novel approach for face recognition using DCT coefficients re-scaling for illumination normalization. In *Proceedings of the 15th International Conference on Advanced Computing and Communications, ADCOM 2007*, pages 535–539, 2007.

[34] Castillo L E Perez C.A. Genetic improvements in illumination compensation by the discrete cosine transform and local normalization for face recognition. In *Proceedings of SPIE - The International Society for Optical Engineering*, volume 7266, 2008.

[35] Garcia-Reyes E Condes-Molleda Y Mendez-Vazquez H. A new combination of local appearance based methods for face recognition under varying lighting conditions. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5197 LNCS:535–542, 2008.

[36] Ward R Du S. Wavelet-based illumination normalization for face recognition. In *Proceedings - International Conference on Image Processing, ICIP*, volume 2, pages 954–957, 2005.

[37] Tan Z.-F. Guo J Nie X.-F. Face illumination compensation based on wavelet transform. *Guangxue Jingmi Gongcheng/Optics and Precision Engineering*, 16(1):150–155, 2008.

[38] Yang L.-P. Gu X.-H. Li W.-H. Gong W.-G. Illumination compensation based on multi-level wavelet decomposition for face recognition. *Guangxue Jingmi Gongcheng/Optics and Precision Engineering*, 16(8):1459–1464, 2008.

[39] Shan S Chen X Gao W Han H. Illumination transfer using homomorphic wavelet filtering and its application to light-insensitive face recognition. In *2008 8th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2008*, 2008.

[40] Hadid A Pietikainen M Ahonen T. Face recognition with local binary patterns. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3021:469–481, 2004.

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

[42] Triggs B Tan X. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Transactions on Image Processing*, 19(6):1635–1650, 2010.

[43] Srisuk S Petpon A. Face recognition with local line binary pattern. In *Proceedings of the 5th International Conference on Image and Graphics, ICIG 2009*, pages 533–539, 2010.

[44] Gao Y Zhao S Liu J Zhang B. Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor. *IEEE Transactions on Image Processing*, 19(2):533–544, 2010.

[45] Huang X Li S Z Wang Y Wu X Zhang G. Boosting Local Binary Pattern (LBP)-based face recognition. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3338:179–186, 2004.

[46] Marcel S Rodriguez Y. Face authentication using adapted local binary pattern histograms. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3954 LNCS:321–332, 2006.

[47] Pietikainen M Ahonen T Hadid A. A discriminative feature space for detecting and recognizing faces. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages II797–II804, 2004.

[48] Liu Q Lu H Tong X Jin H. Face detection using improved LBP under bayesian framework. In *Proceedings - Third International Conference on Image and Graphics*, pages 306–309, 2004.

[49] Kabir Md.H. Chae O Jabid T. Robust facial expression recognition based on local directional pattern. *ETRI Journal*, 32(5):784–794, 2010.

[50] Pietikainen M Hadid A Feng X. Facial expression recognition based on local binary patterns. *Pattern Recognition and Image Analysis*, 17(4):592–598, 2007.

[51] Gong S McOwan P W Shan C. Robust facial expression recognition using local binary patterns. In *Proceedings - International Conference on Image Processing, ICIP*, volume 2, pages 370–373, 2005.

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

[53] Pietikäinen M Mäenpää T Ojala T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

[54] Triggs B Tan X. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4778 LNCS:168–182, 2007.

[55] Li S Z Wang Y Huang X. Shape localization based on statistical method using extended local binary pattern. In *Proceedings - Third International Conference on Image and Graphics*, pages 184–187, 2004.

[56] Kabir H Chaei O Jabid T. Local Directional Pattern (LDP) for face recognition. In *ICCE 2010 - 2010 Digest of Technical Papers International Conference on Consumer Electronics*, pages 329–330, 2010.

[57] Kabir M H Chae O Jabid T. Local directional pattern (LDP) for face recognition. *International Journal of Innovative Computing, Information and Control*, 8(4):2423–2437, 2012.

[58] Gao Y Caelli T Zhao S. High-order circular derivative pattern for image representation and recognition. In *Proceedings - International Conference on Pattern Recognition*, pages 2246–2249, 2010.

[59] Shan S Gao W Chen X Zhang H Zhang W. Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A novel non-statistical model for face representation and recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, volume I, pages 786–791, 2005.

[60] Shan S Chen X Chen J Xie S. Fusing local patterns of gabor magnitude and phase for face recognition. *IEEE Transactions on Image Processing*, 19(5):1349–1361, 2010.

[61] Zhao L Shao Y Ge Y Ke S. Comparison of WHT and HHT methods for analyzing non-Gaussian wind pressure signals. *Tumu Gongcheng Xuebao/China Civil Engineering Journal*, 44(6):61–67, 2011.

