論文 / 著書情報
Article / Book Information

題目(和文)	顔の検出と認識における照明の扱いに関する研究
Title(English)	Study on Illumination Processing in Face Detection and Recognition
著者(和文)	YaoMin
Author(English)	Min Yao
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第9598号, 授与年月日:2014年6月30日, 学位の種別:課程博士, 審査員:長橋　宏,熊澤　逸夫,小池　康晴,山口　雅浩,金子　寛彦
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第9598号, Conferred date:2014/6/30, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Study on Illumination Processing in Face Detection and Recognition

By

Min Yao

Academic Advisor: Prof. Hiroshi Nagahashi

A dissertation submitted for the degree of

Doctor of Engineering

Department of Information Processing

Tokyo Institute of Technology

June 2014

Abstract

Human face is an advantageous biological trait for individual identification. In recent years, face detection and recognition have been widely exploited in various application areas. Although great advances have emerged for their real applications under different conditions, they still suffer from some performance limitations caused by specific uncontrolled factors. Among these factors, the negative influence of illumination is a key problem.

This dissertation addresses the illumination problem in face detection and recognition. With the face detection, we normalize the given image prior to executing the face detector. It is aimed at maintaining the advantages of pre-trained learning-based face detectors while improving their resistance to varying illumination. The prepended normalization largely affects the detection results in difficult lighting conditions, which was rarely studied in the literature. We widely apply a number of normalization techniques that may be initially proposed for different purposes rather than face detection. Meanwhile, we propose a novel normalization method which enhances the local contrast (local facial features) and removes the non-uniform illumination. The experimental results demonstrate the effectiveness of some of the applied normalization techniques. Our method gets the highest scores in terms of F-measure.

With the face recognition, we introduce two proposals to generate effective face representations. The first improves a recently proposed method and incorporates the oriented and scaled facial information. The second is to utilize the multifractal analysis which is able to code the local and global textural regularity. They all directly extract the

illumination insensitive facial features from each single face image. These proposed face representations, both theoretically and experimentally, proved to be insensitive to illumination and discriminative for face classification. We also did a survey on the illumination processing in face recognition.

To my beloved parents, husband and son ...

Acknowledgements

First and foremost, I would like to express my sincere and profound gratitude to Prof. Nagahashi, my supervisor, a respectable, responsible and knowledgeable scholar, for his kindness, patience and understanding, his enormous moral and academic support, constructive comments and guidance throughout the entire period of my doctoral study. Various supports and encouragement were given to me in the daily life as well.

I also wish to extend my great appreciation and sincere thanks to the doctoral Advisory Committee, Prof. Yamaguchi, Prof. Kumazawa, Prof. Koike and Associate Prof. Kaneko for their valuable contribution toward the completion of this dissertation.

I would also like to extend my gratitude to Assistant Prof. Aoki for his estimable suggestions and discussions during this research.

I want to thank my senior Mr. Chamidu for his assistance in my studying algorithms and programming and his kind advice in any aspect of my research.

My Ph.D study would not have been possible without the Monbu-Kagakusho scholarship offered by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT). I even would not have such a great opportunity to study and live in Japan without that financial support. I genuinely acknowledge the generosity of this foundation extended for me.

I also wish to thank all my lab mates for their friendly helps whenever I needed them during my study and daily life. I had a good life in the happy air of our lab.

My heartfelt thanks go to my family for their enduring love, constant moral encouragement, care and concern to me throughout my life. My dearest husband, Da Zhang constantly encouraged me whenever I was overwhelmed with disappointments in both academic and personal life. My son Tianqi Zhang has been a great source of energy and happiness for my life.

Great thanks also go to those who or which help me during my studying and living I have not mentioned here.

Contents

List of Figures

List of Figures

List of Tables

List of Tables

Chapter 1

Introduction

1.1 Problem definition

Face detection and recognition have attracted much academic attention during the past several decades. Significantly increased efforts have gone into the improvement of their effectiveness and efficiency. Their advent and rapid development are largely driven by the growing application demands, such as static matching of controlled photographs as in mug shots matching, credit card verification to surveillance video images, identification for law enforcement, authentication for banking and security system access, and other real/lively applications. Also, the advances in signal analysis techniques, computational powers and availability of feasible modern sensing/analysis/rendering equipment make computers become more and more intelligent and meanwhile guarantee the possibility of reliable face detection and recognition system.

Usually, face detection is considered as the step stone to all facial analysis algorithms and also as the preprocessing step of the face recognition [3, 4]. With a given arbitrary image, the goal of face detection is to tell whether or not any faces exist in the image and, if present, return the location and extent of each face [5, 3]. Face recognition refers to extracting intrinsic attributes of face images with raw pixel intensities and subsequently performing the classification between face images of various personal identities. Usually face recognition is performed to images containing only the face regions. A number of approaches have been proposed to tackle different issues in face detection and recognition. However,

accurate detection and recognition of faces are still limited with problems caused by uncontrolled factors such as complex backgrounds (containing many face-like non-face objects), low resolution or poor quality images, variant face poses, uneven/harsh illumination, various facial expressions, unexpected occlusions, ect. Among these challenges, the varying illumination is one of the most tough problems.

In this dissertation, we address the illumination problem in face detection and recognition. Overcoming the negative effects from varying illumination on the performance of face detection and recognition is the key focus. Particularly speaking, with the face detection, we aim to sufficiently attenuate the illumination in a given image and maintain the important facial features for the face detector. With the face recognition, we attempt to represent the faces in a discriminative manner, which is insensitive to illumination changes; intrinsic facial features reflecting the personal identities are supposed to be preserved and perceptible for the face classifier.

1.2 Motivation

1.2.1 Importance of face detection and recognition

An individual's biological traits cannot be misplaced, forgotten, stolen or forged and thus biometric-based techniques have emerged as the most promising option for recognizing individuals in recent years. Biometric-based technologies include identification based on physiological characteristics (such as face, fingerprints, finger geometry, hand geometry, hand veins, palm, iris, retina, ear and voice) and behavioral traits (such as gait, signature and keystroke dynamics) [6, 7]. face recognition appears to be more advantageous over other biometric methods, since face recognition can be done passively without any explicit action or participation on the part of the user since images can be acquired from a distance by a camera and thus it is customer-friendly. Face detection is a necessity prior to the recognition if images are taken far away.

There are numerous application areas in which face detection and recognition can be exploited [6]:

- Security (access control to buildings, airports/seaports, ATM machines and border checkpoints [8, 9]; computer/ network security [10]; email authentication on multimedia workstations).

- Surveillance (a large number of CCTVs can be monitored to look for known criminals, drug offenders, etc. and authorities can be notified when one is located; for example, this procedure was used at the Super Bowl 2001 game at Tampa, Florida [11]; in another instance, according to a CNN report, two cameras linked to state and national databases of sex offenders, missing children and alleged abductors have been installed recently at Royal Palm Middle School in Phoenix, Arizona [12]).

- General identity verification (electoral registration, banking, electronic commerce, identifying newborns, national IDs, passports, drivers licenses, employee IDs).

- Criminal justice systems (mug-shot/booking systems, post-event analysis, forensics).

- Image database investigations (searching image databases of licensed drivers, benefit recipients, missing children, immigrants and police bookings).

- "Smart Card" applications (in lieu of maintaining a database of facial images, the face-print can be stored in a smart card, bar code or magnetic stripe, authentication of which is performed by matching the live image and the stored template) [13].

- Multimedia environments with adaptive human-computer interfaces (part of ubiquitous or context-aware systems, behavior monitoring at childcare or old peoples centers, recognizing a customer and assessing his needs) [14, 15].

- Video indexing (labeling faces in video) [16, 17].

- Witness face reconstruction [18].

In addition to these applications, the underlying techniques in the current face recognition technology have also been modified and used for related applications such as gender classification [19, 20], expression recognition [21] and facial feature recognition and tracking [22]; each of these has its utility in various domains.

1.2.2 Challenges from illumination problem

Illumination problem is a key bottleneck to the practical application of face detection and recognition systems in real world. Intuitively, varying illumination adds extra noise and unexpected (unnatural) patterns to a given images. This will change or blur the original patterns reflecting the attributes corresponding to different objects including faces and the background. Illumination variations can be caused by the lighting sources, incorrect camera settings, or nearby objects that may cast shadows on the subject's face or reflect additional light. In general, lighting in images can varies in three ways [23]:

1.*Intensity*. Poor lighting conditions may lead to dark, uneven and noisy images; incorrect exposure settings on the camera may lead to abnormally exposed images with lack of intensity contrast.

2.*Color*. Non-white light sources and incorrect camera settings, such as white balance, may generate images with unnatural colors, and, in particular, may render skin an incorrect chrominance.

3.*Location*. Directional lighting may cause the shadows to be cast across the face and result in strong shadow boundaries.

All of the above factors can have strong negative impacts on the performance of face detection and recognition systems. Actually, even human abilities are similarly impaired by shadows and other great illumination changes [24].

For the face detection, it is needed to distinguish between facial and non-facial features. It is theoretically possible to overcome the illumination problem by including sufficient categories of illuminated face images in the training set[1]. However, illuminated facial patterns may vary so largely and make the whole

[1]The widely used and effective face detectors nowadays are usually based on learning algorithms. This will be introduced in detail in the next chapter

Figure 1.1: Original illuminated face image, illumination component L, and reflectance component R (from left to right).

detection system detect many non-facial patterns as faces, since these patterns may be quite similar to the illuminated facial patterns. Some harshly illuminated natural photographs are shown in Fig. 1.1 (a). In these photos, the facial structures are made ambiguous by the illumination effects and may be decided as non-faces. Moreover, color information may also be made useless since difficult lighting conditions are likely to result in incorrect chrominance as mentioned above. Thus face detection techniques that rely on the color information may fail under varying illumination. Some learning-based face detectors adopt so-called illumination invariant features [25–28]. But unfortunately, most of these features are only considered insensitive to monotonic illumination and vulnerable to non-uniform illumination. In fact, it has been pointed out that there is no illumination invariant features [29].

On the other hand, for face recognition, illumination problem is similarly challenging. Figure 1.1 (b) gives some sample images of one subject under varying

illumination, from which it can be noticed that, although they are of the same person, the visual appearances of these images are totally different from each other due to the large illumination changes. It is wrong if one thinks that illumination insensitive face detecting ability is enough and can be continuously extended to the recognition task. The reason is that distinguishing between facial and non-facial features significantly differs from face-to-face matching. Some famous tests of state-of-the-art recognition techniques conducted during the past several years, such as the FERET evaluations [30], FRVT 2000 [31], FRVT 2002 [32] and the FAT 2004 [33], have confirmed that illumination variations is one of the three major problems[1] plaguing current face recognition systems. It has been also proved, both theoretically and experimentally, that variations of facial appearance caused by varying illumination overwhelm those caused by individual identities [34, 35]. Although numerous approaches have been developed to deal with the illumination problem, they still suffer from lacks of efficiency or precision. Recently, representing face images in an illumination insensitive way as a preprocessing step attracts much attention [36–38] due to their simple, straightforward, efficient property. It is an ongoing research topic and many issues present in the existing methods still remain unsolved.

1.3 Objectives and approaches

For the face detection part, a number of well-known learning-based face detectors can achieve extraordinary performance in controlled environments. But face detection under severe illumination conditions is difficult. Possible solutions could be creating illumination invariant features or utilizing skin color information. However, the features and skin colors are not sufficiently reliable as introduced before. Another possible solution is to do illumination normalization prior to executing face detectors. However, researches about the application of normalization to face detection are hardly found in the literature. The study on the influence of prepended normalization on the subsequent face detection has not

[1] The other two major problems are age changes and pose variations.

received the attention it deserves. So our first goal is to maintain the advantages of pre-trained learning-based face detectors while improving the resistance to varying illumination.

We apply a variety of normalization techniques (initially proposed for different purposes rather than face detection) and test them with two famous face detectors (Haar-like face detector and LBP face detector). The comparison results between the detection using normalization and the original detector without normalization will reflect the effectiveness of normalization to the face detection under varying illumination. The major challenge is how capable these normalization techniques are, in suppressing the illumination effects and preserving the important facial patterns. Despite their usefulness in the corresponding areas, they may not be suitable for face detection task. Hence, we propose a new normalization method called Segmentation-based Histogram Truncation and Stretching (SH) which enhances the local contrast (local facial structures) and removes the non-uniform illumination. This normalization-based improvement of the learning-based face detectors is appealing since it is achieved without re-learning.

For the face recognition part, some approaches focus on the image capturing process, that is, the imaging techniques which is more related to the hardware; some build the 3D model using large number of samples; some use primitive image processing methods; some others construct illumination insensitive face representations from the given images. Directly representing each face image is considered simple, straightforward, and efficient. Thus we direct our research to generating effective face representations as our second goal.

To realize this goal, we introduce two proposals. The first is based on the recently proposed method Weberface. We utilize its illumination insensitivity but further improve its computation to incorporate the oriented and scaled information and form our face representations named Oriented Weberface (OWF) and Largely-scaled Weberface (LSWFs) respectively. The second is to apply multifractal analysis for face representation which is supposed to be insensitive to illumination and discriminative for face classification. The $f(\alpha)$ features computed from the multifractal analysis characterize both the local and global textural regularity and we use the $f(\alpha)$-image as the final face representation.

The third goal of this research is to show some noteworthy points for illumination insensitive face recognition. By surveying many existing approaches, we discuss and analyze a theoretical basis, related important hints, unsolved problems, substantial issue, etc. Potential research directions are also given.

1.4 Contributions

The entire research has several contributions listed as follow.

- Wide applications of normalization techniques to face detection under varying illumination: we discussed the influence of prepended illumination normalization on the pre-trained face detector by applying various normalization techniques which might be initially proposed for different purpose rather than for face detection. This study is hardly found in the literature.
- A novel illumination normalization method for face detection: we also proposed an effective illumination normalization method specially designed to enhance local facial features while removing the non-uniform illumination in a given image and facilitate the learning-based face detector.
- Improvement of a recently introduced Weberface (WF) in two ways for illumination insensitive face representation: we proposed two new illumination insensitive face representation methods called Oriented Weberface (OWF) and Largely-scaled Weberfaces (LSWFs), respectively.
- A new face representation method using multifractal analysis: we exploited and analyzed a new face representation which proved, theoretically and experimentally, to be insensitive to illumination variances and discriminative for face classification.
- A survey on the illumination processing in face recognition: some noteworthy considerations were discussed. It is different from any existing survey of this topic which either introduced many algorithms according to their computation methods or gave comparative study with their experimental performance.

1.5 Organization

This dissertation contains seven chapters, of which the first is the introduction. Chapter 2 describes related works for this research, including the prominent algorithms generally proposed for face detection and recognition task and the representative approaches specially developed for tackling the illumination problems. The main description of our studies starts from Chapter 3. This chapter focuses on the application of normalization methods to face detection under varying illumination and presents a novel normalization method for addressing the non-uniform illumination problem. Chapter 4 illustrates the second study that improves a recently proposed illumination insensitive face representation method in two ways for more effective face recognition under difficult lighting conditions. Chapter 5 shows another proposal of face representation which uses multifractal analysis for discriminative and illumination insensitive face recognition. Chapter 6 analyzes and discusses several noteworthy points on illumination processing in face recognition by surveying numerous existing approaches. It particularly provides insights and conclusions in how to design better illumination processing methods. Chapter 3, 4 and 5 could be read separately since they describe relatively independent studies. Some common parts such as the depicts of databases used in the recognition task may be repeated for the ease of understanding. Lastly, Chapter 7 concludes the entire work with remarks, limitations, and future study.

Chapter 2

Related works

In this chapter, we present related works including the prominent algorithms generally proposed for face detection & recognition tasks and the representative approaches specially developed for tackling the illumination problems.

2.1 Face detection

Face detection has been one of the most studied research topics in the past decades. The challenging points associated with face detection can be attributed to variations in scale, angle (in-plane rotation), pose (out-of-plane rotation), facial expression, lighting conditions, occlusions, etc., as seen in Fig. 2.1.

In this section, a taxonomy which groups most of the well-known face detection algorithms and techniques is presented and a summary of newly accomplished researches related to face detection is also given. In Sect. 2.1.3, typical illumination processing methods for face detection are discussed.

2.1.1 Overview

Early works (before year 2000) for face detection can be grouped into two major categories, which were nicely surveyed in [5, 39]. Erik and Low [39] presented a more detailed classification than Yang et al. in [5], although they are pretty similar in general. The following is a modified taxonomy that is also summarized in Fig. 2.2.

Figure 2.1: Examples of face images. Note the huge variations in pose, facial expression, lighting conditions, etc.

The first category assembles what are called rule based methods. They are the methods that rely on searching the input image comprehensively for facial features under a predefined set of rules. This category could be partitioned by itself into two sub categories based on the nature of the rules in use. The rules are either knowledge based rules or feature invariant rules. For the knowledge based methods, the rules that govern the search are arbitrary inherited from the human knowledge of typical faces. Yang et al. [5] stated that this category was developed mainly for face localization, not for face detection. However the feature invariant based methods rely on using a combination of one or more low level features such as edges, gray levels, skin color, texture, or even motion if the input is a video stream. This is done in order to search the input image for a single facial feature or multiple facial features (nose, mouth, or eyebrows) to detect the target face. Methods belong to the rule based category are rather more adequate for face localization, not face detection.

As for the second category, the methods belong to it are called model based methods. They are classified into two sub categories. First, the template matching based methods where a model is pre-specified for a human face includes a predefined structure for the different facial features and their inner-relations. The model is either fixed or parameterized to adapt to the dynamic changes in scale, pose, and shape. The correlation between the input image and the model

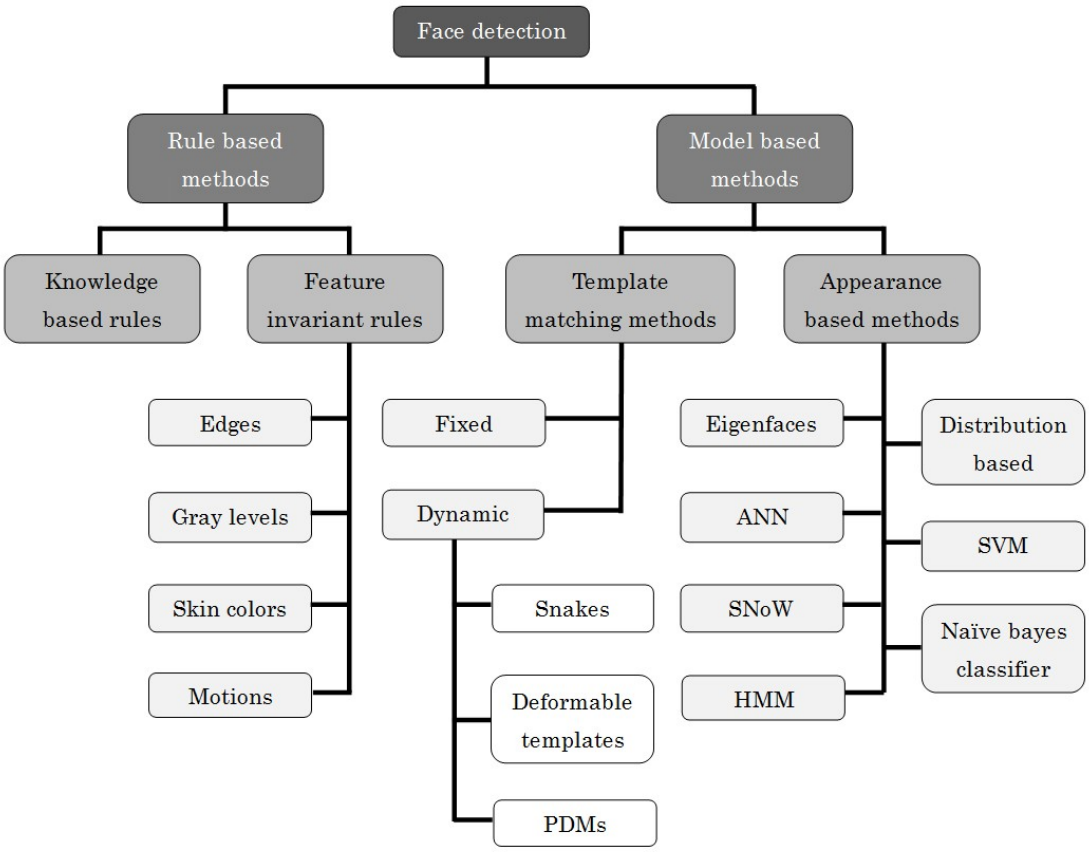


Figure 2.2: Taxonomy of face detection techniques.

is measured for several facial features independently and then a decision is made based on the measured correlation. Examples of the fixed templates can be found in [5]. Examples of parameterized templates that are also known as active shape models include snakes, deformable templates, and point distributed models. Second, appearance based methods (also known as learning-based methods) where a model is learnt of examples from the training samples set. The number of training samples is important because it must be enough to collect as many facial features as possible using one or more feature extraction methods. The classification is done using a specific machine learning technique. Examples of this category include a list of statistical methods and machine learning based techniques such as eigenfaces, distribution-based methods, support vector machines (SVM), artificial neural networks (ANN), sparse network of Winnows (SNoW), naive Bayes

classifiers, hidden Markov model (HMM), etc.

2.1.2 Recent works

This section summarizes some of the recent related works that were published after the comprehensive surveys in [5, 39].

2.1.2.1 Face detection in gray level images

With face detection in gray level images, a series of distinguished researches were introduced by Viola and Jones [40–42] using Haar-like features. Their work is commonly deemed as the seminal one. They started in [41, 40] by introducing a machine learning based framework. They combined three contributions when they presented the integral image representation to compute the features very rapidly. They also used the Adaboost algorithm to select a less number of critical visual features from a large set of initial features. Finally, they presented a combination of cascaded classifiers that are capable of eliminating the regions not of interest during the early stages to save the computational power for regions of actual interest. The framework ran in real time at 15 frames/second accomplishing a detection rate of 81.1% with a 0.02% false positive rate. It was tested using the combination of the MIT [43] and CMU [44, 45] datasets.

Inspired by the Viola and Jones' initial work, many extended and modified versions based on learning algorithms (boosting) were developed. One can find a detailed summary of the recent advances of the learning-based methods in [3]. The general practice is to collect a large set of face and non-face examples, and adopt certain machine learning algorithms to learn a face model to perform classification. There are two key issues in this process: what features to extract, and which learning algorithm to apply.

Different features were adopted for different purposes:

Lienhart and Maydt [46] generalized the feature set of [40] by introducing 45 degree rotated rectangular features, and center-surround features. These features can code more complex textural information. Mita et al. [47] proposed joint Haar-like features, which is based on co-occurrence of multiple Haar-like

features. The authors claimed that the co-occurrence features can better capture the characteristics of human faces, making it possible to construct a more powerful classifier.

To some degree, the above joint Haar-like features resemble a CART tree, which was explored in [48]. It was shown that CART tree based weak classifiers improved results across various boosting algorithms with a small loss in speed.

For multi-view face detection, in [49], three types of features were defined in the detection sub-window. These rectangular features are with flexible sizes and distances and are considered to be non-symmetrical to cater to non-symmetrical characteristics of non-frontal faces. Viola and Jones [42] also proposed a similar feature called diagonal filters which can be computed with 16 array references to the integral image.

Viola et al. [50] further extended the Haar-like feature set to work on motion filtered images for video-based pedestrian detection.

Since haar-like features lack of robustness in handling faces under extreme lighting conditions, modified census transform (MCT) [25] was adopted to generate illumination invariant features for face detection. On each pixel's 3×3 neighborhood, the authors applied a modified census transform that compares the neighborhood pixels with their intensity mean. The results are concatenated into an index number representing the pixel's local structure. During boosting, the weak classifiers are constructed by examining the distributions of the index numbers for the pixels. Similar features using local binary patterns (LBP) can be also found in [26, 27] under a Bayesian and a boosting framework, respectively.

To explore possibilities for further improvements of performance, more and more complex features were proposed in the literature. For instance, Liu and Shum [51] studied generic linear features, which is defined by a mapping function. Note the Haar-like feature set is a subset of linear features. Another example is the anisotropic Gaussian filters in [52]. In [53], the linear features were constructed by prelearning them using local non-negative matrix factorization (LNMF), which is still sub-optimal. Instead, Liu and Shum [51] proposed to search for the linear features by examining the Kullback-Leibler (KL) divergence of the positive and negative histograms projected on the feature during boosting (hence the name Kullback-Leibler boosting). In [54], the authors proposed to

apply Fisher discriminant analysis and more generally recursive nonparametric discriminant analysis (RNDA) to find the linear projections.

Another popular complex feature for face/object detection is based on regional statistics such as histograms. Levi and Weiss [55] proposed local edge orientation histograms, which computes the histogram of edge orientations in subregions of the test windows. Later, Dalal and Triggs [56] proposed a similar scheme called histogram of oriented gradients (HOG), which became a very popular feature for human/pedestrian detection [57–61]. Zhang et al. [62] proposed another histogram-based feature called spatial histograms, which is based on local statistics of LBP. HOG and LBP were also combined in [63], which achieved excellent performance on human detection with partial occlusion handling.

Huang et al. [64] proposed a sparse feature set in order to strengthen the features' discrimination power without incurring too much additional computational cost. The performance of the multi-view face detector trained in [64] using sparse features was very good.

There have also been many features that attempted to model the shape of the objects. For instance, Opelt et al. [65] composed multiple boundary fragments to weak classifiers and formed a strong "boundary-fragment-model" detector using boosting. Shotton et al. [66] learned their object detectors with a boosting algorithm and their feature set consisted of a randomly chosen dictionary of contour fragments. A very similar edgelet feature was proposed in [67], and was used to learn human body part detectors in order to handle multiple, partially occluded humans. In [68], shapelet features focusing on local regions of the image were built from low-level gradient information using AdaBoost for pedestrian detection.

On the other hand, various Boosting learning algorithms were also widely studied:

In the original face detection paper by Viola and Jones [40], the standard AdaBoost algorithm [69] was adopted. In a number of follow-up works [70, 49, 47, 71], researchers advocated the use of Real- Boost. Both Lienhart et al. [46] and Brubaker et al. [48] compared three boosting algorithms: AdaBoost, RealBoost and GentleBoost, though they reach different conclusions as the former recom-

mended GentleBoost while the latter showed RealBoost works slightly better when combined with CART based weak classifiers.

In [49], the authors proposed FloatBoost, which attempted to overcome the monotonicity problem of the sequential AdaBoost Learning.

Viola and Jones [40] trained each node independently. A number of follow-up works showed that there is indeed information in the results from the previous nodes, and it is best to reuse them instead of starting from scratch at each new node. For instance, in [72], the authors proposed to use a "chain" structure to integrate historical knowledge into successive boosting learning.

To set proper thresholds for the intermediate nodes, Viola and Jones [73] proposed a new scheme called asymmetric AdaBoost, which artificially increase the weights on positive examples in each round of AdaBoost such that the error criterion biases towards having low false negative rates. Masnadi-Shirazi and Vasconcelos [74] further proposed a more rigorous form of asymmetric boosting based on the statistical interpretation of boosting [75] with an extension of the boosting loss.

Wu et al. [76] proposed to decouple the problems of feature selection and ensemble classifier design in order to introduce asymmetry.

In [77], the authors proposed to use a a ratio test to determine the rejection thresholds. In [78], a dynamic cascade was proposed, which assumes that the false negative rate of the nodes changes exponentially in each stage, following the idea in [79]. The approach is simple and ad hoc, though it appears to work reasonably well.

To train a cascade detector with billions of examples without explicitly setting the intermediate thresholds, the authors in [79] proposed a scheme that starts with a small set of training examples, and adds to it new samples at each stage that the current classifier misclassifies. In [78] and [80], the authors proposed to use importance sampling to help address the large data set issue.

Moreover, the issues of fast features selection and detection attract much attention. McCane and Novins [81] proposed a discrete downhill search scheme to limit the number of features compared during feature selection. Brubaker et al. [48] studied various filter schemes to reduce the size of the feature pool, and showed that randomly selecting a subset of features at each iteration for feature

selection appears to work reasonably well. In the sparse feature set in [64], the authors limited the granules to be in square shape, which is very efficient to compute in both software and hardware through building pyramids for the test image. Schneiderman [82] designed a feature centric cascade to speed up the detection. A similar approach was deployed in [83] to speed up their locally assembled binary feature based detector.

More complex structures of learning scheme were developed as well. Parallel cascade [71], Detector-pyramid [49], Decision tree I [42] and Decision tree II [84–86] were proposed for multiview face detection.

Some recent works went one step further and did not maintain a fixed subcategory label for the training examples any more. For instance, Kim and Cipolla [87] proposed an algorithm called multiple classifier boosting, which is a straightforward extension of the multiple instance boosting approach in Viola et al. [88].

Here we listed the approaches corresponding to each of the two issues as in the following tables. Table 2.1 shows seven types of features and examples of each type. Table 2.2 listed a bunch of challenges faced by boosting learning in face detection application and the corresponding schemes.

Apart from the boosting learning algorithms inspired by the seminal work of Viola and Jones [40], there were still other learning schemes proposed to achieve good performances such as Bayesian features, SVM, Neural networks, etc. They are generally summarized in Table 2.3.

2.1.2.2 Face detection in color images

Color is an important feature of human faces. Using skin color as a feature to detect a face has several advantages. Color processing is much faster than processing other facial features. Color information is also scale- and pose-invariant. A survey on pixel-based skin color detection techniques can be found in [118].

In [119] Sandeep and Rajagopalan proposed a fast algorithm using a color histogram for human skin in HSV space combined with edge information. No standard dataset or well-known dataset was used but the histogram analysis was done using over 450,000 pixels drawn from different internet sources, so that they belong to people of different races. Specific detection rates were not mentioned.

Table 2.1: Features for face/object detection [3].

Feature Type	Representative Works
Haar-like features and its variants	Haar-like features [40] Rotated Haar-like features [46] Rectangular features with structure [49, 42] Haar-like features on motion filtered image [50]
Pixel-based features	Pixel pairs [89] Control point set [90]
Binarized features	Modified census transform [25] LBP features [26, 27] Locally assembled binary feature [83]
Generic linear features	Anisotropic Gaussian filters [52] LNMF [53] Generic linear features with KL boosting [51] RNDA [54]
Statistics-based features	Edge orientation histograms [56, 55] Spectral histogram [91] Spatial histogram (LBP-based) [62] HOG and LBP [63] Region covariance [92]
Composite features	Joint Haar-like features [47] Sparse feature set [64]
Shape features	Boundary/contour fragments [65, 66] Edgelet [67] Shapelet [68]

Table 2.2: Face/object detection schemes to address challenges in boosting learning [3].

Challenges	Representative Works
General boosting schemes	AdaBoost [40] RealBoost [70, 49, 47, 71] GentleBoost [48, 46] FloatBoost [49]
Reuse previous nodes' results	Boosting chain [72] Nested cascade [71]
Introduce asymmetry	Asymmetric Boosting [54, 93, 74] Linear asymmetric classifier [76]
Set intermediate thresholds during training	Fixed node performance [46] WaldBoost [77] Validation data-based method [48] Exponential curve [78]
Set intermediate thresholds after training	Greedy search [94] Soft cascade [79] Multiple instance pruning [80]
Speed up training	Greedy search in feature space [81] Random feature subset [48] Forward feature selection [71] Use feature statistics [93]
Speed up testing	Reducing number of weak classifiers [49, 95] Feature centric evaluation [82, 83] Caching/selective attention [96] etc.
Multiview face detection	Parallel cascade [71] Pyramid structure [49] Decision tree [42, 84] Vector valued boosting [64, 86]
Learning without subcategory labels	Cluster and then train [97] Exemplar-based learning [98] Probabilistic boosting tree [99] Cluster with selected features [100] Multiple classifier/category boosting [101, 87]

Table 2.3: Other schemes for face/object detection [3].

General Approach	Representative Works
Template matching	Antiface [102]
Bayesian	Bayesian discriminating features [103]
SVM-speed up	Reduced set vectors and approximation [104, 105] Resolution based SVM cascade [106]
SVM-multiview face detection	SVR based pose estimator [107] SVR fusion of multiple SVMs [108] Cascade and bagging [109] Local and global kernels [110]
Neural networks	Constrained generative model [111] Convolutional neural network [112, 113]
Part-based approaches	Wavelet localized parts [114, 115] SVM component detectors adaptively trained [116] Overlapping part detectors [117]

Lim et al. proposed in [120] to use only the color information for face region segmentation. With the input image, they used the predefined skin color boundary and comprehensive color image normalization (composing of row and column normalizations). Pixels that fall within the color boundary are denoted as skin color and finally face candidates can be segmented.

Ghimire et al. [121] combined the skin color and edge information of the input image. They first did the enhancement in HSV color space and then operated segmentation in YCrCb and RGB space. The segmentation result was combined with the edge detection to get the face location.

In [122], Wang et al. combined the Gaussian skin color model in YCrCb space with template matching to do face detection.

In [123] Lin introduced an approach for the detection of faces in color images that is composed of two stages. The first involves converting the input RGB image into a binary image based on color segmentation. In the second stage, 4-connected components are found and labeled. The center of each block and the 3 other blocks are scanned to find the combination that forms an isosceles triangle. These components are potential face regions. These regions were then

tested using a multilayer feedforward neural network.

Other related works used color information either as a pre-filter or post-filter and passed the filtered results to the Haar-like face detector (that is Viola and Jones' work) or LBP face detector. Examples can be found in [124–128].

2.1.3 Illumination processing in face detection

One possible solution to the illumination problem in face detection is to normalize the color if the input image is with colors. Many face detection approaches utilize skin color to locate faces. Thus, they address the illumination problem from the color perspective.

In [120], in order to get rid of the influence of illumination to the image color, the color were normalized based on the collection of prior information of normal skin colors and illuminated skin colors from existing databases. Considering that slightly unbalanced light source, shading, and light intensity do not alter the chrominance characteristics of the skin-color model, illuminated skin colors in chrominance space were extended compared to the skin color under normal lighting condition. Then a row and column normalization were taken so as to remove the lighting geometry and illumination effect.

In [121], to normalize the color of a given image, the image was firstly transferred to be a uniform lighting environment based on a non-linear transformation. In this image enhancement process, the V channel of the HSV space was used. Then the RGB and YCrCb color spaces were used to determine the skin color candidates by setting the color ranges.

In [122], a simple color normalization was firstly operated to remove the color offset via the approach proposed by Gray World [129]. Rather than only using the chrominance components in YCrCb color space, the authors then transferred the original space to a switched YCr'Cb' taking into consideration of the Y component. The Cr' and Cb' were compared with the pre-trained Gaussian skin color model.

Another solution is to adopt illumination-invariant features in the learning-based framework. As have been introduced in the overview, the learning-based

algorithms dominated most of the recent advances in face detection. Sometimes, they use specific features to overcome the illumination problem.

In [25], modified census transform (MCT) was used to describe features. MCT is computed by comparing the neighborhood pixels with their intensity mean. The results are concatenated into an index number representing the pixel's local structure. MCT represents the relative values rather than the absolute values and thus is considered insensitive to monotonic illumination.

In [26], a feature set that improves the original local binary patterns (LBP) [130] called ILBP was proposed to generate more reasonable textural information for face detection task. ILBP, similar to MCT, also proved insensitive to monotonic illumination changes. Another feature set called multiple local binary patterns (MLBP) was introduced in [27] which preserves the illumination insensitivity of the original LBP while providing more complex textural information. Hence, MLBP was used to reduce the required number of features and speed up the training process.

Recently, in [28], the authors combined the characteristics of Haar-like features and LBP features and produced an illumination insensitive feature set named Haar local binary patterns (HLBP). Its insensitivity to illumination mainly relies on the merits of LBP features. It yielded higher true positive rate (TPR) than MCT when more false positives (FP) were allowed. It performed as well as Haar and got better performance than Haar for the cases with adverse imaging conditions.

A third way to overcome the illumination influence in face detection could be executing normalization before the detection. It is found in the literature that some primitive image processing method such as histogram equalization (HE) was applied [40, 131]. However, there is no systematic study about the impacts of prepended normalization on the face detection.

2.2 Face recognition

2.2.1 Overview

Face recognition presents a challenging problem in the field of image analysis and computer vision, and has received a great deal of attention over the past few years because of its many applications in various domains [132–134].

According to a survey about the face recognition techniques [135], face recognition method can be mainly divided into three categories based on the face data acquisition methodology: methods that operate on intensity images; those that deal with video sequences; and those that require other sensory data such as 3D information or infra-red imagery. Here the first category is introduced. Face recognition methods for intensity images can further fall into two main categories: feature-based and holistic [1, 136, 137].

2.2.1.1 Feature-based methods

Feature-based approaches first process the input image to identify and extract (and measure) distinctive facial features such as the eyes, mouth, nose, etc., as well as other fiducial marks, and then compute the geometric relationships among those facial points, thus reducing the input facial image to a vector of geometric features. Standard statistical pattern recognition techniques are then employed to match faces using these measurements.

Early works carried out on automated face recognition were mostly based on these techniques. One of the earliest such attempts was by Kanade [138], who employed simple image processing methods to extract a vector of 16 facial parameters which were ratios of distances, areas and angles (to compensate for the varying size of the pictures) and used a simple Euclidean distance measure for matching to achieve a peak performance of 75% on a database of 20 different people using 2 images per person (one for reference and one for testing).

Brunelli and Poggio [1], building upon Kanade's approach, computed a vector of 35 geometric features (Fig. 2.3) from a database of 47 people (4 images per person) and reported a 90% recognition rate.

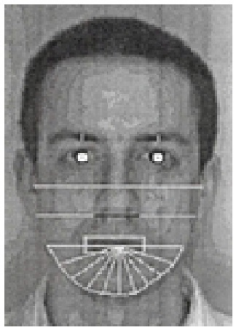


Figure 2.3: Geometrical features (white) used in the face recognition experiments [1].

More sophisticated feature extraction techniques involve deformable templates, Hough transform methods, Reisfeld's symmetry operator, Graf's filtering and morphological operations. However, all of these techniques rely heavily on heuristics such as restricting the search subspace with geometrical constraints [139].

Another well-known feature-based approach is the elastic bunch graph matching method proposed by Wiskott et al. [140]. This technique is based on Dynamic Link Structures [141]. A graph for an individual face is generated as follows: a set of fiducial points on the face are chosen. Each fiducial point is a node of a full connected graph, and is labeled with the Gabor filters' responses applied to a window around the fiducial point. Each arch is labeled with the distance between its correspondent fiducial points. A representative set of such graphs is combined into a stack-like structure, called a face bunch graph. Once the system has a face bunch graph, graphs for new face images can then be generated automatically by Elastic Bunch Graph Matching. Recognition of a new face image is performed by comparing its image graph to those of all the known face images and picking the one with the highest similarity value. Other recent variations of this approach replace the Gabor features by a graph matching strategy [142] and HOG (Histograms of Oriented Gradients [143]).

2.2.1.2 Holistic methods

Holistic approaches attempt to identify faces using global representations, i.e., descriptions based on the entire image rather than on local features of the face.

These schemes can be subdivided into two groups: statistical and AI approaches.

Statistical method

In the simplest version of the holistic approaches, the image is represented as a 2D array of intensity values and recognition is performed by direct correlation comparisons between the input face and all the other faces in the database.

Turk and Pentland [144, 145] realized that projections along eigenpictures could be used as classification features to recognize faces. They employed this reasoning to develop a face recognition system and recognizes particular faces by comparing their projections along the eigenfaces to those of the face images of the known individuals, which is called as PCA, a famous and dominant method in face recognition.

As stated by Moses et al. [146], "the variations between the images of the same face due to illumination and lighting direction are almost always larger than image variations due to a change in face identity". Therefore, they propose using Fishers Linear Discriminant Analysis (LDA) [147], which maximizes the ratio of the between-class scatter and the within-class scatter and is thus purportedly better for classification than PCA.

Numerous variations and extensions of the standard eigenfaces and the Fisherfaces approaches have been suggested since their introduction.

AI method

AI approaches utilize tools such as neural networks and machine learning techniques to recognize faces.

Weng et al. [148] made use of a hierarchical neural network which was grown automatically and not trained on the traditional gradient descent method. They reported good results on a database of 10 subjects.

Recently, Zhang et al. [149] proposed an approach in which a similarity function is learned for describing the level of confidence that two images belong to the same person. The facial features are selected by obtaining Local Binary Pattern (LBP) [130] histograms of the subregions of the face image and the Chi-square

distances between the corresponding LBP histograms are chosen as the discriminative features. The AdaBoost learning algorithm, introduced by Freund and Schapire [69], is then applied to select the most efficient LBP features as well as to obtain the similarity function in the form of a linear combination of LBP feature-based weak learners.

SVM [150] and Hidden Markov models [151] have also been employed for face recognition by several other researchers and have been shown to yield good results.

2.2.2 Illumination processing in face recognition

Face recognition performance are likely to be severely degraded if faces under consideration are posed in varying illumination. It has been shown that the appearance variations of faces caused by illumination are more significant than those by individual identities [34, 35]. To solve the illumination problem of face recognition, numerous illumination processing approaches have been proposed. These approaches can be generally classified into two categories [152]: active approaches and passive approaches. Surveys about both of these categories can be found in [152, 153].

Active illumination processing approaches usually overcome the illumination problem by applying specific imaging techniques in order to capture face modalities (still face images or facial features) which are insensitive to illumination variances. Imaging techniques which capture 3D facial information (i.e., depth data) [134, 154–156] and Infrared (IR) imagery including Thermal-IR [157–160] and Near-IR [161–163] belong to this category.

On the other hand, passive approaches deal with face images already contaminated by the illumination. Related surveys of this class can be found in [164, 165]. They are further grouped into three categories. The first category employs simple operation built on primitive image processing theories such as histogram equalization (HE) [166], gamma correction (GC) [167], edge map, image gradient, etc. These methods have limited performance with large illumination variances.

The second category tries to model the face illumination using a large number of training samples, and thus is time-or-memory-consuming. Illumination cone

model [168] is an example.

The third category attempts to compute the illumination insensitive face representation as a preprocessing step. The majority of these approaches use the Lambertian reflectance model [169] which treats the illuminated face image $I(x, y)$ as a product of the reflectance component $R(x, y)$ and the illumination component $L(x, y)$: $I(x, y)$=$R(x, y)L(x, y)$. The goal is to obtain the component $R(x, y)$ which is related to only the intrinsic facial features. Some methods measure $R(x, y)$ by first estimating $L(x, y)$. For instance, multiscale retinex method (MSR) [170], adaptive single-scale retinex method (ASR) [171] and self-quotient image model (SQI) [172] estimate $L(x, y)$ as a smoothed version of the face image with a common assumption that $L(x, y)$ corresponds to low frequencies. Nonsubsampled contourlet transform-based method (NSCT) [173] and logarithmic total variation model (LTV) [174] estimate $L(x, y)$ by regarding it as a relatively large-scale component, compared to $R(x, y)$. Wavelet-based method (WA) [175] and wavelet-denoising method (WD) [176] factorize the given image into some parts of different resolutions and estimate $L(x, y)$ as the low resolution component. Some other methods directly compute the face representation only pertaining to the component $R(x, y)$. Gradientface (GRF) [177] and Weber-face (WF) [178] are two examples, both of which assume that $L(x, y)$ has slow spatial changes.