[62] H Bogucka. On the impact of the impulse noise on the WHT/OFDM transmission. *IEEE Communications Letters*, 9(1):37–39, 2005.

[63] C.-F. Chan. Efficient implementation of class of isotropic quadratic filters by using Walsh-Hadamard transform. *Electronics Letters*, 35(16):1306–1308, 1999.

[64] Saniie J Cardoso G. Performance evaluation of DWT, DCT, and WHT for compression of ultrasonic signals. In *Proceedings - IEEE Ultrasonics Symposium*, volume 3, pages 2314–2317, 2004.

[65] Urbaszek A Huber J Schaldach M Molina A. Novel, low-complexity method for intracardiac signal compression in implantable devices. In *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, volume 1, pages 95–96, 1997.

[66] Prost R Choi T.-Y. Jung H.-Y. A unified mathematical form of the Walsh-Hadamard transform for lossless image data compression. *Signal Processing*, 63(1):35–43, 1997.

[67] Matsumoto Shuichi Hamada Takahiro. WHT-based composite motion compensated NTSC interframe direct coding. *IEEE Transactions on Communications*, 44(12):1711–1719, 1996.

[68] Deodhar R S Patnaik L M Dhavale S.V. Walsh Hadamard Transform based robust blind watermarking for digital audio copyright protection. *Communications in Computer and Information Science*, 250 CCIS:469–475, 2011.

[69] Uehira K Yanaka K Ishikawa Y. Practical evaluation of illumination watermarking technique using orthogonal transforms. *IEEE/OSA Journal of Display Technology*, 6(9):351–358, 2010.

[70] Samokhval V A Podenok L P Sadykhov R.Kh. Face recognition algorithm on the basis of truncated Walsh-Hadamard transform and synthetic discriminant functions. In *Proceedings - Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 219–222, 2004.

[71] Liu H Wang Q Yang J Guo Z. A fast algorithm of face detection for driver monitoring. In *Proceedings - ISDA 2006: Sixth International Conference on Intelligent Systems Design and Applications*, volume 2, pages 267–271, 2006.

[72] Osman I Yahia M Hassan M. Walsh-Hadamard Transform for Facial Feature Extraction in Face Recognition. *World Academy of Science, Engineering and Technology*, 29:194–198, 2007.

[73] Espinosa-Duro V Ortega J A Faundez-Zanuy M. An efficient face identification method in a transformed domain. In *Proceedings - International Carnahan Conference on Security Technology*, pages 281–284, 2007.

[74] Z. Lian and M.J. Er. An efficient illumination normalization method in a transformed domain. In *11th International Conference on Control, Automation, Robotics and Vision, ICARCV 2010*, pages 884–889, 2010.

[75] Z. Lian and M.J. Er. Illumination normalisation for face recognition in transformed domain. *Electronics Letters*, 46(15):1060–1061, 2010.

[76] Er M J Wu S Chen W. PCA and LDA in DCT domain. *Pattern Recognition Letters*, 26(15):2474–2482, 2005.

[77] Liao H.-Y.M. Ko M.-T. Lin J.-C. Yu G.-J. Chen L.-F. New LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, 33(10):1713–1726, 2000.

[78] Liu Q Lu H Ma S Huang R. Solving the small sample size problem of LDA. In *Proceedings - International Conference on Pattern Recognition*, volume 16, pages 29–32, 2002.

[79] Wang Y Li S Z Tan T Liu W. Null space approach of fisher discriminant analysis for face recognition. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3087:32–44, 2004.

[80] Z. Lian, M.J. Er, and Y. Liang. A novel efficient local illumination compensation method based on dct in logarithm domain. *Pattern Recognition Letters*, 33(13):1725–1733, 2012.

[81] Zhang Xiaoyu Podilchuk Christine. Face recognition using DCT-based feature vectors. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, volume 4, pages 2144–2147, 1996.

[82] Venkatesh Y V Ramasubramanian D. Encoding and recognition of faces based on the human visual model and DCT. *Pattern Recognition*, 34(12):2447–2458, 2001.

[83] Liu C Zhang Y. Efficient face recognition method based on DCT and LDA. *Journal of Systems Engineering and Electronics*, 15(2):211–216, 2004.

[84] Xiao Y Khorasani K Ward R K Ma L. A new facial expression recognition technique using 2-D DCT and K-means algorithm. In *Proceedings - International Conference on Image Processing, ICIP*, volume 5, pages 1269–1272, 2004.

[85] Zhang D Jing X.-Y. A face and palmprint recognition approach based on discriminant DCT feature extraction. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34(6):2405–2415, 2004.

[86] Batista L V Poel J V D Matos F.M.D.S. Face recognition using DCT coefficients selection. In *Proceedings of the ACM Symposium on Applied Computing*, pages 1753–1757, 2008.

[87] Baker S Bsat M Sim T. The CMU Pose, Illumination, and Expression Database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618, 2003.