Chapter 3

Illumination normalization-based face detection

3.1 Introduction

As mentioned in Chapter 2, to solve the illumination problem in face detection, one solution is to adopt illumination-invariant features such as MCT [25], LBP [26–28], etc. They use the corresponding features to learn and build a face and non-face class model or a boosted classifier and then confirm the patterns of a given image as faces or non-faces. However, these feature-focused modifications based on learning algorithms are still vulnerable to varying illumination. Figure 3.1 shows two input images of the same object (background and face) but with different illumination conditions and their corresponding LBP features. We can see the the so-called illumination-invariant features are not able to overcome the non-uniform illumination. Since the learnt face & non-face class model and the boosted classifier are fixed, there is also a need of "re-learning" when it is wanted to improve the detection performance and the learning process is a heavy task.

Another solution tries to utilize skin color information to balance and compensate the illumination [179, 124, 125]. However, skin colors vary largely in different illumination which can be seen from Fig. 3.2. It is difficult to model the extremely illuminated skin colors without prior knowledge. Moreover, the

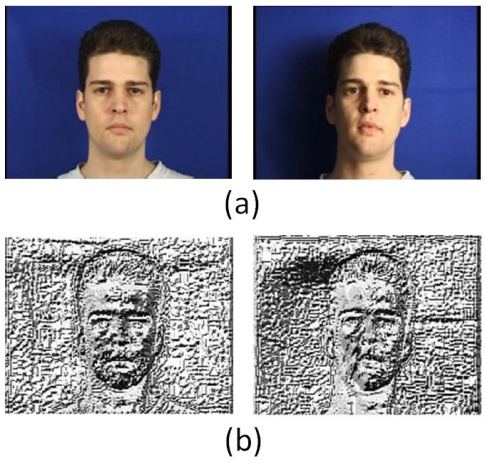


Figure 3.1: Input images of the same object (background and face) (a) and their corresponding LBP features (b).

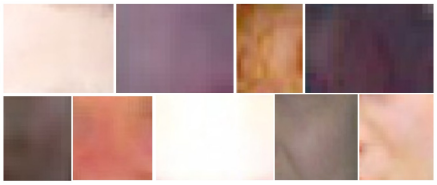


Figure 3.2: Various skin colors under different lighting conditions.

illumination of the test images used in these researches are still even enough and well controlled.

A third way is to do illumination normalization as a preprocessing step. The famous Histogram Equalization (HE) has been applied to face detection [40, 131]. However, to our best knowledge, normalization methods besides the primitive global image processing techniques, have been rarely tried to solve the illumination problem in face detection. Researches about the application of normalization to face detection are hardly found in the literature.

In this study, we focus on the manipulation of combining the illumination normalization and two famous learing-based face detectors (Haar-like and LBP face detectors). It is aimed at maintaining the advantages of pre-trained learning-based face detectors while improving the resistance to varying illumination. Figure 3.3 shows the flowchart of the general framework.

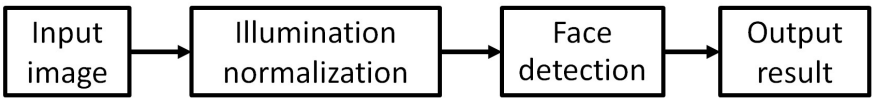


Figure 3.3: General framework.

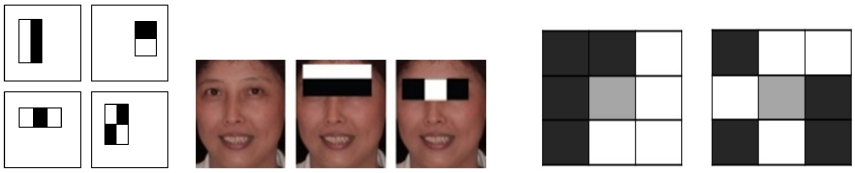
The rest of this chapter is organized as follows. Section 3.2 introduces the learning-based face detectors used in this work. In Sect. 3.3, descriptions of the related normalization methods are provided. In Sect. 3.4, a proposed illumination normalization method is explained in detail. Then the experiments are presented in Sect. 3.5, followed by the final summary in Sect. 3.6.

3.2 Face detectors used in this work

A learning-based face detector refers to the face detection technique which constructs a classifier based on a specific learning algorithm using sophisticatedly selected facial features. Widely used learning algorithms and facial features are boosting-based algorithms [3, 40, 94, 180, 77, 181] and neighborhood comparison (intensity contrast) illustrators [40, 25, 130]. The basic concept of these learning-based methods is to calculate corresponding facial features in multi-scaled analysis windows and to use a specific boosting algorithm for training out of an effective classifier. But being subject to the training samples and efficiency requirement, the learning-based face detector tends to be weak in detecting faces under varying illumination.

In this work, we utilize two learning-based (boosting-based) face detectors [182], namely, Haar-like face detector and LBP face detector[1]. They are advantageous for the richness of feature set, efficient feature selection and fast evaluation of used features. Figure 3.4 illustrates some Haar-like features (see (a)) and LBP features (see (b)). As seen from these features, they largely depend on the discriminability of neighbourhood comparison values. However, harsh illumination makes these patterns useless and makes these two detectors incapable under

[1]Note that the implementation of LBP face detector in [182] uses Multi-block LBP (MLBP) features. For more details about the MLBP features, refer to [27]. In this study, we call MLBP features as LBP features for simplicity.



(a) (b)

Figure 3.4: (a) are some Haar-like features shown in stick figures (left) and of real faces (right); (b) are some LBP features.

varying illumination. Correspondingly, we apply various normalization methods, trying to gain proper contrasts and preserve these patterns in the given image.

3.3 Related normalization methods

The normalization methods applied in this work can be generally grouped into three categories: histogram equalization variants, histogram remapping methods, and face illumination normalization techniques. In this section we will give an overview of them.

3.3.1 Histogram equalization variants

Histogram equalization (HE) is a commonly used method for image contrast enhancement. It obtains a uniform histogram for the output image. However, HE often results in lacks of local enhancement and yields undesirable artifacts and noise [183, 184]. Later, adaptive histogram equalization (AHE) [185] was introduced to improve the local contrast enhancement. But it also suffers from the noise amplification in "flat" regions and "ring" artifacts at strong edges [186] . Contrast-limited AHE (CHE) was proposed to reduce these drawbacks. On the other hand, many bi-histogram equalization methods have been proposed to overcome the problems of HE such as brightness preserving bi-histogram equalization (BBHE) [187], equal area dualistic sub-image histogram equalization (DSIHE), minimum mean brightness error bi-histogram equalization (MMBEBHE). They divide the histogram of an input image into two parts based on the mean value, median value, and minimum mean difference threshold, respectively and then equalize each part independently. In [183], it was considered that the histogram

division approaches in these methods may cause more annoying side effects and thus a new normalization method called range limited bi-histogram equalization (RHE) was proposed. RHE uses Otsu method to choose a proper threshold for histogram division and then does the range limitation and equalization to achieve minimum absolute mean brightness error. Among the above mentioned methods, we will examine HE, AHE, CHE, and RHE.

3.3.2 Histogram remapping methods

Since many normalization methods are based on the image histogram, in [188], a more general concept of histogram remapping was discussed, which remaps the input histogram to a target distribution of different shapes. In this work, we will test three histogram remapping methods with target distribution shape of normal distribution (ND), lognormal distribution (LN), and exponential distribution (EX) [188].

3.3.3 Face illumination normalization techniques

Face detection is highly related to face recognition and numerous face illumination normalization methods were developed to tackle illumination problem in face recognition [152, 153, 189]. In this paper, we will investigate four face illumination normalization techniques for face detection. They are ASR (adaptive single-scale retinex) [171], DCT (discrete cosine transform-based method) [190], TT (Tan and Triggs' method) [191], and WF (Weber face) [178]. All of them are classical or state-of-the-art methods. They are particularly selected in our tests due to their abilities in 1) maintaining most of original visual patterns (unlike GRF [177]), and 2) removing many illumination effects (unlike the early work WA [175]).

3.4 Segmentation-based half histogram truncation and stretching

As illumination has different effects in different regions, a possible way to enhance local facial structures is to firstly partition the input image into several

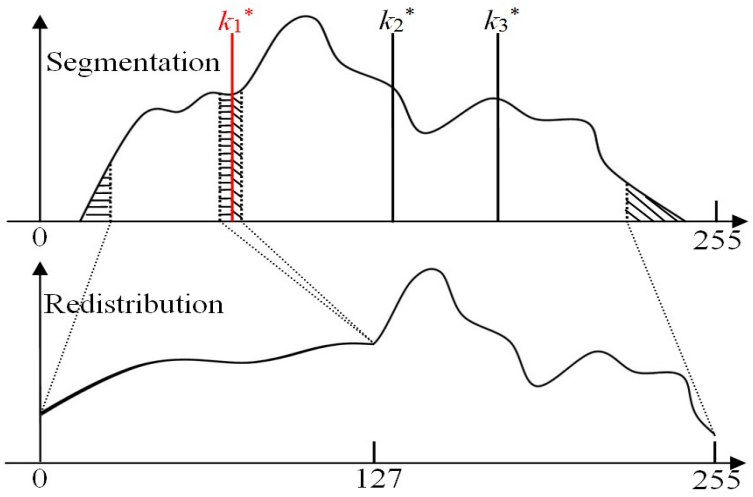


Figure 3.5: Illustration of the proposed method. Segmentation of the input image histogram (top) and output histogram after redistribution (bottom).

regions, and then execute normalization within each region separately. Hence the proposed method is formally defined by two procedures: image segmentation and histogram redistribution. We call it segmentation-based half histogram truncation and stretching (SH).

In [192], face images are roughly partitioned into four geometrically regular parts for illumination invariant face recognition. Yet regular regions are somewhat arbitrary and difficult to conquer the non-uniform illumination problem. AHE, CHE and some bi-histogram equalization methods also adopt the idea of local normalization, but they will lead to undesired artifacts, saturated brightness or lacks of local contrast enhancement. SH uses Otsu method [193] for image segmentation, similar to RHE [183], but the difference is that SH applies multi-segmentation and redistribution of the histogram to a large range. Furthermore, during the redistribution, histogram truncation is used to enlarge the range rather than limiting the range in RHE. Figure 3.5 gives a simple illustration of SH.

3.4.1 Image segmentation

Given an input image $I(x, y)$ with n pixels and a total number of L grey levels, the probability of pixel occurrence is denoted by $p_r(r_q) = n_q/n$, where n_q $(q =$

$0, 1, ..., L-1$) is the number of pixels with r_q grey level.

If intensities of this image are segmented into N spans $(S_1, S_2, \ldots, S_N)$, a series of thresholds $\boldsymbol{k} = \{k_1, k_2, \ldots, k_{N-1}\}$ are required, in which k_1 is the largest grey value in S_1 and so on to k_{N-1} being the largest grey value in S_{N-1} . Then the cumulative probability ω_m, mean level μ_m for each span and the mean intensity of the whole input image μ_T are given by

$$\omega_m = \sum_{q \in S_m} p_r(r_q), \tag{3.1}$$

$$\mu_m = \sum_{q \in S_m} r_q p_r(r_q)/\omega_m, \tag{3.2}$$

and

$$\mu_T = \sum_{q=0}^{L-1} r_q p_r(r_q). \tag{3.3}$$

Otsu method selects the optimal $\boldsymbol{k}^* = \{k_1^*, k_2^*, \ldots, k_{N-1}^*\}$ that maximizes the inter-class variance σ_B^2:

$$\boldsymbol{k}^* = \underset{1 \le k_1 < \cdots < k_{N-1} < L}{\arg\max} \{\sigma_B^2(\boldsymbol{k})\}, \tag{3.4}$$

where $\sigma_B^2 = \sum_{m=1}^{N} \omega_m (\mu_m - \mu_T)^2$. After this segmentation, intensity values within each span are more uniform than those in the global level, and so is the illumination.

3.4.2 Histogram redistribution

Previously, Otsu method segments the input image into N intensity spans; there are N–1 thresholds $k_1^*, k_2^*, \ldots, k_{N-1}^*$ in ascending order. These spans are separated into two groups: the lowest intensity span and the rest intensity spans. The redistribution step contains the histogram truncation and stretching [194], which is operated in the two groups of spans independently. Truncation means to cut off a specific percentage of the lower and upper ends of one span group so as to enlarge the intensity range and remove spurious effects caused by few very bright or dark pixels. For the purpose of enlarging the range in the dark regions, the lowest intensity span is stretched to the low half band of the target distribution and the

rest intensity spans to the high half band. In this way, the target distribution can cover the entire band. There are two assumptions in this redistribution.

1) Faces are probably with low intensity values (i.e., in the dark region) under varying illumination.

2) Even if faces are not wholly in the dark region, the facial organs are probably with low intensity values (i.e., in the dark region) under varying illumination.

In this sense, the stretched redistribution of the lowest intensity span to the low half band is supposed to be able to emphasize the facial structures.

As mentioned above, the region belonging to the lowest intensity span can be labelled corresponding to grey levels $\boldsymbol{r}_{\text{low}} = \{r_0, r_1, \ldots, r_{k_1^*-1}\}$ and the rest regions to grey levels $\boldsymbol{r}_{\text{high}} = \{r_{k_1^*}, r_{k_1^*+1}, \ldots, r_{L-1}\}$. L denotes the total number of grey levels of the input image. The number of pixels to be cut off at both ends of the histogram within the lowest intensity span can be calculated by

$$C_{\text{low}} = \sum_{q=0}^{k_1^*-1} n_q \times \gamma_1, \tag{3.5}$$

where, γ_1 indicates the cut-off percentage. With histogram redistribution, the goal is to get a new intensity value $r'_q(x, y)$ of target pixel (x, y) with initial intensity r_q :

$$r'_q = \begin{cases} 0, & r_q \leq r_s \\ \left(\frac{r_q - r_s}{r_t - r_s}\right) \times 127, & r_s < r_q < r_t \\ 127, & r_q \geq r_v \end{cases}, \tag{3.6}$$

where r_s and r_t are the cut-off intensity values of the lower and upper end of the histogram within the lowest intensity span, respectively. They satisfy the following equations.

$$\sum_{q=0}^{s} n_q \leq C_{\text{low}} < \sum_{q=0}^{s+1} n_q \tag{3.7}$$

$$\sum_{q=t}^{k_1^*-1} n_q \leq C_{\text{low}} < \sum_{q=t-1}^{k_1^*-1} n_q \tag{3.8}$$

It can go through a similar process within the rest intensity spans. Firstly

calculate the number of pixels to be cut off at the end by

$$C_{\text{high}} = \sum_{q=k_1^*}^{L-1} n_q \times \gamma_2, \tag{3.9}$$

where γ_2 indicates the cut-off percentage within the intensity spans under consideration. Then find the cut-off intensity values r_u and r_v, letting

$$\sum_{q=k_1^*}^{u} n_q \leq C_{\text{high}} < \sum_{q=k_1^*}^{u+1} n_q, \tag{3.10}$$

$$\sum_{q=v}^{L-1} n_q \leq C_{\text{high}} < \sum_{q=v-1}^{L-1} n_q. \tag{3.11}$$

Finally the new pixel intensity value $r_q(x,y)'$ to the high half band can be got by

$$r'_q = \begin{cases} 128, & r_q \leq r_u \\ \left(\frac{r_q - r_u}{r_v - r_u}\right) \times 127 + 128, & r_u < r_q < r_v \\ 255, & r_q \geq r_v \end{cases}. \tag{3.12}$$

Figure 3.6 gives some visual samples using different normalization methods. It can be seen that some of the existing methods may change the original facial patterns, some may introduce intensity saturation, and some are likely to result in lacks of contrast enhancement. Our method can efficiently compensate the non-uniform illumination and appropriately enhance the local facial structures.

3.5 Experiments

In the experiments, we used the pre-trained implementation of Haar-like face detector and LBP face detector [182]. This would allow us to understand how illumination normalization can affect the original learning-based face detectors without re-learning. Two databases are used for assessment: extended Yale B database [168] and natural database.

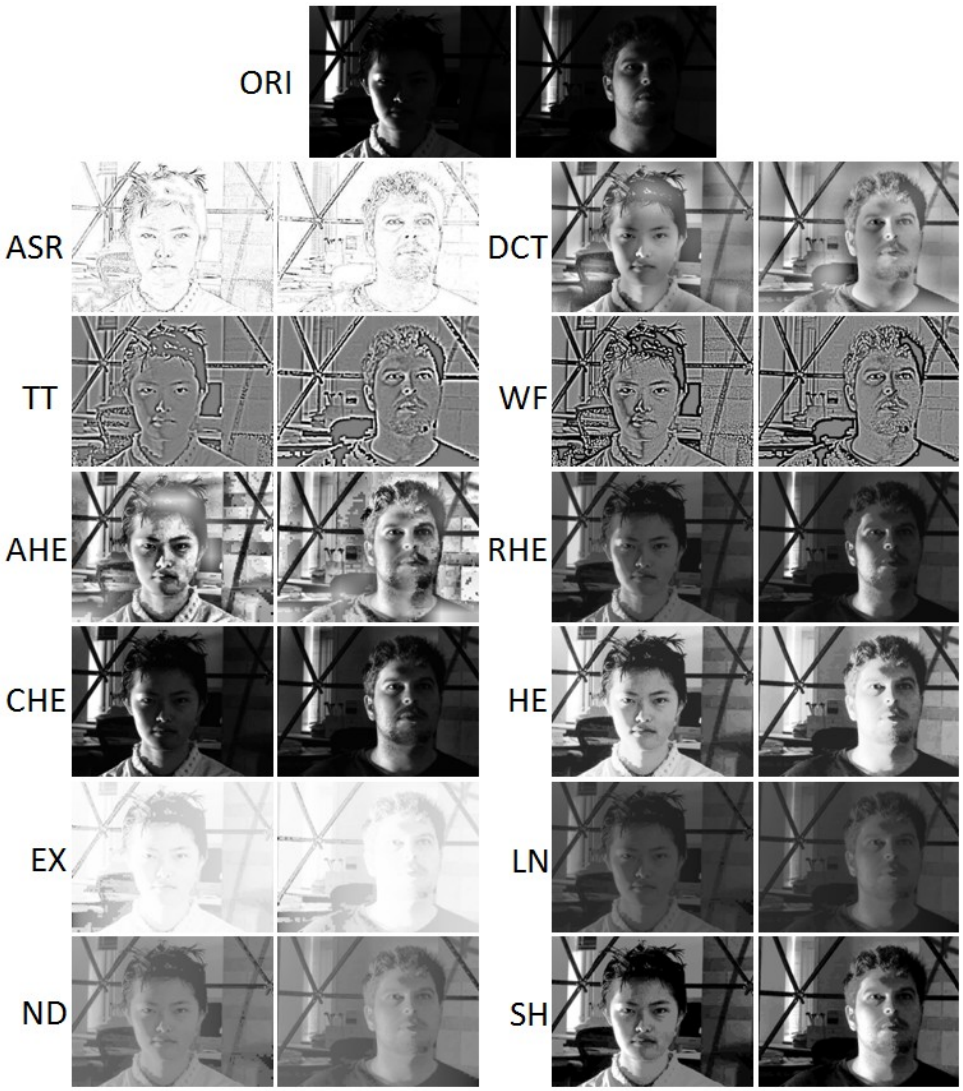


Figure 3.6: Sample images and their corresponding processed images using different normalization methods.

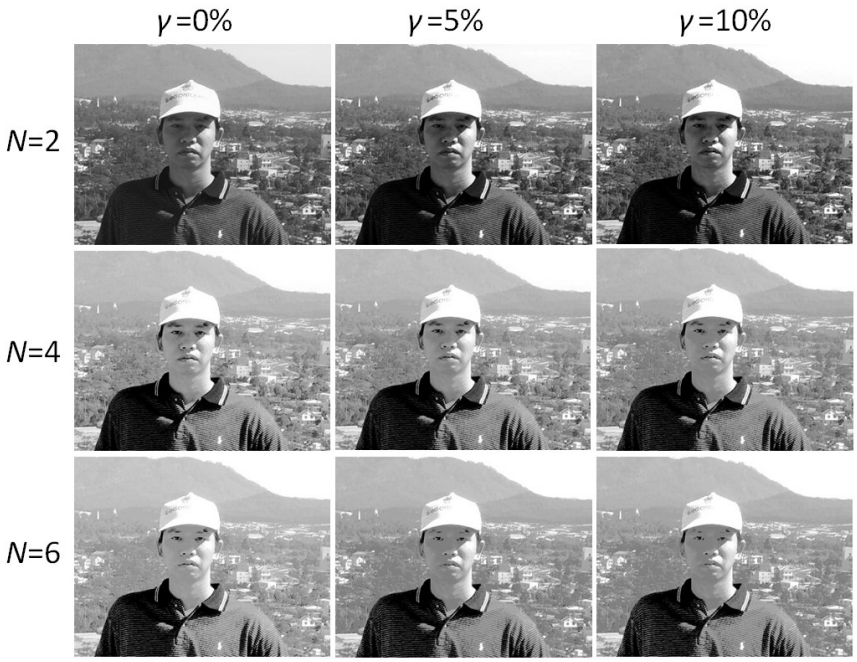


Figure 3.7: Images processed by SH using different parameters.

3.5.1 Parameter selection

The parameters in the applied normalization methods were set as recommended or used in the reference papers. In our method, N is the number of spans segmented by Otsu method; γ_1 and γ_2 denote the percentage of pixels to be cut off within the lowest intensity span and within the rest intensity spans of the input image, respectively. If N is too small, the dark region which probably contains faces or facial organs can not be enhanced enough. If it is too large, the faces or facial organs may not be included in the low intensity span and thus will not be appropriately enhanced as well. γ_1 and γ_2 are used to further enlarge the local enhancement but there will be losses of useful details if they are too large. Figure 3.7 shows some output images processed by our method using different parameters where $\gamma_1 = \gamma_2 = \gamma$. Figure 3.8 shows the relation between the detection rate of Haar-like face detector and the parameters based on the natural database. Since $N = 4$ with $\gamma = 5$ yields the highest rate, we used these values through our experiments.

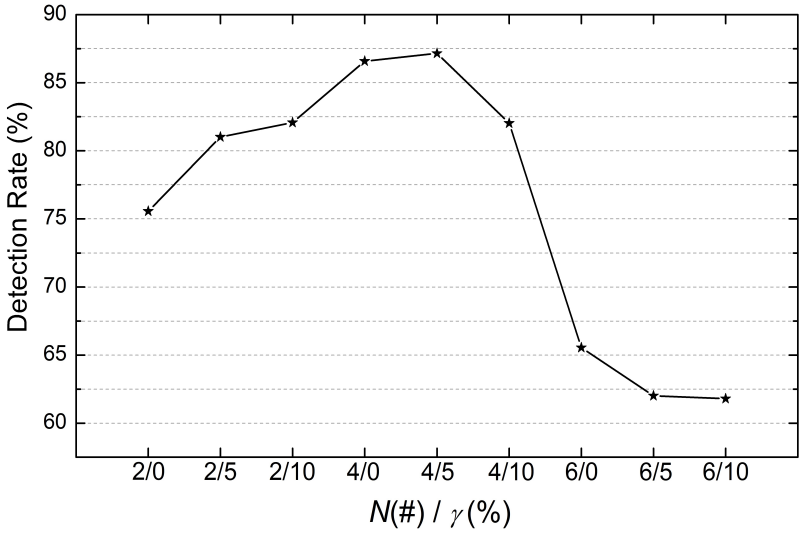


Figure 3.8: Detection rate versus N and γ.

3.5.2 Evaluation protocol

In pattern recognition and information retrieval, the performance of the classifier is assessed through various statistical measures such as precision, recall (sensitivity), specificity, F-measure, etc. These measures are computed by analyzing the outcomes and the actual classification. Face detection task is a two class classification problem. So, in this study, we compute the observed results of two classes (face or non-face) as shown in Table 3.1. Recall and Precision are calculated by

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{3.13}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{3.14}$$

where TP, FN, and FP are the number of true positives, false negatives, and false positives, respectively. Therefore, Recall reflects the ability to detect true faces and Precision reflects the resistance to face-like non-face patterns. Furthermore, F-measure attempts to give a single score that accounts for both precision and

Table 3.1: Outcomes of two class classification

	Actual class (Observation)	
Predicted class (Expectation)	TP Correct result	FP Unexpected result
	FN Missing result	TN Correct absence of result

recall:

$$\text{F-meaure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{3.15}$$

Accurate classifiers anticipate 100% for both precision and recall, and higher the values of F-measure.

3.5.3 Results on Extended Yale B database

Extended Yale B database contains 16128 gray images of 28 human subjects with 9 poses and 64 illumination conditions (576 images for each subject). It is a popular database and is commonly used in the testing task of face recognition and detection. Based on this database, we evaluate how these normalization methods perform with different subjects and different light sources.

Test with Different Subjects. Since the illumination conditions for each subject are similar, we did test on 5 selected subjects (subject 1, 3, 10, 11, and 16) with totally 2880 faces, which vary in gender, region (western/eastern), skin color, and facial structure. Sample images of the 5 selected subjects are shown in Fig. 3.9. Tables 3.2 and 3.3 show the comparative results (in the form of Recall/Precision) with Haar-like face detector and LBP face detector, respectively. Figure 3.10 shows the overall performance distribution with different subjects. The results of the original face detectors (ORI) are also given as the baseline.

From Table 3.2, we can see that the performance of each method varies largely for different subjects. Generally speaking, subject 3, 10, and 16 get similar performance; subject 1 has the best recall; the recall of subject 11 is much lower than the other subjects. ASR reaches very high results for subject 11 who is with the

Figure 3.9: Sample images of subject 1, 3, 10, 11 and 16 (from left to right).

Table 3.2: Results (Recall(%)/Precision(%)) using Haar-like face detector on different subjects

Subject	1	3	10	11	16
ORI	87.15 / 99.60	74.48 / 96.62	72.22 / 99.76	59.20 / 98.27	80.73 / 97.28
ASR	92.36 / 63.26	98.78 / 62.60	98.26 / 97.42	99.83 / 99.65	99.13 / 98.96
DCT	97.57 / 98.60	59.72 / 93.99	78.99 / 100.00	44.79 / 98.47	63.72 / 99.73
TT	63.54 /100.00	10.59 / 100.00	30.38 / 100.00	19.44 / 100.00	24.65 / 100.00
WF	95.49 / 98.21	68.40 / 97.77	67.19 / 99.49	69.44 / 100.00	73.44 / 99.76
AHE	99.13 / 93.00	80.21 / 82.49	84.72 / 93.67	77.43 / 88.14	83.68 / 93.05
RHE	99.65 / 99.31	93.58 / 98.36	90.63 / 93.72	84.03 / 97.38	93.23 / 98.17
CHE	97.92 / 99.65	83.51 / 98.36	88.54 / 99.61	75.35 / 99.09	88.19 / 97.32
HE	99.83 / 98.97	92.36 / 96.90	92.36 / 90.94	85.24 / 81.83	94.79 / 98.56
EX	100.00 / 90.57	96.18 / 80.76	91.49 / 94.28	89.76 / 95.04	94.62 / 99.63
LN	100.00 / 99.65	93.06 / 98.17	96.01 / 98.40	83.68 / 96.59	96.01 / 98.22
ND	100.00 / 100.00	94.44 / 96.28	96.53 / 97.37	86.11 / 87.94	96.35 / 97.54
SH	100.00 / 99.83	94.62 / 97.85	96.70 / 98.76	85.59 / 99.20	96.70 / 98.93

black skin. But its precision is not stable. By contrast, histogram equalization variants, histogram remapping methods, and our method have relatively stable performance. The face illumination normalization techniques DCT, TT, and WF have rather poor performance, even worse than the original Haar-like face detector for some subjects. In Table 3.3, the rank of the performance for different subjects is similar to that with Haar-like face detector. This demonstrates that subjects have a specific influence on detection results. This time, ASR get unsatisfactory results for the entire subjects; DCT, TT, and WF can not improve the original LBP face detector at all. Still, histogram equalization variants, histogram remapping methods, and our method obtain stable and significant improvements of the original face detector for all the subjects (except AHE).

Test with Different Light Sources. All face images of the five subjects were divided into 5 subsets according to different lighting angles, resulting in subset 1 (0° to 12°, 315 images), subset 2 (13° to 25°, 540 images), subset 3 (26° to 50°, 540 images), subset 4 (51° to 77°, 630 images), and subset 5 (above

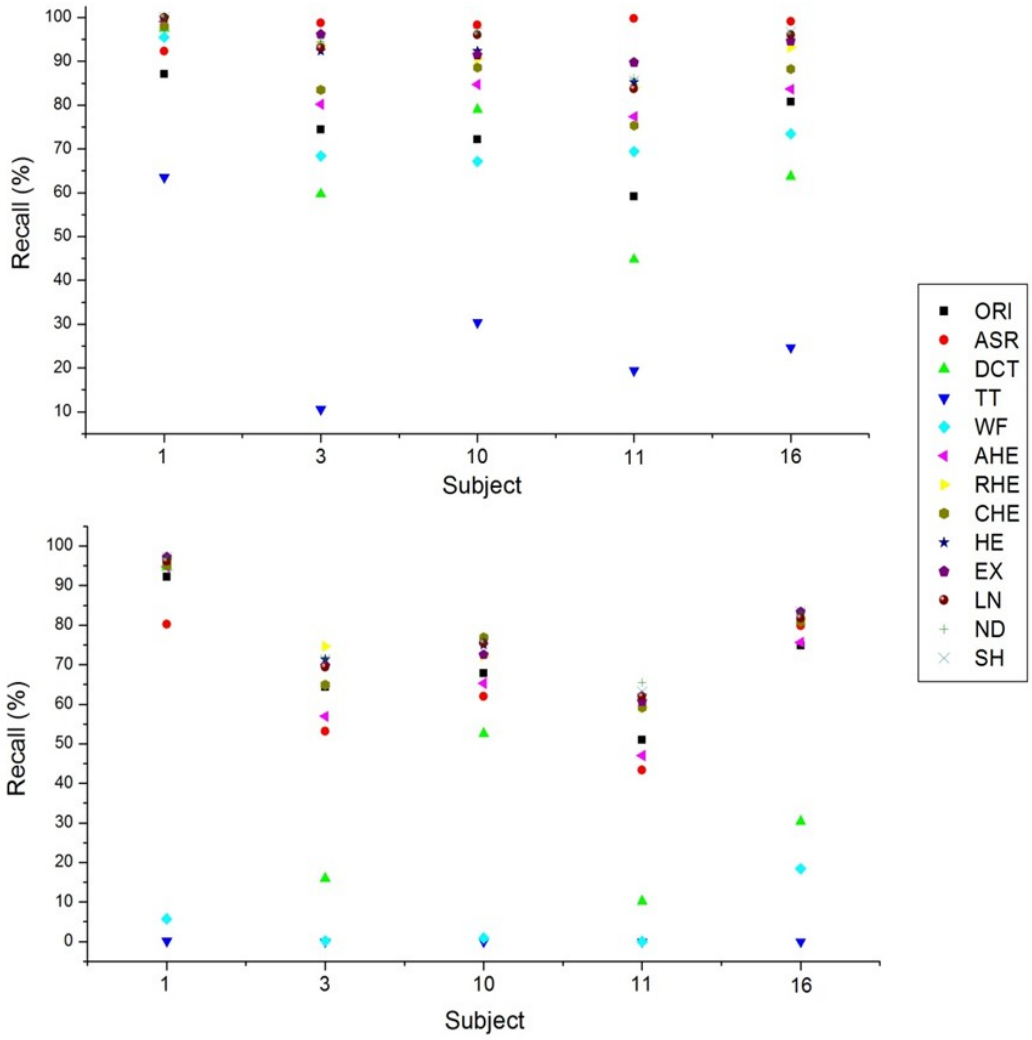


Figure 3.10: Overall performance distribution with different subjects using Haar-like face detector (above) and LBP face detector (bottom).

Table 3.3: Results (Recall(%)/Precision(%)) using LBP face detector on different subjects

Subject	1	3	10	11	16
ORI	92.19 / 94.32	64.41 / 83.56	67.88 / 90.09	51.04 / 62.42	74.83 / 83.85
ASR	80.21 / 83.70	53.13 / 91.62	61.98 / 84.60	43.40 / 92.94	79.86 / 91.27
DCT	94.62 / 85.42	15.97 / 31.08	52.60 / 77.30	10.24 / 48.36	30.38 / 68.36
TT	0.17 / 25.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00
WF	5.73 / 100.00	0.17 / 100.00	0.87 / 100.00	0.00 / -	18.40 / 96.36
AHE	94.79 / 65.23	56.94 / 55.97	65.28 / 73.29	47.05 / 46.64	75.69 / 57.75
RHE	96.53 / 93.45	74.65 / 87.93	72.22 / 81.09	59.20 / 70.60	81.60 / 91.98
CHE	95.14 / 85.09	64.93 / 75.86	76.91 / 85.19	59.20 / 57.80	80.90 / 80.76
HE	96.35 / 94.55	71.35 / 89.15	75.00 / 91.53	62.15 / 73.06	82.99 / 90.87
EX	97.22 / 95.73	69.79 / 92.41	72.57 / 90.48	60.76 / 79.37	83.33 / 92.49
LN	96.18 / 94.86	69.44 / 90.50	75.52 / 85.29	61.98 / 72.86	81.77 / 92.53
ND	97.05 / 94.75	71.53 / 91.35	76.39 / 81.78	65.45 / 75.86	83.16 / 91.06
SH	97.22 / 95.40	71.35 / 92.57	75.69 / 90.27	63.19 / 76.79	83.33 / 93.75

Figure 3.11: Sample images from subset 1 to 5 (from left to right).

78°, 855 images). Sample images of one subject from subset 1 to 5 are shown in Fig. 3.11. As can be seen from the overall performance distribution displayed in Fig. 3.12 and statistical results shown in Tables 3.4 and 3.5, the more extreme the illumination condition is, the worse the general performance of all the methods becomes and, however, the more improvements of the original face detectors can be achieved. The recall of ASR is high and stable with Haar-like face detector under different light sources but low with LBP face detector, compared with histogram equalization variants, histogram remapping methods, and our method. The other face illumination normalization techniques except ASR are unreliable for all of the lighting conditions.

Tables 3.6 and 3.7 give the average performance using Haar-like face detector and LBP face detector, respectively. Since the detection rates of TT with both detectors are extremely bad, its precisions are meaningless and listed just for reference. The case of WF with LBP face detector is similar. It can be noticed that most of the face illumination normalization techniques are not effective in

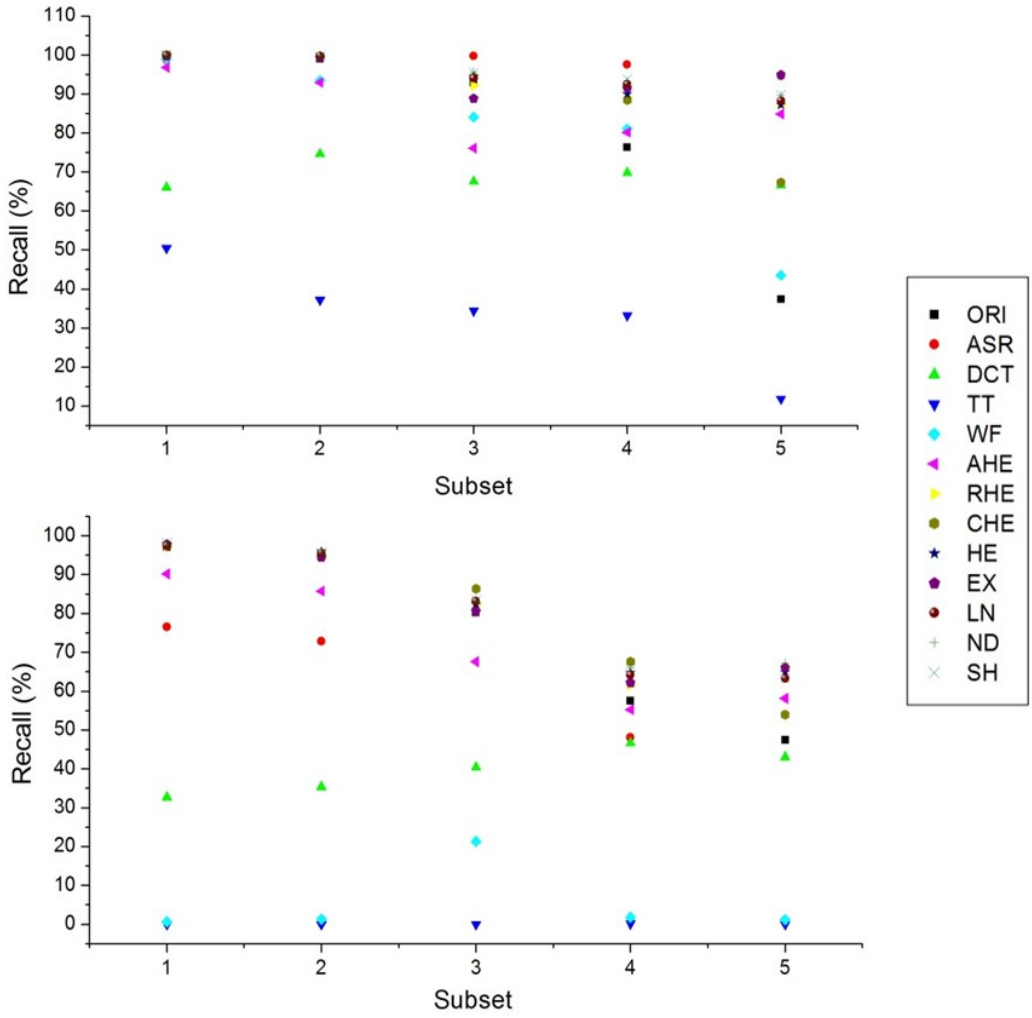


Figure 3.12: Overall performance distribution with different light sources using Haar-like face detector (above) and LBP face detector (bottom).

Table 3.4: Results (Recall(%)/Precision(%)) using Haar-like face detector on different light sources

Subset	1	2	3	4	5
ORI	100.00 / 97.52	99.07 / 99.07	92.96 / 98.82	76.35 / 98.36	37.43 / 96.97
ASR	98.73 / 81.63	99.81 / 81.30	99.81 / 79.62	97.62 / 80.92	94.62 / 80.58
DCT	66.03 / 96.30	74.63 / 98.53	67.59 / 99.18	69.84 / 99.10	66.67 / 97.60
TT	50.48 / 100.00	37.22 / 100.00	34.44 / 100.00	33.17 / 100.00	11.81 / 100.00
WF	99.05 / 99.36	93.52 / 99.02	84.07 / 98.91	81.11 / 99.22	43.51 / 98.41
AHE	96.83 / 88.15	92.96 / 90.78	76.11 / 86.89	80.16 / 91.65	84.91 / 91.55
RHE	99.68 / 97.82	99.44 / 97.99	92.22 / 99.01	89.05 / 96.56	87.25 / 96.38
CHE	100.00 / 98.75	99.81 / 99.45	94.44 / 99.42	88.41 / 98.24	67.37 / 98.29
HE	100.00 / 94.59	99.63 / 94.72	94.44 / 98.46	90.00 / 92.95	87.25 / 88.81
EX	99.68 / 92.08	99.26 / 93.06	88.89 / 90.91	91.59 / 91.59	94.97 / 90.52
LN	100.00 / 99.37	99.81 / 99.26	94.07 / 98.07	92.54 / 97.82	88.30 / 97.55
ND	100.00 / 97.22	100.00 / 97.12	95.74 / 98.29	93.33 / 95.77	89.71 / 92.97
SH	100.00 / 99.37	99.63 / 98.90	95.56 / 98.85	93.81 / 98.50	89.82 / 99.10

Table 3.5: Results (Recall(%)/Precision(%)) using LBP face detector on different light sources

Subset	1	2	3	4	5
ORI	97.14 / 84.30	94.63 / 87.20	80.19 / 85.74	57.46 / 79.56	47.49 / 78.53
ASR	76.51 / 90.94	72.78 / 88.91	80.93 / 91.23	48.10 / 86.57	53.92 / 84.59
DCT	32.70 / 63.19	35.37 / 69.96	40.37 / 67.91	46.67 / 71.88	43.04 / 68.40
TT	0.00 / 0.00	0.00 / 0.00	0.00 / 0.00	0.16 / 16.67	0.00 / 0.00
WF	0.63 / 100.00	1.30 / 77.78	21.30 / 98.29	1.75 / 100.00	1.17 / 100.00
AHE	90.16 / 64.25	85.74 / 65.67	67.59 / 62.82	55.24 / 51.71	58.13 / 57.06
RHE	97.46 / 87.22	94.44 / 89.01	82.22 / 86.05	61.59 / 81.51	65.96 / 83.68
CHE	97.46 / 78.32	94.81 / 80.00	86.30 / 79.66	67.62 / 75.94	53.92 / 71.36
HE	97.78 / 89.53	95.93 / 90.56	82.41 / 87.77	64.76 / 87.74	64.91 / 85.65
EX	97.78 / 94.48	94.44 / 91.56	80.56 / 90.06	62.22 / 89.09	66.08 / 88.84
LN	97.46 / 90.03	95.56 / 90.69	83.15 / 88.04	64.13 / 86.32	63.27 / 83.62
ND	97.78 / 91.12	95.93 / 89.00	83.33 / 86.87	66.03 / 85.42	67.25 / 84.93
SH	98.10 / 91.42	95.56 / 90.53	83.33 / 89.29	65.71 / 89.22	65.73 / 90.06

Table 3.6: Average performance using Haar-like face detector

Method	Recall (%)	Precision (%)	F-measure
ORI	74.76	98.31	84.93
ASR	97.67	80.72	88.39
DCT	68.96	98.27	81.04
TT	29.72	100.00	45.82
WF	74.79	98.99	85.21
AHE	85.03	90.17	87.53
RHE	92.22	97.40	94.74
CHE	86.70	98.81	92.36
HE	92.92	93.27	93.09
EX	94.41	91.49	92.93
LN	93.75	98.25	95.95
ND	94.69	95.85	95.27
SH	94.72	98.91	**96.77**

improving the original detectors except for ASR. ASR accomplishes the highest recall with Haar-like face detector. But the precision and its performance with LBP face detector is not as comparable as the histogram-based methods and our method. When looking at the F-measure, face illumination normalization techniques have the lowest performance generally. Histogram equalization variants perform better. Histogram remapping methods and our method obtain equally good results, ranking the top, where our method yields the best score (slightly better than the histogram remapping methods) with both detectors. On the whole, the improvements of LBP face detector are not as large as Haar-like face detector as LBP features are more robust to illumination than Haar-like features.