[88] Bolme D Draper B A Teixeira M Beveridge J.R. The CSU face identification evaluation system : IIIts purpose, features, and structure. *Machine Vision and Applications*, 16(2):128–138, 2005.

[89] J Demsar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.

[90] O J Dunn. Multiple comparisons among means. *Journal of the American Statistical Association*, 56:52–64, 1961.

[91] Lam K.-M. Xie X. An efficient illumination normalization method for face recognition. *Pattern Recognition Letters*, 27(6):609–617, 2006.

[92] Z. Lian, M.J. Er, and J. Li. A novel local illumination normalization approach for face recognition. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6676 LNCS(PART 2):350–355, 2011.

[93] Z. Lian and M.J. Er. Local relation map: A novel illumination invariant face recognition approach. *International Journal of Advanced Robotic Systems*, 9, 2012.

[94] Stiefelhagen R Ekenel H.K. Automatic frequency band selection for illumination robust face recognition. In *Proceedings - International Conference on Pattern Recognition*, pages 2684–2687, 2010.

[95] F Mueller. Distribution shape of two-dimensional DCT coefficients of natural images. *Electronics Letters*, 29(22):1935–1936, 1993.

[96] Hamilton C H Macklem M S McMichael J S Swart N R Bauschke H.H. Recompression of JPEG images by requantization. *IEEE Transactions on Image Processing*, 12(7):843–849, 2003.

[97] Lynch W E Vincent A Sorial H. Estimating Laplacian parameters of DCT coefficients for requantization in the transcoding of MPEG-2 video. In *IEEE International Conference on Image Processing*, volume 1, pages 956–959, 2000.

[98] Hamilton C H Macklem M S McMichael J S Swart N R Bauschke H.H. A requantization-based method for recompressing JPEG images. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, volume 3, pages III/2489–III/2492, 2002.

[99] Goodman J W Lam E.Y. A mathematical analysis of the DCT coefficient distributions for images. *IEEE Transactions on Image Processing*, 9(10):1661–1666, 2000.

[100] Purczynski J Krupinski R. Modeling the distribution of DCT coefficients for JPEG reconstruction. *Signal Processing: Image Communication*, 22(5):439–447, 2007.

[101] S Nadarajah. Gaussian DCT coefficient models. *Acta Applicandae Mathematicae*, 106(3):455–472, 2009.

[102] S Nadarajah. GG DCT coefficient models. *International Journal of Wavelets, Multiresolution and Information Processing*, 8(5):793–812, 2010.

[103] Zhang M Xiong Z Zhou T. Generalised Gaussian distribution to early detect zero quantised discrete cosine transform coefficients in H.264/AVC video encoding. *IET Image Processing*, 4(6):473–485, 2010.

[104] Barlaud Michel Mathieu Pierre Daubechies Ingrid Antonini Marc. Image coding using wavelet transform. *IEEE Transactions of Image Processing*, 1(2):205–220, 1992.

[105] David L Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, 1995.

[106] Kolaczyk E D McCulloch R E Chipman H.A. Adaptive Bayesian wavelet shrinkage. *Journal of the American Statistical Association*, 92(440):1413–1421, 1997.

[107] Yu Bin Vetterli Martin Chang S.Grace. Image denoising via lossy compression and wavelet thresholding. In *IEEE International Conference on Image Processing*, volume 1, pages 604–607, 1997.

[108] Sapatinas T Silverman B W Abramovich F. Wavelet thresholding via a Bayesian approach. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 60(4):725–749, 1998.

[109] Yu B Vetterli M Chang S.G. Spatially adaptive wavelet thresholding with context modeling for image denoising. *IEEE Transactions on Image Processing*, 9(9):1522–1531, 2000.

[110] Yu B Vetterli M Chang S.G. Adaptive wavelet thresholding for image denoising and compression. *IEEE Transactions on Image Processing*, 9(9):1532–1546, 2000.

[111] Z. Lian, M.J. Er, and Y. Cong. Local line derivative pattern for face recognition. In *Proceedings - International Conference on Image Processing, ICIP*, pages 1449–1452, 2012.

[112] Z. Lian, M.J. Er, and J. Li. A novel face recognition approach under illumination variations based on local binary pattern. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6855 LNCS(PART 2):89–96, 2011.

[113] Wechsler H Huang J Rauss P J Phillips P.J. The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5):295–306, 1998.

[114] Yang A Y Ganesh A Sastry S S Ma Y Wright J. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.

[115] Do T T Tran T D Chen Y. Robust face recognition using locally adaptive sparse representation. In *Proceedings - International Conference on Image Processing, ICIP*, pages 1657–1660, 2010.

[116] Xu R Song S Xia H. Robust facial expression recognition via sparse representation over overcomplete dictionaries. *Journal of Computational Information Systems*, 8(1):425–433, 2012.