3.5.4 Results on Natural Database

The images in Extended Yale B database were taken under controlled light sources. In order to evaluate the illumination insensitive face detection under natural environments, a natural database was created. It contains 840 images with 932 faces. These images were collected from the internet, Bao database [195], and the real-world photographs. Most of the faces are illuminated in extreme or

Table 3.7: Average performance using LBP face detector

Method	Recall (%)	Precision (%)	F-measure
ORI	70.07	83.18	76.06
ASR	63.72	88.18	73.98
DCT	40.76	68.90	51.22
TT	0.03	2.86	0.07
WF	5.03	97.32	9.57
AHE	67.95	59.81	63.62
RHE	76.84	85.41	80.90
CHE	75.42	76.91	76.16
HE	77.57	88.09	82.50
EX	76.74	90.50	83.05
LN	76.98	87.46	81.88
ND	78.72	87.13	82.71
SH	78.16	90.04	**83.68**

non-uniform way and meanwhile embedded in complicated natural backgrounds. Figure 3.13 shows several detection results of our method and the results of the original face detectors. The performance statistics using the two face detectors are given in Tables 3.8 and 3.9, respectively. From Table 3.8, it can be seen that most of the normalization methods (except DCT, TT, and WF) with Haar-like face detector generate much better results than the original detector. Considering the F-measure, SH and EX are the best two methods and HE comes the third. In Table 3.9, the original LBP face detector demonstrates much higher recall than Haar-like face detector and the improvements achieved by the normalization methods are not so significant. But SH and EX still get the highest values of F-measure. These results demonstrate that normalization as a preprocessing is highly potential for improving the pre-trained face detectors. With regard to the general performance, our method is desirable for illumination normalization in face detection.

Figure 3.13: Sample results by using the original Haar-like face detector, Haar-like face detector with our normalization method, the original LBP face detector, and LBP face detector with our method (from top row to bottom row).

Table 3.8: Performance statistics using Haar-like face detector

Method	Recall (%)	Precision (%)	F-measure
ORI	72.86	96.28	82.95
ASR	82.89	96.44	89.15
DCT	55.62	94.40	70.00
TT	41.08	92.05	56.80
WF	36.80	92.90	52.71
AHE	85.33	94.45	89.66
RHE	77.51	95.20	85.44
CHE	82.89	95.76	88.86
HE	84.60	96.78	90.28
EX	86.55	96.59	91.30
LN	79.83	96.03	87.18
ND	82.40	95.87	88.63
SH	87.16	97.41	**92.00**

Table 3.9: Performance statistics using LBP face detector

Method	Recall (%)	Precision (%)	F-measure
ORI	89.36	95.06	92.12
ASR	73.35	94.94	82.76
DCT	78.73	95.69	86.38
TT	12.22	87.72	21.46
WF	20.90	88.14	33.79
AHE	90.46	93.32	91.87
RHE	89.61	94.34	91.91
CHE	90.10	94.97	92.47
HE	91.56	95.05	93.28
EX	91.81	96.41	94.05
LN	89.73	94.59	92.10
ND	90.10	94.49	92.24
SH	92.30	96.79	**94.49**

3.5.5 Discussions

It is noteworthy that EX and ASR with Haar-like face detector get good recalls when the faces are with black skin or under harsh illumination as the conditions in subset 5. It is because that they largely suppress the influence from the harsh illumination or dark skin colors and maintain the Haar-like features.

ASR, DCT, TT, and WF were all initially proposed for illumination normalization in face recognition. However, DCT, TT, and WF perform poorly in face detection and ASR with LBP face detector also yields bad result. Although these four techniques can result in similar face images of one subject after normalizing the illumination, which facilitate face recognition, they are not suitable for face detection. If one wants to make use of the illumination suppression capability of these techniques, a possible way is to apply them to the training samples as well. Nevertheless, the computational cost will be very high and extensive implementations will be very difficult.

On the other hand, the relatively high improvements achieved by histogram equalization variants, histogram remapping methods, and our method indicate that these methods can balance the image illumination and get proper contrasts. Especially, SH enhances the local facial structures more than purely removing the illumination; it is advantageous in terms of stability. It is also supposed that the truncation operation in SH helps decrease the false positives and the results seem to prove so. But unfortunately, SH may sacrifice some details of the image for better local enhancement and this leads to false negatives in some cases.

Another notable thing is that TT yields extremely bad results in face detection while DCT has relatively "good" results (better than TT and WF). However, in face recognition, TT has the best performance among the four face normalization techniques while DCT performs the worst (according to [38]). What's more, AHE, EHE, and CHE were all developed to improve HE in the field of image enhancement. However, HE outperforms them in face detection under varying illumination. Then, we can draw an interesting conclusion that effective normalization methods in face recognition or image enhancement are not necessarily useful in face detection. In summary, illumination processing in face detection is different from that in face recognition and image enhancement. Face recognition

is more focusing on eliminating the illumination to make faces of the same subject more similar to each other. Image enhancement aims to improve the image visual quality without distortions (detail losses). For face detection, emphasizing faces to distinguish them from the background (including illumination) is the core work.

3.6 Summary

In this study, we applied and evaluated a variety of existing normalization methods for face detection under varying illumination. Some of them significantly improved the performance of the pre-trained face detectors (Haar-like face detector and LBP face detector). An interesting conclusion was also made that effective normalization methods in face recognition or image enhancement are not necessarily useful in face detection. Meanwhile, we proposed a new normalization method to compensate non-uniform illumination and enhance the local facial structures. This method proved to be effective in raising the performance of the original face detectors under varying illumination. In the future, we shall study on better trade-off between the local contrast enhancement/illumination removement and the loss of useful information.

Chapter 4

Modified Weberface for face representation

4.1 Introduction

Among the many approaches that use the Lambertian reflectance model to process the illumination in face recognition, some obtain the illumination insensitive component $R(x, y)$ by firstly estimating the illumination component $L(x, y)$ and then solving the inverse computation $R(x, y) = I(x, y)/L^*(x, y)$. It is argued in [177] that, compared to this indirect measure of $R(x, y)$, the direct representation of faces only related to $R(x, y)$ is more effective and robust for illumination insensitive face recognition and a new method called Gradientface (GRF) was proposed in correspondence. It yields better performance than a lot of earlier works. In [178], Weberface (WF) was developed and achieved satisfactory performance as comparable as Gradientface. But the operation employed in Weberface considered only one small scale and ignored the detailed information of face images oriented along different directions which is supposed to be important for face classification.

In this chapter, regarding the limitations of the conventional WF, we propose to improve it in two ways to further exploit the capacity for suppressing the illumination influence while maintaining useful facial features. The oriented Weberface (OWF) and largely-scaled Weberfaces (LSWFs) are created. The former one calculates eight directional face images separately and then concatenates them to

obtain the final output. The latter are produced to get face representations at proper scales. We compare our methods with the conventional Weberface and several other state-of-the-art methods based on three databases. Experimental results show that the proposed methods can achieve fairly encouraging results, outperforming other methods.

In the rest of this chapter, we start with a brief introduction about the conventional Weberface in Sect. 4.2. In Sect. 4.3, we explain and analyze the proposed methods in detail. In Sect. 4.4, the experimental results are presented and discussed. The summary is given in Sect. 4.5.

4.2 Weberface review

Weberface [178] is inspired by the Weber's law which supposes that the relative value is more constant than the absolute value. Concretely speaking, it hypothesizes that the ratio between the smallest sensible change in a stimulus ($\Delta I_{\min}$) and the stimulus with noise (I) is a constant:

$$\frac{\Delta I_{\min}}{I} = k. \tag{4.1}$$

When applied to the illumination insensitive representation of faces, the smallest sensible change is described by the local variation and the noised stimulus corresponds to the illuminated face image. The resultant constant k is the expected illumination insensitive representation. Along this line, Weberface is given by

$$\mathrm{WF} = \arctan\left(\alpha \sum_{i=0}^{P-1} \frac{x_c - x_i}{x_c}\right), \tag{4.2}$$

where x_c denotes the center pixel and x_i $(i = 0, 1, ..., P-1)$ are its neighboring pixels. Therefore, $x_c - x_i$ describes a local variation. P denotes the total number of the pixels in the neighborhood and is a parameter controlling the extent of the local intensity contrast. The arctangent operation is used to avoid the extremely large output. The conventional Weberface sets P as 9, which demonstrates a 3×3 mask for the local operation. The mask is shown in Fig. 4.1. Also, $\alpha = 4$ proved to be the best according to the experiments in [178].

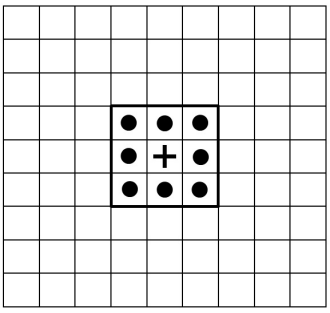


Figure 4.1: The local operation area using the 3×3 mask.

4.3 Proposed methods

In this section, we present the two proposed modifications about the conventional Weberface which generate the oriented Weberface and largely-scaled Weberfaces. These two methods are explained and analyzed in the following consecutive subsections separately.

4.3.1 Oriented Weberface (OWF)

From (4.2), one can know that the difference of intensities within a local neighborhood is critical for maintaining the intrinsic facial features and removing the illumination. Facial features and illumination vary in different way along different directions. However, the conventional Weberface summarizes all the results of the subtraction and division within the neighborhood along different directions together. As a consequence, the facial details oriented in variant directions are blurred.

To overcome this drawback of the conventional Weberface, we compute the Weberface along eight different directions. Let O_i $(i = 1, 2, ..., 8)$ denote the eight directional face images, then they can be expressed by

$$O_i = \arctan\left(\alpha \frac{x_c - x_i}{x_c}\right)\Bigg|_{i\in\{1,2,3,4,5,6,7,8\}}, \tag{4.3}$$

where x_c denotes the center pixel, x_i denotes one of the neighboring pixels of x_c. Then these eight directional face images are concatenated to form the final

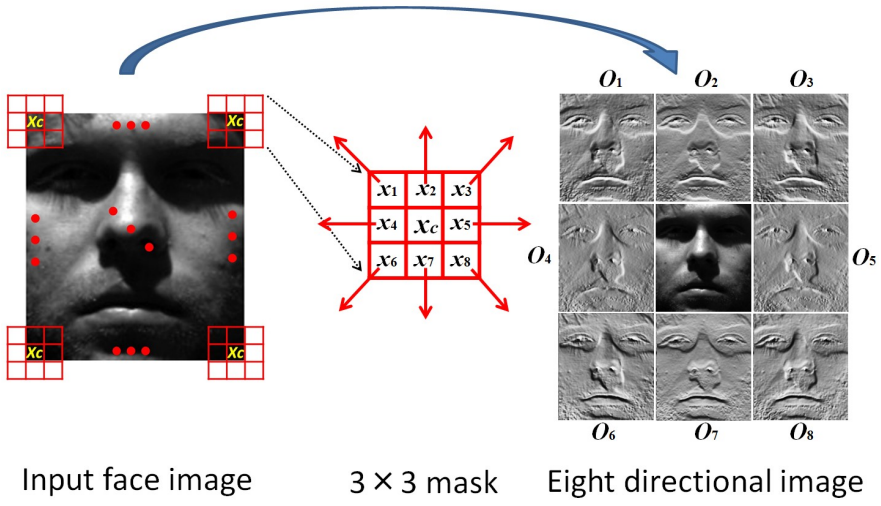


Figure 4.2: The local operation area using the 3×3 mask.

illumination insensitive face representation. We call this representation as an oriented Weberface (OWF) and state it as

$$\begin{aligned} \text{OWF} &= \oplus\{O_i\} \\ &= \oplus\left\{\arctan\left(\alpha\frac{x_c - x_i}{x_c}\right)\right\}\Big|_{i\in\{1,2,3,4,5,6,7,8\}}, \end{aligned} \tag{4.4}$$

where $\oplus\{\cdot\}$ denotes the concatenating operation.

Figure 4.2 illustrates the proposed OWF. The arrows in this figure show the eight directions and the right-most image gives the eight directional face images O_1 to O_8 of the given image corresponding to the eight directions. Since OWF is based on the Weber's law, it keeps all the merits of Weberface. For example, each of the directional face images is computed as a ratio and thus it is robust to multiplicative noise. On the other hand, the oriented patterns have already had good applications to the face recognition in the literature. For instance, the oriented facial information was also considered in [196] and a new method called oriented local histogram equalization (OLHE) was successfully developed. It was applied to maintain facial features while compensating varying illumination.

We have mentioned in the introduction that many illumination invariant facial

features are associated to the reflectance model which is expressed as

$$I(x, y) = R(x, y)L(x, y), \tag{4.5}$$

where $I(x, y)$ denotes the image pixel with illumination, $R(x, y)$ is the reflectance only depending on the intrinsic facial features which is considered to be illumination invariant, and $L(x, y)$ denotes the illumination component at the present pixel (x, y). Next, we would like to prove that OWF is only related to $R(x, y)$ and verify that OWF can represent faces in an illumination insensitive way. According to (4.3), each of the directional face images can be rewritten as

$$O_i = \arctan\left(\alpha \frac{I(x, y) - I(x + \Delta x, y + \Delta y)}{I(x, y)}\right), \tag{4.6}$$

where $i = 1, 2, ..., 8$ and $(\Delta x, \Delta y)$ equals to $(-1, -1)$, $(-1, 0)$, $(-1, 1)$, $(0, -1)$, $(0, 1)$, $(1, -1)$, $(1, 0)$, and $(1, 1)$, respectively when i changes from 1 to 8.

From (4.5), we have

$$I(x + \Delta x, y + \Delta y) = R(x + \Delta x, y + \Delta y)L(x + \Delta x, y + \Delta y). \tag{4.7}$$

Since L is commonly assumed to vary very slightly, it is approximately constant within local neighborhood, that is

$$L(x + \Delta x, y + \Delta y) \approx L(x, y). \tag{4.8}$$

Then the following deduction can be made:

$$\begin{aligned} O_i &\approx \arctan\left(\alpha \frac{R(x, y)L(x, y) - R(x + \Delta x, y + \Delta y)L(x, y)}{R(x, y)L(x, y)}\right) \\ &\approx \arctan\left(\alpha \frac{R(x, y) - R(x + \Delta x, y + \Delta y)}{R(x, y)}\right). \end{aligned} \tag{4.9}$$

It can be noticed from (4.9) that each of the directional face images of OWF only depends on the reflectance component R. By concatenating, the final representation of OWF is also only related to R. This indicates that our method is illumination insensitive. Besides the characteristic of being insensitive to il-

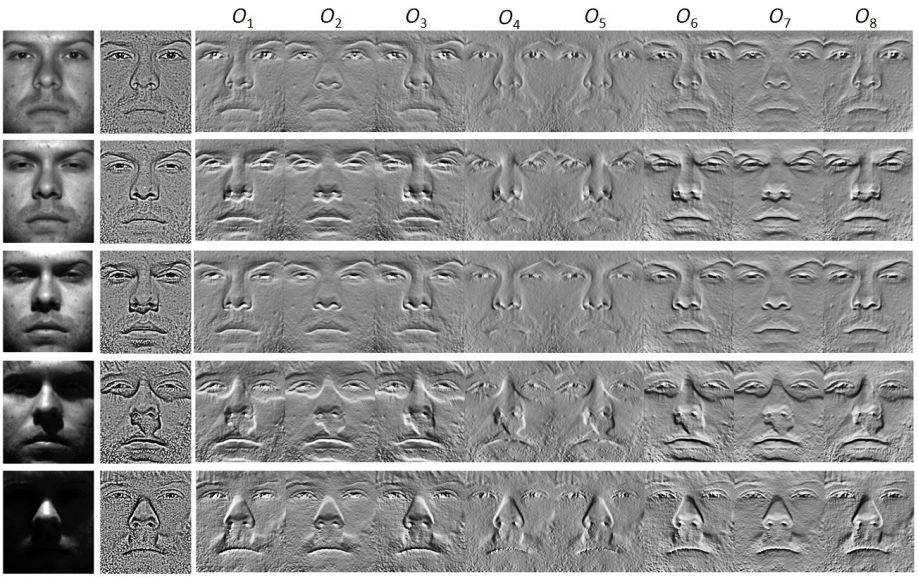


Figure 4.3: The input face images (1st column), the conventional Weberface (2nd column) of corresponding faces, and the eight directional face images of OWF (3rd column) of corresponding faces.

lumination changes, OWF contains the detailed facial features along various directions. These features are useful for discriminating between different persons, which was mixed and missed in the conventional Weberface. On the other hand, note that illumination has different influences along different directions. For example, shadow is likely to enforce negative effects along the boundaries. In this case, separate computation of directional face images can disperse the negative effects. Some visual samples of OWF are shown in Fig. 4.3. It can be seen that our method is able to absorb the specific influence of illumination such as negative effects of shadow boundaries around eyes and nose.

4.3.2 Largely-scaled Weberfaces (LSWFs)

Theoretically, the illumination component L varies slowly in the local area, and this indicates that L remains approximately the same within the small neighborhood. Based on this theoretical premise, Weberface was proved to be only related to illumination invariant component R. It can be also concluded from this premise that the smaller area the operation is executed in, the more illumination-

insensitive the face representation is. Thus, the conventional Weberface uses the operation within a 3×3 squared neighborhood.

However, in real applications, the illumination conditions are far more complex than the simulations given by a physical model. In fact, the "local" operation in illumination insensitive feature extraction does not necessarily correspond to the "nearest" neighborhood. One evidence is found in local normalization (LoN) [197]. LoN adopted the common assumption about the invariability of L in local small area, but finally demonstrated that the best performance was achieved by using the block of size 7 among the sizes of 3, 5, 7, 9, 11, 13. Moreover, in [198], the analysis of pattern description scales of the proposed methods called Weber local descriptor (WLD) was provided. The developed multiscale WLD improved the discrimination of the original one. Hence, keeping this fact in mind, we are motivated to apply more largely-scaled various masks to gain face representations at proper scales which can realize a good balance between illumination compensation and facial feature maintenance.

Here, three largely-scaled Weberfaces are derived using the masks shown in Fig. 4.4. We name them WF2, WF3, and WF4 corresponding to the mask dimension of 5, 7, and 9, respectively. They are used to characterize the illumination insensitive patterns with local salient facial features in different granularities. In Fig. 4.4, the pixels represented by the solid black dots are taken as the neighborhood pixels and the center cross represents the pixel under consideration. Actually, simply enlarging the value of p in (4.2) is a way to make scales larger as well. However, this is likely to overlap the smaller scale mask into the larger one and mix the contributions of differently scaled patterns to insensitivity to varying illumination. The computation of these largely-scaled Weberfaces is almost the same with the conventional one. But we find that the increase of the mask size is often accompanied by the larger intensity difference visually. Since the parameter is used to adjust the intensity difference between neighboring pixels, we slightly decrease the α value for our method with larger mask dimension, that is, $\alpha = 3$ for WF2, $\alpha = 2$ for WF3, $\alpha = 1$ for WF4. Figure 4.5 gives several visual samples using the conventional Weberface and the largely-scaled Weberfaces, from which we can see that the largely-scaled Weberfaces could reduce most illumination effects and make the important facial features mare salient than the conventional

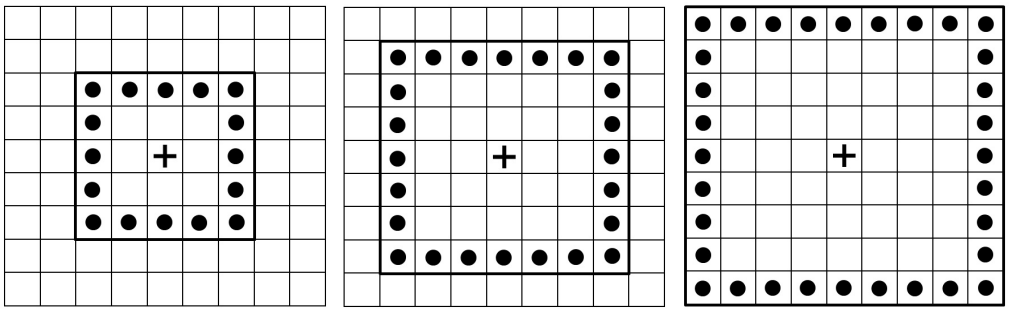


Figure 4.4: Masks used in the largely-scaled Weberfaces with the dimension of 5, 7, and 9, respectively (from left to right).

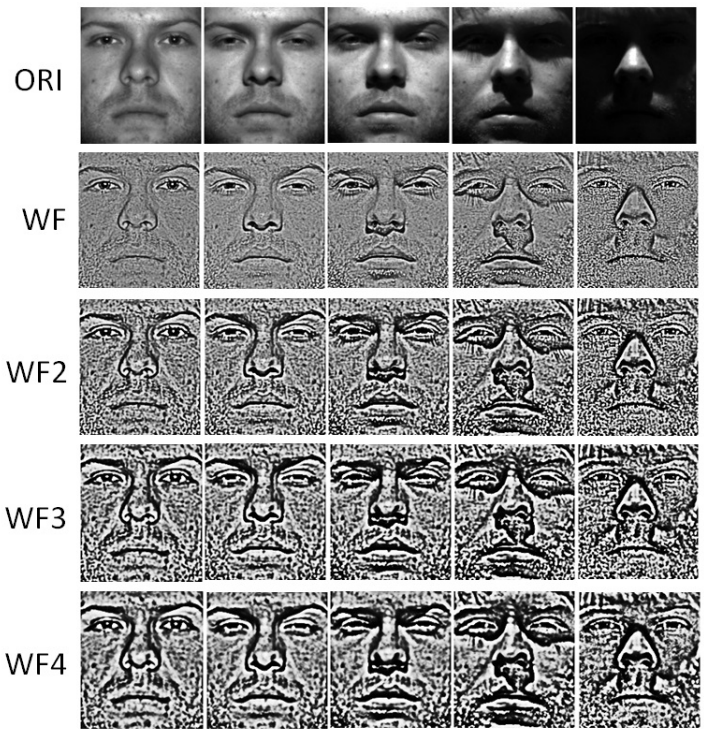


Figure 4.5: The original face images and their corresponding images processed by the conventional WF, largely-scaled WF2, WF3, WF4 (from top row to bottom row).

Weberface. According to the analysis of the largely-scaled Weberfaces, they also keep the merits of the conventional one and are only related to the component R.

4.4 Experiments

4.4.1 Setting

In order to evaluate the proposed methods, some experiments were conducted. Assessment is based on three famous databases with large illumination variations, namely CMU-PIE database [199], Yale B Face Database [168] and Extended Yale B Face Database [200, 201]. Several other methods were included for comparison in our experiments. The parameters in the tested methods are set as recommended or used in the original papers.

4.4.2 Evaluation protocol

During the experiments, we used the raw pixels rather than other learning methods, such as PCA and LDA to evaluate the performance of different methods. This is because the learning methods (PCA, LDA, etc.) have some capability for handling illumination while direct matching method does not have. Therefore, the results achieved by using the raw pixel values can directly show the capability for handling illumination of different methods. The one nearest neighborhood rule with three distance measures–L_1 norm, L_2 norm, and χ^2 distance measure, is used as the classifier. These distance measures are defined by the following three equations, respectively.

$$L_1(\mathbf{X}, \mathbf{Y}) = \sum_{i,j} |X_{i,j} - Y_{i,j}| \tag{4.10}$$

$$L_2(\mathbf{X}, \mathbf{Y}) = \sqrt{\sum_{i,j} (X_{i,j} - Y_{i,j})^2} \tag{4.11}$$

$$\chi^2(\mathbf{X}, \mathbf{Y}) = \sum_{i,j} \frac{(X_{i,j} - Y_{i,j})^2}{2\,(X_{i,j} + Y_{i,j})} \tag{4.12}$$

It is worth noting that the smaller output of each distance measure means the higher similarity. We selected the best performance result from these three distance measures for each of the tested methods as its result for comparison.

The results of the original face images without any processing (ORI) are also given as the baseline.

4.4.3 Results on Yale B

Yale B Face Database is composed of 10 subjects with 9 poses and 64 illumination conditions per pose. We used the cropped version of this database [201] and each face image is of size 168×192. We chose only the frontal faces in our experiments and thus there are totally 640 illuminated face images. They were divided into five subsets according to the lighting angles. The five subsets are: subset 1 (0°∼12°), subset 2 (13°∼25°), subset 3 (26°∼50°), subset 4 (51°∼77°), and subset 5 (above 78°). According to this subdivision, there are 70, 120, 120, 140, 190 images in subset 1 to 5, respectively.

Based on these subsets, two experiments were devised. The first experiment was conducted to use all the images from subset 1 as the gallery images, and the other images from subset 2 to 5 as the probe images. We compared OWF, WF2, WF3, WF4 with the conventional Weberface (WF) [178] and several other state-of-the-art methods including HE [166], DCT [190], WA [175], SQI [172], ASR [171] and GRF [177]. Figure 4.6 shows 10 faces under various illumination conditions and the corresponding illumination insensitive face representations using different methods. Table 4.1 gives the comparative results for different subsets. The distance measure used to get the highest result for each method is also shown correspondingly. As can be seen, the proposed methods achieve extraordinary performance for face images under harsh illumination conditions of subsets 4 and 5 and the recognition rates for these subsets are all above 99%. As for the average performance, OWF significantly increases the original image with no processing, from 42.81% to 99.83%. WF2, WF3 and WF4 even yield 100.00% recognition rates. They outperform HE, DCT, WA, and SQI greatly. They even get better results than ASR, GRF and the conventional Weberface.

Another experiment based on this database was designed to use one image per subject from subset 1 as the gallery images (totally 10 images). These images are with the most neutral lighting condition. Then the rest images in subset 1 and all images of subsets from 2 to 5 were used as the probes for testing.

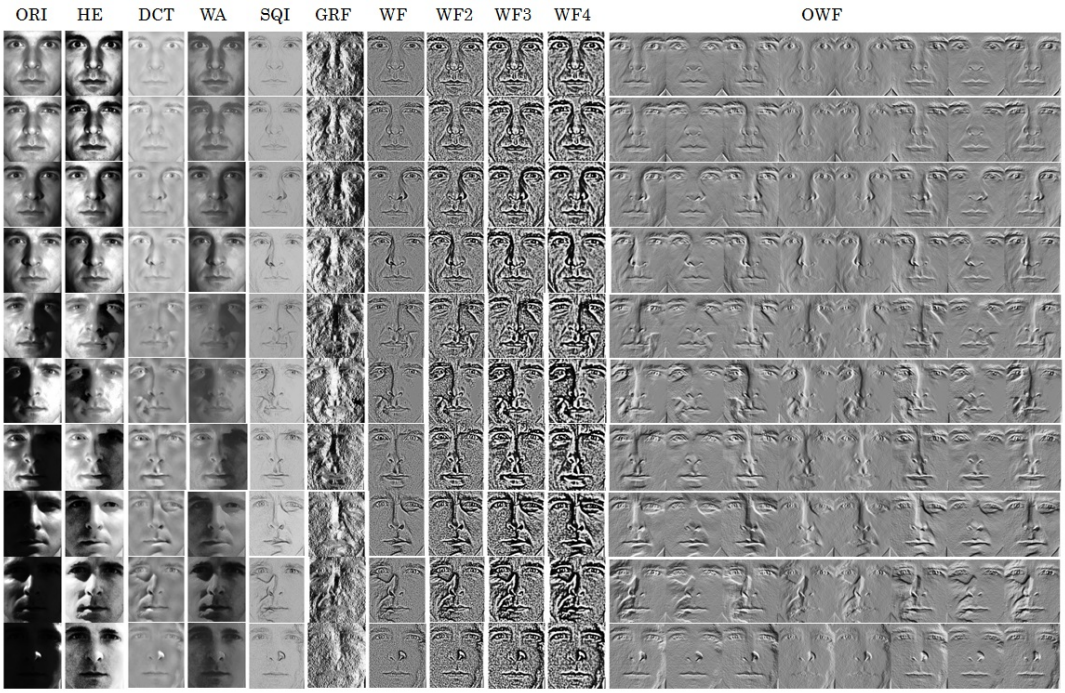


Figure 4.6: Results of different methods on face images with various illumination in Yale B Face Database. The images are (from left column to right column) face images processed with nothing, HE, DCT, WA, SQI, GRF, WF, WF2, WF3, WF4 and OWF (eight directional face images concatenated together).

Table 4.1: Recognition rates (%) on Yale B face database with subset 1 as the galleries.

Method		Subset2	Subset3	Subset4	Subset5	Avg.
ORI	χ^2	95.83	58.33	22.86	14.21	42.81
HE	χ^2	100.00	90.83	41.43	63.16	71.40
DCT	L_2	100.00	94.17	50.00	16.84	58.77
WA	χ^2	95.83	71.67	26.43	26.84	50.70
SQI	χ^2	90.83	87.50	88.57	90.53	89.47
ASR	L_1	99.17	100.00	98.57	99.47	99.30
GRF	L_1	98.33	100.00	92.14	84.21	92.45
WF	L_1	97.50	100.00	98.57	100.00	99.12
OWF	L_2	100.00	100.00	99.29	100.00	99.83
WF2	L_1	100.00	100.00	100.00	100.00	**100.00**
WF3	$L_1/L_2/\chi^2$	100.00	100.00	100.00	100.00	**100.00**
WF4	$L_1/L_2/\chi^2$	100.00	100.00	100.00	100.00	**100.00**

Table 4.2: Recognition rates (%) on Yale B face database with one image per subject in subset 1 as the galleries.

Method		Subset1	Subset2	Subset3	Subset4	Subset5	Avg.
ORI	χ^2	100.00	95.00	50.83	20.71	15.26	46.51
HE	χ^2	100.00	98.33	72.50	47.14	54.74	69.05
DCT	L_2	98.33	94.17	69.17	25.71	14.21	50.48
WA	χ^2	86.67	86.67	60.83	24.29	16.84	46.83
SQI	χ^2	61.67	65.00	45.00	52.86	48.95	53.34
ASR	L_1	100.00	100.00	100.00	96.43	96.32	98.10
GRF	L_1	100.00	97.50	95.00	79.29	63.16	82.86
WF	L_1	90.00	97.50	93.33	94.29	92.11	93.65
OWF	L_2	100.00	100.00	100.00	95.00	98.42	98.41
WF2	L_1	100.00	100.00	97.50	99.29	99.47	99.21
WF3	L_1/χ^2	100.00	100.00	100.00	100.00	100.00	**100.00**
WF4	L_1/χ^2	100.00	100.00	100.00	100.00	100.00	**100.00**

This experiment is more challenging than the previous one since the reference information is much less. The statistic results are shown in Table 4.2. The proposed methods generate the best results again. Note that the results of GRF and WF drop largely from those of the first experiment with the decrease of 9.59% and 5.47%, respectively. By contrast, OWF and WF2 are only 1.42%, 0.79% less than the previous results while WF3 and WF4 both keep the 100.00% recognition rates. ASR also obtains satisfactory results, but is still not as comparable as our methods. These encouraging results validate the effectiveness of the proposed methods when applied for illumination insensitive face recognition.

4.4.4 Results on Extended Yale B

Extended Yale B Face Database consists of 38 subjects, which is an extended version of Yale B Face Database. Each of these subjects was taken under the same conditions as in Yale B and the image size is 168×192 [200, 201]. We also divided this database into 5 subsets according to the various lighting angles. But the number of images in each subset is greatly increased. There are 266, 456, 456, 532, 722 images in subsets from 1 to 5, respectively. With the enlarged

Table 4.3: Recognition rates (%) on Extended Yale B face database with subset 1 as the galleries.

Method		Subset2	Subset3	Subset4	Subset5	Avg.
ORI	L_2	90.13	41.89	5.45	2.63	30.01
ASR	L_2	98.90	99.56	91.54	90.72	94.50
GRF	L_1	99.34	99.78	85.90	59.14	82.73
WF	L_1	98.03	99.78	90.98	93.21	95.06
OWF	L_1	100.00	100.00	98.31	98.20	98.98
WF2	L_1	99.78	99.78	97.74	96.26	98.11
WF3	L_1	100.00	99.78	98.87	97.92	98.98
WF4	L_1	100.00	99.78	99.44	98.61	**99.35**

numbers of subjects, we further assessed the proposed methods and generated more persuasive results.

We compared the proposed methods with the conventional Weberface, ASR and GRF based on the same two experiments with those carried out on Yale B Face Database. In the first experiment, subset 1 was used for reference and the other subsets were used for testing. The second experiment used 38 neutrally illuminated images (one image per subject) from subset 1 as the galleries and the rest images in subset 1 together with subsets 2 to 5 for testing. The corresponding comparative results are shown in Tables 4.3 and 4.4. In Table 4.3, it can be seen that the proposed methods get the highest results and the performance of ASR drops largely compared with that on Yale B face database. The second experiment is rather challenging because the number of subjects to be discriminated become more and available reference information is less. However, from Table 4.4, OWF improves the conventional WF by 14.20% and still maintains the high recognition rate above 90%. The largely-scaled Weberfaces, especially WF3, also significantly outperform the other methods. These results again verify the capability of the proposed methods in illumination insensitive face representation and illustrate their advantages over WF.

Table 4.4: Recognition rates (%) on Extended Yale B face database with one image per subject from subset 1 as the galleries.

Method		Subset1	Subset2	Subset3	Subset4	Subset5	Avg.
ORI	χ^2	95.18	91.45	21.49	4.70	2.77	32.46
ASR	L_1	85.09	99.34	80.48	80.64	73.82	82.54
GRF	L_1	85.96	99.34	84.43	67.67	30.06	67.29
WF	L_1	65.35	97.15	67.98	76.13	73.82	76.86
OWF	L_1	89.47	100.00	86.40	90.79	89.06	**91.06**
WF2	L_1	78.51	99.56	77.41	89.47	82.69	86.01
WF3	L_1	82.46	100.00	81.58	93.42	86.15	89.18
WF4	L_1	82.89	91.89	79.61	90.60	90.86	88.10

Table 4.5: Average recognition rates (%) on CMU-PIE face database.

Method	ORI	ASR	GRF	WF	OWF	WF2	WF3	WF4
	L_2	L_2	L_1	L_1	L_1	L_1	L_1	L_1
Avg.	28.47	91.62	85.18	92.21	96.05	97.95	**98.46**	98.04

4.4.5 Results on CMU-PIE

CMU-PIE face database consists of 68 subjects under large variations in illumination, pose and expression, totally 41368 face images. The illumination subset ("C27", 1428 images) of 68 subjects under 21 different illumination directions was chosen in our experiments. All the images were cropped to the size of 161×161. One image per one subject (totally 68 images) were chosen as the galleries each time and the others were used as the probes. Figure 4.7 shows the recognition rates of different methods versus variant gallery images. The results of ORI are excluded from this figure for convenient display purpose due to its rather poor performance. It is noteworthy that the largely-scaled Weberfaces rank the top and can even get good results with harshly illuminated gallery images such as No.1 to 3 and No. 14 to 17. OWF is not as effective but generates better results than ASR, GRF and WF. The average recognition rates of these methods are given in Table 4.5, which also demonstrates the effectiveness of our methods.

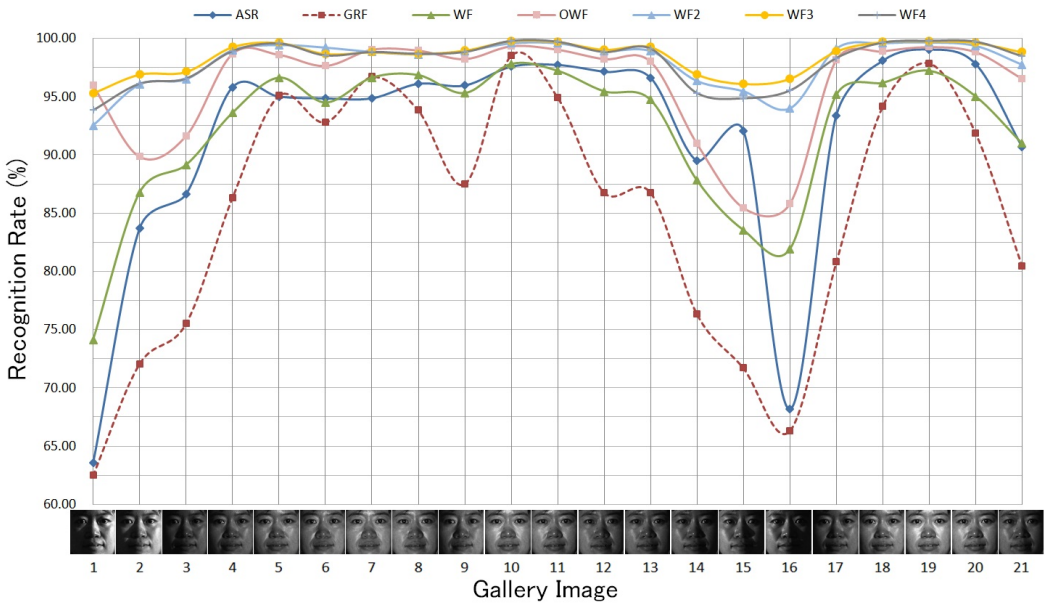


Figure 4.7: Recognition rates of variant methods on CMU-PIE database using different gallery images (face images of 21 illumination conditions) shown in the horizontal axis.

4.5 Summary

Face representation for recognizing faces under varying illumination is a task attempting to not only get representation robust to illumination, but also maintain the intrinsic facial features as much as possible. In this study, considering that the conventional Weberface ignores the facial information oriented along various directions and only adopts a simple scale operation, we proposed to improve it in specific ways. First, we introduced the oriented Weberface (OWF) which makes use of eight directional face images based on the Weber's law and concatenates these images to build an effective face representation insensitive to illumination. It is only related to the reflectance component R according to the reflectance model. Most importantly, it can maintain the detailed facial features oriented along various directions while possessing the ability to cushion some negative influence of illumination. On the other hand, we exploited the effectiveness of operations in larger scales. This was inspired by the fact that the "local" operation in illumination insensitive feature extraction does not necessar-

ily correspond to the "nearest" neighbourhood. As a consequence, we presented three largely-scaled Weberfaces (LSWFs) which characterize the illumination insensitive patterns with local salient facial features in different granularities. We tested the effectiveness of our methods based on three famous databases. Our methods improved the original face images by recognition rates above 50%. They also outperformed the conventional Weberface and several state-of-the-art methods significantly. These encouraging results achieved by our methods confirmed their robustness to illumination changes and justified their contributions to illumination insensitive face recognition. However, the computational cost and memory cost of OWF are rather high and how to combine each of the advantages of OWF and LSWFs remains unsolved. Solving these problems should be the future direction of our research.

Chapter 5

Multifractal analysis-based face representation

5.1 Introduction

To tackle the illumination problem in face recognition, it can be observed that the illumination component is usually associated to the terms of "low frequency", "large scale", "low resolution", or "slow spatial change" which describe the similar textural behavior substantially. In [202], Struc et al. utilized an image denoising technique called non-local means (NLM) for illumination insensitive face recognition. This method constructs the face representation by coding the global textural similarity using the information of the local dissimilarity. The final representation proved to be effective in de-illumination.

In this study, we propose a new illumination insensitive face representation based on multifractal analysis, hence the name of the method Multifractal-Face (MFF). A multifractal spectrum is able to derive global image regularity computed with the textural irregularity in the local neighborhood which is similar to the property of NLM. We utilize this illumination insensitivity to represent faces. On the other hand, the multifractal spectrum is considered discriminative for classification task, which is supposed to be suitable for the subsequent face recognition.

In Sect. 5.2, the introduction of multifractal analysis is given. In Sect. 5.3, the usage of multifractal analysis in our study is discussed in detail. Section 5.4

assesses our methods through experiments on two benchmark databases and Sect. 5.5 concludes this study.

5.2 Multifractal analysis

5.2.1 Fractal fundamentals

The notion "fractal" was firstly introduced by Mandelbrot to describe the non-regular structure of complex natural objects and phenomena [203]. Fractal concept is mainly based on the idea of self-similarity. To be self-similar, an object contains more or less similar parts arising at different scales. This phenomena is precisely illustrated by the Koch's curve as shown in Fig. 5.1. It begins with a line (Fig. 5.1 (a)). In the first step, the middle third is replaced by an equilateral triangle and the baseline is removed (Fig. 5.1 (b)). This procedure is repeated to generate Fig. 5.1 (c) and (d), respectively. In the limit, there is a strictly self-similar structure; each fourth of this structure is a re-scaled copy (by a factor of 1/3) of the entire structure.

Mandelbrot showed that the boundary length of a fractal object can be expressed by a homogeneous power law.

$$L(r) = Nr^{D_H} \tag{5.1}$$

where $L(r)$ is the boundary length measured using a straight line segment of length r, N is the number of such segments required to step along the boundary and D_H is an exponent called Hausdorff dimension.

Rearrangement of the relationship in (5.1) with the limits gives,

$$D_H = \lim_{r \to 0} \frac{\log(N)}{\log(1/r)} \tag{5.2}$$

We have illustrated that Koch's curve is composed of four copies scaled by a factor of 1/3. The D_H for the Koch's curve can therefore be calculated as $log(4)/log(1/3) \approx 1.262$, which is a fractal number. Similar phenomenon can be also described by using the other self-similar structures, e.g. Cantor set and Sierpinski gasket.

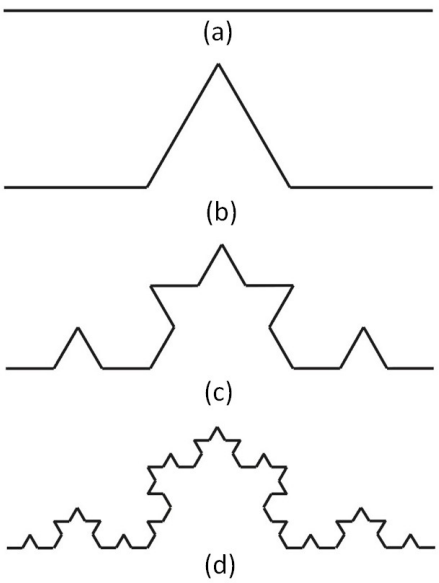


Figure 5.1: (a) Straight line. (b) The middle third of (a) is raised to produce an equilateral triangle. The same process is repeated to generate (c) and (d), respectively.

This non-integer exponent, also called fractal dimension (FD), have been widely used to described the complex objects in all scales.

To compute the FD, box-counting is the most frequently used and most popular method due to its efficiency, accuracy, and ease of implementation [204].

The box-dimension, D_B is defined as follows,

$$D_B = \lim_{\varepsilon \to 0} \frac{\log N_\varepsilon(S)}{\log(1/\varepsilon)} \tag{5.3}$$

where, ε is the side length of the box and $N_\varepsilon(S)$ is the number of boxes required to completely cover the object being measured.

The limit $\varepsilon \to 0$ cannot be applied to natural objects, therefore the gradient of the graph of $\log(N_\varepsilon(S))$ against $\log(1/\varepsilon)$ is calculated to estimate the box-counting dimension or FD. In particular, we compute the gradient of the linear regression line that is estimated for the different sizes of boxes. Figure 5.2 illustrates the computation of box-counting method using Koch's curve [205]. The sizes of the boxes and the number of boxes required to cover the curve are given

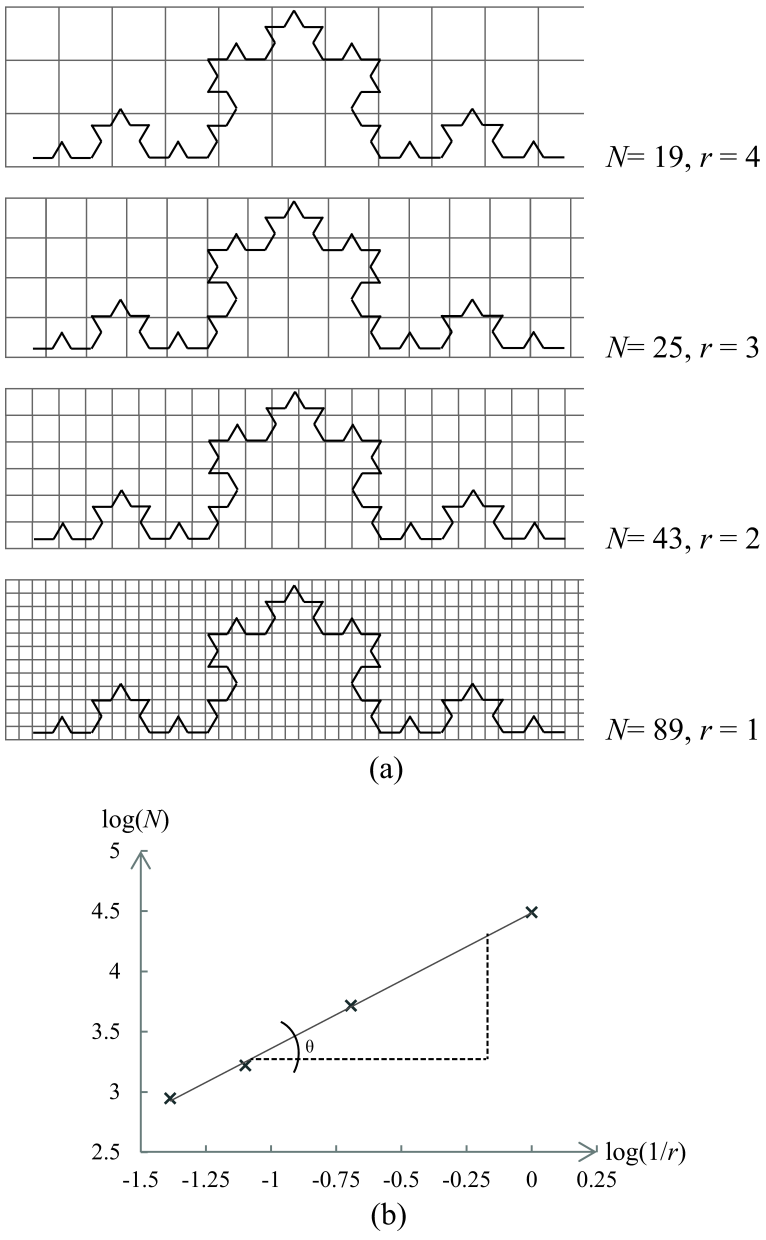


Figure 5.2: FD computation using box-counting method for Koch curve. (a) Different sizes of grids covering the structure. N denotes the total number of boxes required to cover the structure. r denotes the relative size of side length of the box. (b) FD estimation: plot of $\log(N)$ against to $\log(1/r)$.

by r and N, respectively. The plot $\log(N)$ against $\log(1/r)$ is shown at the bottom. The numerical value of the gradient $(\tan^{-1}\theta)$ of the linear regression line is 1.265, which is approximately equals to the value 1.262 computed before.

5.2.2 Multifractal computation

Deterministic structures that are mathematically generated by recursively applying the same rule (e.g. Koch curve, Cantor set and Sierpinski gasket) can be characterized by the same FD in all scales. However, most natural structures are non-deterministic (mathematically non-regular in generating the self-similar parts) and they do not have strict fractal behavior. Therefore, a single FD may not be adequate to characterize such structures. Multifractal analysis is a generalization of fractal analysis which characterizes irregular natural structures as a spectrum of FDs, i.e. multifractal spectrum (MFS). It has demonstrated its power in signal/image classification [206], edge detection, segmentation, characterization [204], retrieval [207], denoising [208], etc., especially in the medical field.

The main idea behind multifractal analysis is to make a description of the singularities of some measure over a region. More precisely, multifractal analysis computes local dimensions of every point in the structure using a finite measure. The entire set of local dimensions is partitioned into non-empty subsets, in which similar or equal local dimensions are grouped. The points corresponded to the local dimensions in each subset that derive a sub-structure, which contains similar structural behavior. Therefore, FD of each substructure is computed and their inter-relationships are examined.

Generally, in the application to digital images, the multifractal computation is performed in two steps.

In the first step, one may find the local irregularity at a point (x, y) using a function $\mu(\cdot)$ called "multifractal measure", which is defined in terms of the Hölder Exponent $h_\mu(x, y)$,

$$h_\mu(x, y) = \lim_{\varepsilon \to 1} \frac{\log(\mu(W_\varepsilon(x, y)))}{\log(\varepsilon)}, \tag{5.4}$$

where $W_\varepsilon(x,y)$ stands for the window of side length ε, centered at (x,y). The output values of $\mu(\cdot)$ inside a series of given windows, $\varepsilon=2k+1$ $(k=0,1,2,...,m)$, are used to form a log-log plot as $\log(\mu(W_\varepsilon(x,y)))$ against $\log(\varepsilon)$. $h_\mu(x,y)$ is then estimated as the slope of the linear regression line through the plotted points. After similar computations for all the pixels, a mapping can be established from the original intensity image to a matrix, which we call α feature matrix (or α-image) reflecting the local textural behavior.

In the next step, the entire range of α values is quantized into R discrete sub-ranges, let α_r be all the α values in the rth sub-range. A binary image I_{α_r} can be formed from α_r by setting the lower and upper limits, $\alpha_{r\mathrm{Low}}$ and $\alpha_{r\mathrm{Up}}$, for the rth sub-range:

$$I_{\alpha_r}(x,y)=\begin{cases}1, & \alpha_{r\mathrm{Low}}\leq\alpha(x,y)<\alpha_{r\mathrm{Up}}\\ 0, & \text{otherwise}\end{cases}. \tag{5.5}$$

This resultant binary image has the same dimension as the α-image. The MFS is calculated by obtaining the FD for each I_{α_r}. To get the FD, a variety of algorithms have been developed [209]. Here, we employ the box-counting algorithm. By covering the whole binary image with a grid of side length ε', one evaluates the number of non-empty boxes $N_{\varepsilon'}$ and estimates the FD of this image as

$$\mathrm{FD}=-\lim_{\varepsilon'\to 1}\frac{\log(N_{\varepsilon'})}{\log(\varepsilon')}, \tag{5.6}$$

which is the box-counting dimension of the binary image. During the computation, $N_{\varepsilon'}$ varies with different values of ε'. In our study, we set $\varepsilon'=1,2,4,...,16$ and display them with the corresponding $N_{\varepsilon'}$ in the log-log plot. The slope of the linear regression line is obtained as the FD. For every binary image, the same operation is taken, and the MFS is finally formed. We can match a FD value corresponding to an α, leading to a $f(\alpha)$ feature matrix (or $f(\alpha)$-image) representing the multifractal features of the given image.

5.3 Multifractal-Face (MFF)

The aim of illumination insensitive face representation is to preserve intrinsic facial features while sufficiently suppressing the illumination cast to the face. So distinguishing the difference between facial features and illumination plays the key role to separate them. As mentioned in the introduction, many existing methods take advantages of the assumed property of illumination and obtain the reflectance component by decomposing the input face image according to the textural information from *frequency*, *scale*, *resolution*, *spatial change speed* or *similarity degree*.

Considering that the illumination can be deemed as the *regular* texture in local region, we utilize the $f(\alpha)$-image which contains both the local and global information of image regularity [210, 206], as an illumination insensitive face representation. In addition, from the Lambertian reflectance model, the face illumination is cast in a multiplicative way. The computation of the ratio of logarithms in (5.4) does not change due to multiplicative changes [211], which also proves that the $f(\alpha)$-image is robust to uneven illumination in face images. What's more, it can be noticed that multifractal analysis provides a distinctive description of the global distribution of local textural behaviors. Thus, the $f(\alpha)$-image is supposed to be capable of extracting intrinsic and discriminative facial features for subsequent face recognition.

We strictly follow the definition of the $f(\alpha)$-image and compute it via the previously presented process which is illustrated in Fig. 5.3. The resultant face representation is named as Multifractal-Face (MFF). We require to carefully decide the intensity function $\mu(\cdot)$ in (5.4) which is used as the multifractal measure. The classification effectiveness heavily depends on the choice of this measure function. Here we investigate three multifractal measures to observe the disparity of the pixel intensities from three different viewpoints. They are Maximum: μ_{Max}, Minimum: μ_{Min}, and Summation: μ_{Sum}. These measures are called capacity measures which were proposed in [212]. They are respectively defined by

$$\mu_{\text{Max}}(x, y) = \max_{(m,n)\in\Omega} g(m, n), \tag{5.7}$$

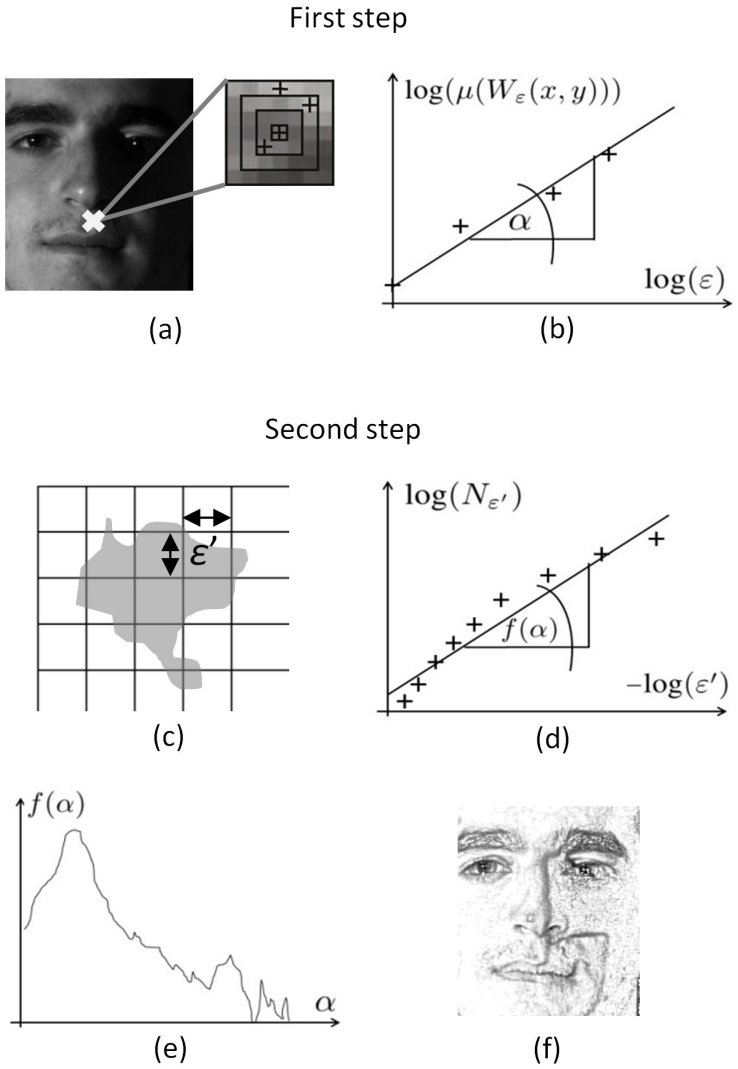


Figure 5.3: Computation of $f(\alpha)$ features. (a) Input image and outcomes ("+") of multifractal measure of a pixel ("×"). (b) Estimation of h_μ (α). (c) Process of non-zero-box counting. (d) Estimation of $f(\alpha)$. (e) Formation of multifractal spectrum. (f) Output face representation ($f(\alpha)$-image).

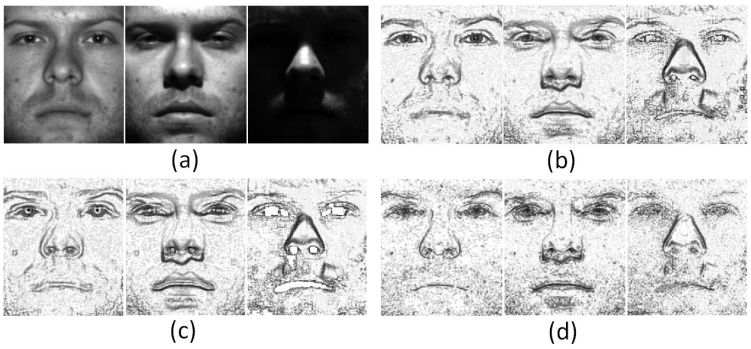


Figure 5.4: Sample face images (a), MFF_{Max} (b), MFF_{Min} (c), and MFF_{Sum} (d).

$$\mu_{\text{Min}}(x, y) = \min_{(m,n)\in\Omega^*} g(m, n), \tag{5.8}$$

$$\mu_{\text{Sum}}(x, y) = \sum_{(m,n)\in\Omega} g(m, n), \tag{5.9}$$

where $\mu_{(\cdot)}(x, y)$ denotes the corresponding measure at point (x, y). Ω is the local window with side length ε centered at the current pixel (x, y). Ω^* stands for all of the non-zero pixels within Ω. $g(m, n)$ is the pixel intensity at point (m, n).

Figure 5.4 shows some samples of MFF using different measures named MFF_{Max}, MFF_{Min}, and MFF_{Sum}, respectively. We can see that MFF_{Max} seems better at maintaining salient facial details while compensating illumination, which is also verified by the experimental results.

5.4 Experiments

To evaluate the proposed face representation method, some experiments were conducted on two benchmark databases: CMU-PIE Database [199] and Extended Yale B Face Database [200, 201]. Several other methods were also implemented for comparison in our experiments, namely HE, GRF [177], ASR [171], NLM [202], and WF [178]. The parameters in all of these algorithms were set as recommended in the original paper.

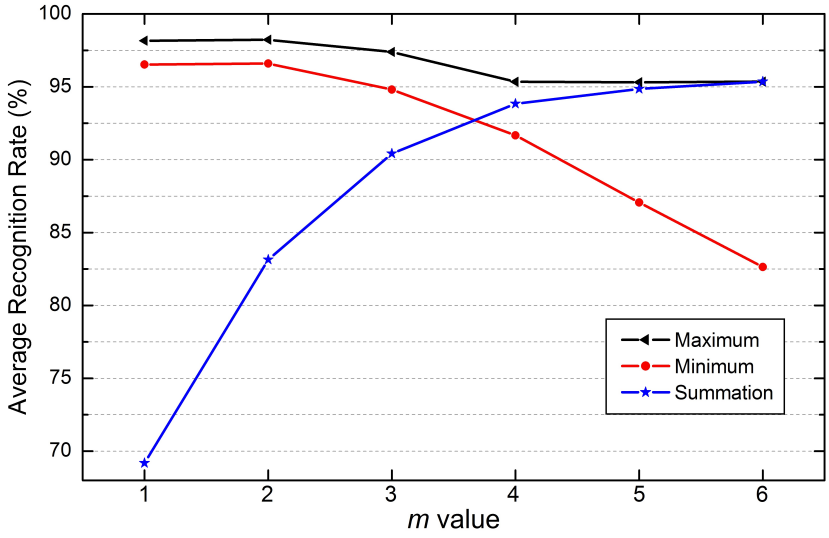


Figure 5.5: Average recognition rate versus m value.

With regard to the recognition protocol, we use the one nearest neighborhood rule with three distance measures: L_1 norm, L_2 norm, and χ^2 distance measure. We selected the best result from the three distance measures for each of the tested methods as its result for comparison. The results of the original face images without any processing (ORI) are also given as the baseline.

5.4.1 Results on CMU-PIE

CMU-PIE database consists of 68 subjects under large variations in illumination, pose and expression, totally 41368 images. The illumination subset including 68 subjects under 21 different illumination directions was chosen in our experiments. All the images were properly cropped to the size of 161×161 with only the face region. One image per subject (totally 68 images) was chosen as the galleries each time and the others were used as the probes.

There is a parameter m needed to be set in our method (refer to (5.4)). It decides the number of local windows (m+1) that are used to estimate the Hölder Exponent h_μ. It also decides the maximum size of the window ($2m+1$). The window size ε should not be too large if μ_{Max} or μ_{Min} are used as the measure function. Otherwise, there may be no variation of h_μ with ε when meeting the

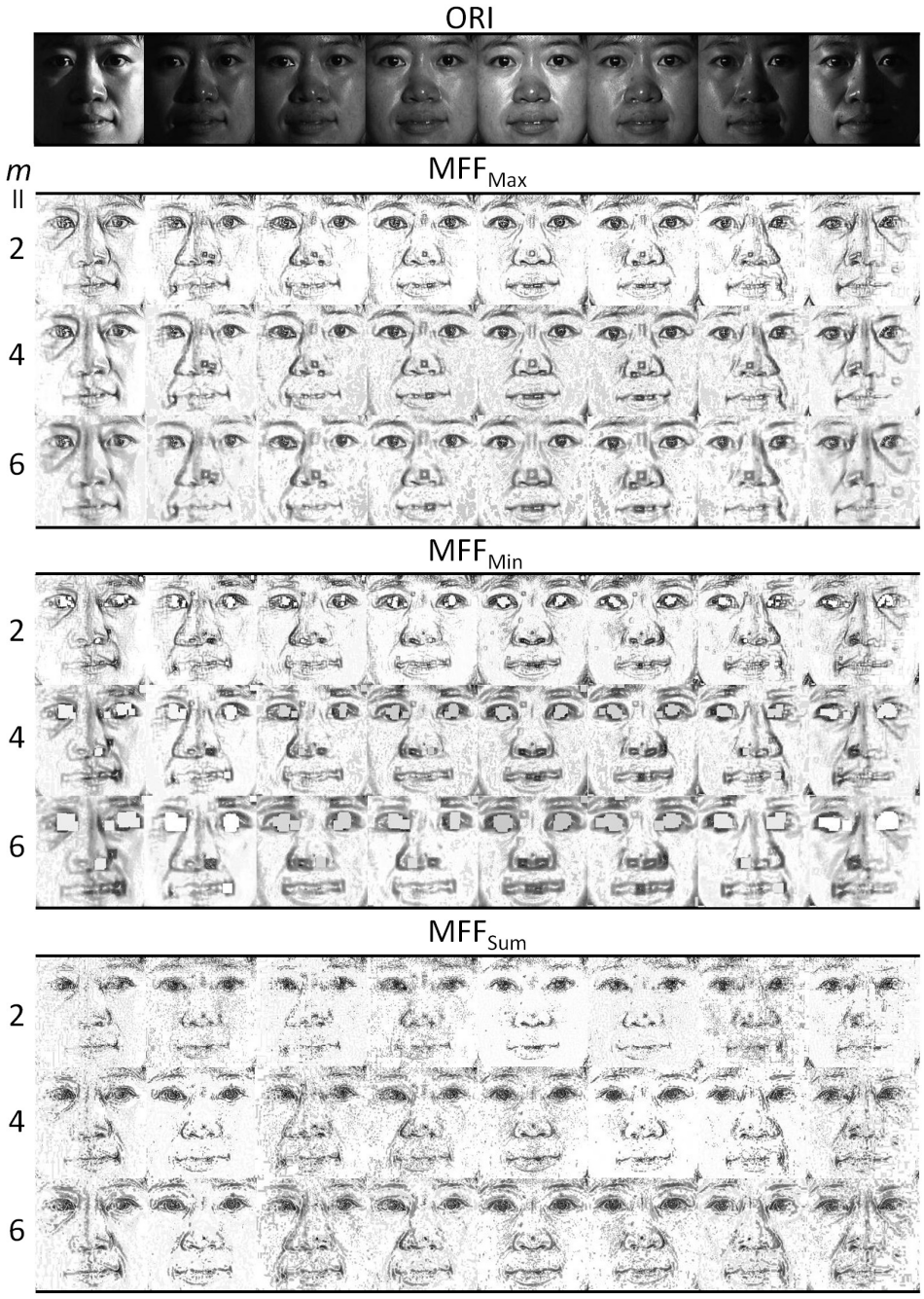


Figure 5.6: Sample results of $\mathrm{MFF}_{\mathrm{Max}}$, $\mathrm{MFF}_{\mathrm{Min}}$, and $\mathrm{MFF}_{\mathrm{Sum}}$ with m set as 2, 4, 6, respectively.

global maximum or minimum intensity in the neighborhood of a pixel, which leads to inaccurate estimation of h_μ [207]. We tested six cases with m value changing from 1 to 6. Figure 5.5 shows the relationship between the average recognition rate and the varying m value. The performance of MFF_{Max} and MFF_{Min} decreases with the increase of the window size. This may be a good justification of the above statement.

Actually, selecting the m value is to realize a nice trade-off between the facial feature maintenance and illumination redundancy. Figure 5.6 gives visual samples of MFF with different settings. We can see that when m increases, both the facial features and illumination effects (such as shadow boundaries) become more extrusive. From Fig. 5.5, we found the local highest rate when μ_{Max} is used with $m = 2$ (i.e. 3 local windows) which realizes a better trade-off. Since MFF_{Min} and MFF_{Sum} are relatively not reliable in terms of both stability and recognition accuracy, we apply MFF_{Max} with $m = 1, 2, 3$ (MFF1, MFF2, MFF3) in the following experiments.

Figure 5.7 shows the recognition rates of different methods versus variant gallery images. The results of ORI are excluded from this figure for the purpose of convenient display due to its rather poor performance. It is noteworthy that our methods get high and stable results even with harshly illuminated gallery images. The statistical results shown in Table 5.1 also demonstrates the effectiveness of our methods, in which MFF2 achieves the best performance.

5.4.2 Results on Extended Yale B

Extended Yale B Face Database is composed of 38 subjects with 9 poses and 64 illumination conditions per pose. It is the extended version of, and contains more subjects than, the famous Yale B Face Database [168] and therefore more challenging. We used the cropped version of this database and each face image is of size 168×192. We chose only the frontal faces in our experiments and thus there are totally 2432 illuminated face images. They were divided into five subsets according to the lighting angles: S1 (0°∼12°), S2 (13°∼25°), S3 (26°∼50°), S4 (51°∼77°), and S5 (above 78°). As a result, there are 266, 456, 456, 532, 722

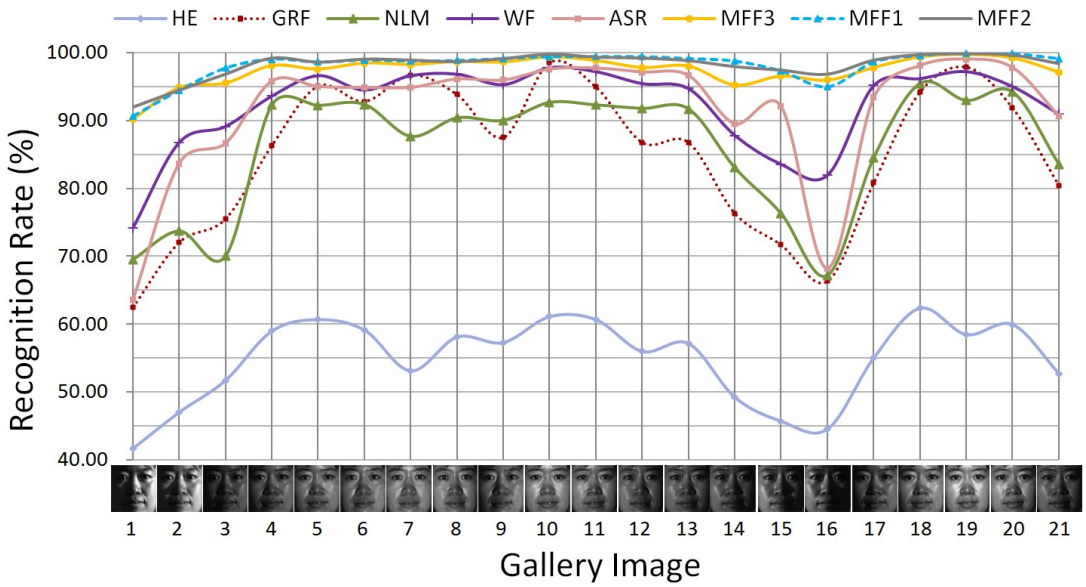


Figure 5.7: Recognition rates versus galleries under different illumination conditions visually shown along the horizontal axis.

Table 5.1: Recongition rates (%) on CMU-PIE.

Method		Avg.
ORI	L_2	28.47
HE	χ^2	54.77
GRF	L_1	85.18
NLM	χ^2	85.88
ASR	L_2	91.62
WF	L_1	92.21
MFF3	L_2	97.38
MFF1	L_2	98.16
MFF2	L_2	**98.22**

Table 5.2: Recongition rates (%) on Extended Yale B.

Method		S1	S2	S3	S4	S5	Avg.
ORI	χ^2	95.18	91.45	21.49	4.70	2.11	32.46
HE	χ^2	98.68	94.30	44.74	17.11	28.53	48.29
GRF	L_1	85.96	99.34	84.43	67.67	30.06	67.29
NLM	χ^2	89.91	99.12	73.03	64.29	59.70	73.64
WF	L_1	65.35	97.15	67.98	76.13	73.82	76.86
ASR	L_1	85.09	99.34	80.48	80.64	73.82	**82.54**
MFF3	L_2	97.81	99.78	88.16	72.93	58.73	79.03
MFF1	L_1	98.25	100.00	91.01	79.89	55.96	80.37
MFF2	L_1	94.74	100.00	87.50	79.51	60.39	80.62

images in S1 to S5, respectively. The sample face images processed by different methods are displayed in Fig. 5.8.

The 38 most neutrally illuminated images (one image per subject) from S1 are used as the galleries and the rest images in S1 with all images in S2 to S5 as the probes. Table 5.2 gives the recognition rates of each tested method on different probe subset. It can be seen that MFF1, MFF2, MFF3 improve the average recognition rate of ORI from 32.46% to 80.37%, 80.62%, and 79.03%, respectively. They also outperform HE, GRF, NLM and the recently proposed WF significantly. Although ASR has averagely better results by around 2%, its recognition accuracy on S1, S2, and S3 are much worse than our methods. Since S1, S2, and S3 contain the images with less harsh illumination, results on these subsets reflect the ability to maintain discriminant facial features. This indicates that MFF, based on multifractal analysis, is more discriminative for face recognition. In summary, illumination insensitive face recognition demands twofold considerations: insensitivity to illumination and discriminative ability for classification. MFF achieves a satisfactory balance between them, compared with other methods.

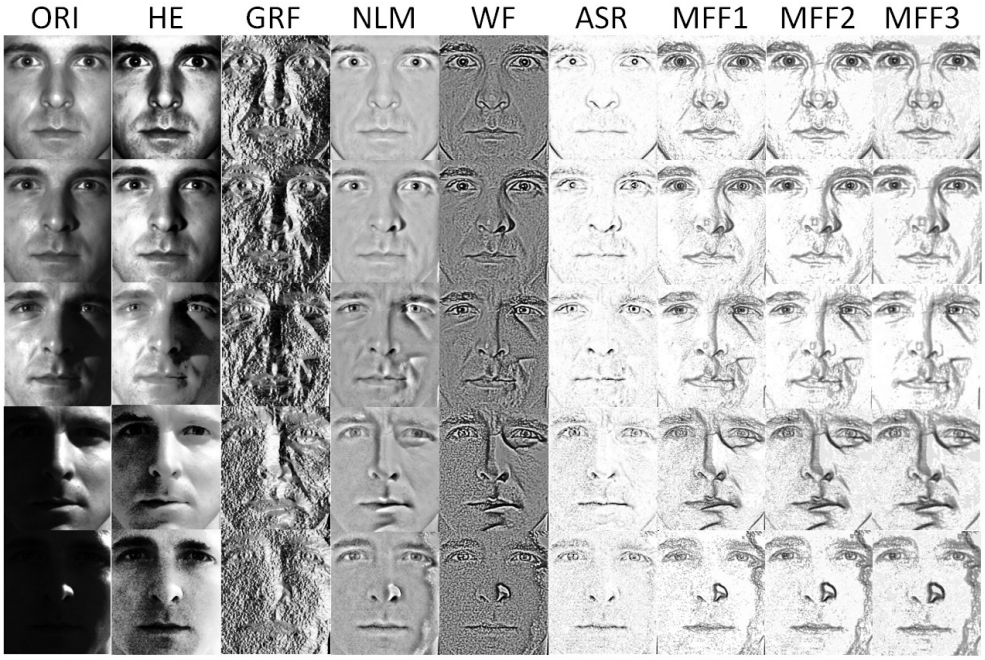


Figure 5.8: Sample results using various methods.

5.5 Summary

Multifractal analysis provides rich and discriminative textural information and has successful applications in various classification tasks. Inspired from this, we construct a face representation named Mulrifractal-Face (MFF) by extracting $f(\alpha)$ features. MFF generates the description of image regularity in local and global level, which is considered discriminative in both face recognition and illumination removal. Its robustness to illumination and discriminative ability are confirmed by the experiments, where satisfactory face recognition accuracy is achieved, compared with several state-of-the-art methods.

Chapter 6

Survey on illumination processing in face recognition

6.1 Introduction

As can be observed from the available literature, most of the surveys on the illumination processing in face recognition only gave an introduction about the methodological classification with an overview of existing algorithms [152, 153] or evaluations of several methods based on the recognition performance [213, 188, 214]. However, among the numerous approaches for illumination processing in face recognition, some are based on the same model, some share common fundamentals, and some possess special trains of thinking. Moreover, there also exist unsolved problems in most of these approaches. Analysis of these points seems necessary and inspiring, which is also the purpose of this study.

In this chapter, we summarize and analyze several noteworthy points for the illumination processing of face images via reviewing various representative approaches. Specifically, we discuss about a principle of the reflectance model and its implementations in various approaches. Then we consider about the contributions of basic processing methods to the illumination normalization, assess the feature description domains and feature scales, and describe an unsettled common problem. We also introduce an essential question of what we should actually normalize. By these summary and analysis, it is motivated to arouse relevant academic interests and enlighten possible directions for improvements or

creation of novel methods to illumination insensitive face recognition. The active approach and illumination modeling are not included because the former one is less pertaining to algorithms and the latter is usually computationally expensive or requiring much prior knowledge. The approaches discussed in this study mainly range from illumination correction to illumination insensitive features extraction. Same approaches may co-occur in different sections. This is inevitable and does not matter since what we would like to focus on is not a specific approach but the viewpoints of interests for illumination processing.

In the rest of this chapter, we start with a brief introduction of the commonly used reflectance model in Sect. 6.2. Section 6.3 gives the analysis of how various approaches correspond to the reflectance model. Discussions about different considerations are shown in Sect. 6.4. An essential question is discussed in Sect. 6.5. In Sect. 6.6, some experimental results of summative evaluations are presented. The final conclusion is made in Sect. 6.7.

6.2 Reflectance model

The reflectance model is a widely used concept and foundation for illumination processing in face recognition. It expresses a given face image $I(x, y)$ under some illumination condition as a product [215]:

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

where R is the reflectance and L is the luminance at point (x, y). Usually, the reflectance component R only depends on the intrinsic characteristics of the object (i.e., face) and L is the unexpected illumination factor that does not determine the visual appearance of the object. Therefore, the goal is to maintain the features of R and get rid of the effect of L. Figure 6.2 displays an example of attempting to separate the component R and L from a given face image. In most researches, the component L is commonly assumed to vary slowly in spatial level. This assumption is also verified by the human face model.

When it comes to the computer graphic application, a human face can be modeled as a combination of a series of small and flat facets [216, 217] such

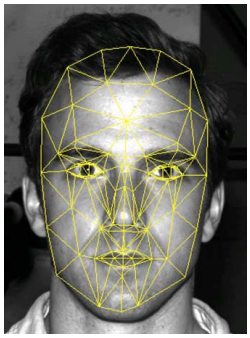


Figure 6.1: A face image and its corresponding CANDIDE-3 model.

as CANDIDE-3 face model [218] which is composed of a sequence of triangular facets. Figure 6.1 is an example. The area of each facet W is small enough to be considered as a planar patch and thus the illumination is approximately fixed within the facet. Then the reflectance model can be also expressed as:

$$I(x, y) = A \cdot R(x, y), \qquad (x, y) \in W_{N \times N}, \tag{6.2}$$

where A is a constant which denotes the illumination within squared small area W, and N is the dimension of the area.

6.3 Correspondence to refelctance model

Except for the illumination correction methods such as histogram equalization, many approaches[1] that cope with the illumination in face recognition are based on the reflectance model. Thus, it seems interesting to give an overview of various representative approaches complied with this model. Explanations of how they follow this model are also included.

6.3.1 Estimation of luminance (L)

As aforementioned, the goal of illumination processing in face recognition is to maintain the facial intrinsic properties represented by R and remove the illumination effect L. As seen from the reflectance model, the illumination problem can

[1] They are referred to as extraction of illumination insensitive features.

Figure 6.2: Original illuminated face image, illumination component L, and reflectance component R (from left to right).

be solved by first estimating L and then calculating the de-illuminated version I' of the input face image by

$$I'(x,y) = R(x,y) = \frac{I(x,y)}{L(x,y)}. \tag{6.3}$$

Sometimes the logarithmic form of the image is taken. In this case, I' can be obtained by first estimating the L or $\ln(L)$ and then computing

$$I'(x,y) = R(x,y) = \exp\left(f(x,y) - u(x,y)\right), \tag{6.4}$$

where $f = \ln(I)$, $u = \ln(L)$, or

$$I'(x,y) = v(x,y) = f(x,y) - u(x,y), \tag{6.5}$$

where $v = \ln(R)$.

Self quotient image (SQI). Based on the assumption that L varies little spatially, SQI [213, 172] employs a low pass filter G to estimate the low frequency component L by

$$L(x,y) = G(x,y) * I(x,y), \tag{6.6}$$

and use (6.3) for the final measure of I'.

Adaptive single scale retinex (ASR). ASR proposed by Park et al. [171] estimates L by iteratively convolving the input image with a 3×3 weighted mask as

$$L^{(t+1)}(x,y) = \max\{L'^{(t+1)}(x,y), L^{(t)}(x,y)\}, \quad 0 < t \leq T \tag{6.7}$$

with

$$L'^{(t+1)}(x,y) = \frac{1}{N(x,y)} \sum\sum_{(x,y)\in W_{3\times3}} L^{(t)}(x,y)w(x,y),$$

$$N(x,y) = \sum\sum_{(x,y)\in W_{3\times3}} w(x,y),$$

where $W_{3\times3}$ denotes the local square mask, $w(x,y)$ is the weight coefficient, and T is the total number of iterations. The coefficient is calculated via combining two measures of the illumination discontinuity at each pixel. This aims to preserve the shadow boundaries in L. Finally, the measure of I' is obtained through (6.5).

Logarithmic total variation (LTV). In [174], the illumination effect and facial features were distinguished as large scale features and small scale features, respectively. Thus, the authors first operated decomposition in the logarithmic image using TV-L^1 model [219] and used the decomposed large scale features as the estimation of $\ln(L)$:

$$u = \underset{u}{\operatorname{argmin}} \int |\nabla u| + \lambda \|f - u\|_{L^1}, \tag{6.8}$$

where $\int |\nabla u|$ is the total variation (TV) of u, λ is a scalar threshold on the image scale, $f = \ln(I)$, $u = \ln(L)$. At last, the measure of I' can be calculated by (6.4). In the final result, the large scale features are excluded.

Discrete cosine transform-based method (DCT). DCT [190] is another classic method that decomposes the input face image into multiple scales (i.e., frequencies). Also based on the assumption that L is of low frequency, the authors defined the first n low frequency components as L estimation. They subtracted these components from the discrete cosine transformed image of the logarithmic form of the input image. To sum up, first, represent the logarithmic image $f = \ln(I)$ by $M \times N$ two dimensional cosine transformation:

$$f(x,y) = \sum_{s=0}^{M-1} \sum_{t=0}^{N-1} D(s,t), \tag{6.9}$$

and then estimate the $\ln(L)$ by

$$u(x,y) = \sum_{i=1}^{n} D(s_i, t_i), \tag{6.10}$$

where

$$D(s,t) = \alpha(s)\alpha(t)C(s,t)\cos\left[\frac{\pi(2x+1)s}{2M}\right]\cos\left[\frac{\pi(2y+1)t}{2N}\right].$$

With the $f(x,y)$ and $u(x,y)$ described in (6.9) and (6.10), (6.5) is eventually used to achieve the measure of I'.

Wavelet denoising-based method (WD). WD is applied in [176] to generate a nonlinear estimation of $\ln(L)$ using the wavelet coefficients of the logarithmic image. Labeling the subbands of the wavelet transform as HH_k, HL_k, LH_k, and LL_k (k denotes the decomposition level), the first three represent the fine details of the input face image and the last one is the low resolution residual. By manipulating soft thresholding to the coefficients of HH_k, HL_k, LH_k, while keeping the coefficients of LL_k unaltered, u ($=\ln(L)$) is well estimated:

$$u(x,y) = IW(th_T(Wf_i), Wf_{LL_k}), \tag{6.11}$$

where $i \in \{HH_k, HL_k, LH_k\}$, Wf denotes the wavelet coefficient of logarithmic image $f = \ln(I)$ and its subscript denotes the corresponding subband, $th_T(\cdot)$ is the thresholding function which is executed when Wf is the coefficient of the detail subbands, and $IW(\cdot)$ is the inverse transform with the operated wavelet coefficients. Then measure of I' can be obtained via (6.5), where, LL_k components are removed.

6.3.2 Representation only related to reflectance (R)

It seems easy and understandable to estimate the illumination component L because of the presence of the common assumption that L spatially varies slowly. However, finding an estimation of L is only a middle step of the whole procedure and it is very difficult to accurately estimate L just based on the assumption,

especially for images under extreme illumination. It cannot be avoidable to incur errors in the estimation of L, as discussed in [177].

Actually, the approaches that result in face representation only related to R can be regarded as illumination insensitive. They omit the step of estimating L and directly compute the R-related representation of the input face image. In the following analysis, we will introduce several approaches of this kind and give the deduction about how they are only related to R.

Gradientface (GRF). In [177], it was argued that direct face representation only related to illumination insensitive component R is more robust than measuring R by first estimating L. Accordingly, they produced an illumination insensitive measure called gradient face as:

$$GRF = \arctan\left(I_y/I_x\right), \quad GRF \in [0, 2\pi), \tag{6.12}$$

where I_x and I_y are the differentiations of image I along the x and y direction, respectively.

Proof: Based on the reflectance model, consider any pixel at (x, y) and its neighbor at $(x + \Delta x, y)$:

$$I(x + \Delta x, y) = R(x + \Delta x, y)L(x + \Delta x, y). \tag{6.13}$$

Subtract (6.1) from (6.13) and refer to the assumption that L is approximately smooth:

$$\begin{aligned} & I(x + \Delta x, y) - I(x, y) \\ = \;& R(x + \Delta x, y)L(x + \Delta x, y) - R(x, y)L(x, y) \\ \approx \;& R(x + \Delta x, y)L(x, y) - R(x, y)L(x, y) \\ \approx \;& (R(x + \Delta x, y) - R(x, y))L(x, y). \end{aligned} \tag{6.14}$$

Transforming (6.14) into derivative form, we gain $I_x \approx L(x, y)R_x$ and similarly $I_y \approx L(x, y)R_y$. By dividing these two equations, one can obviously notice that

$$GRF = \arctan\left(I_y/I_x\right) \approx \arctan\left(R_y/R_x\right),$$

where GRF is only related to R.

Weberface (WF). In [178], based on the Weber's law, an illumination insensitive measure which is only related to R was also developed. The output representation named Weber face is

$$WF = \arctan \left(\alpha \sum_{\Delta x=-1}^{1} \sum_{\Delta y=-1}^{1} \frac{I(x,y) - I(x+\Delta x, y+\Delta y)}{I(x,y)} \right), \tag{6.15}$$

where α is a parameter for adjusting the contrast within the neighborhood.

Proof: According to the reflectance model, we have

$$\begin{aligned} & I(x+\Delta x, y+\Delta y) \\ = \; & R(x+\Delta x, y+\Delta y) L(x+\Delta x, y+\Delta y). \end{aligned} \tag{6.16}$$

Since L has slow spatial changes, it can be rewritten as

$$I(x+\Delta x, y+\Delta y) \approx R(x+\Delta x, y+\Delta y) L(x,y). \tag{6.17}$$

Replace each of $I(x,y)$ and $I(x+\Delta x, y+\Delta y)$ in (6.15) by (6.1) and (6.17):

$$WF \approx \arctan \left(\alpha \sum_{\Delta x=-1}^{1} \sum_{\Delta y=-1}^{1} \frac{R(x,y) - R(x+\Delta x, y+\Delta y)}{R(x,y)} \right),$$

which also depends on only R.

Local normalization (LoN). This approach was proposed in [197]. The underlying idea is that the finally processed image should be of local zero mean and with unit variance within a small area W. The illumination insensitive representation is given as

$$LoN = \frac{I(x,y) - E(I(x,y))}{Var(I(x,y))}, \qquad (x,y) \in W_{N \times N}, \tag{6.18}$$

where $E(\cdot)$ and $Var(\cdot)$ are the mean and variance operation within the local small area W, respectively. N denotes the dimension of the small area.

Proof: According to (6.2), (6.18) can be restated as

$$\begin{aligned} LoN &= \frac{A\cdot R(x,y) - E(A\cdot R(x,y))}{Var(A\cdot R(x,y))} \\ &= \frac{A\cdot R(x,y) - A\cdot E(R(x,y))}{A\cdot Var(R(x,y))} = \frac{R(x,y) - E(R(x,y))}{Var(R(x,y))}, \end{aligned}$$

from which we can also see that LoN is merely related to R.

6.3.3 Other processing

Instead of estimating the luminance L in the first stage or building the representation related to R directly, some other approaches aim to enhance the component R or suppress the effect of L in their own ways. They also share the assumption that L is probably of low frequency and R is relatively of high frequency.

Homomorphic filtering-based method (HF). HF was designed in [220] in order to reduce the contribution of L. It applies a high pass filter to the logarithmic image $f = \ln(I)$ for the purpose of excluding component L in the filtered output. The resultant representation can be expressed by

$$I'(x,y) = \exp\left(\Im^{-1}\{\Im\{f\}\cdot H\}\right), \tag{6.19}$$

where $\Im\{\cdot\}$ means Fourier transform and H denotes a Fourier transformed version of a high pass filter such as Butterworth filter [184] in [220].

Wavelet-based method (WA). WA [175] also relies on the wavelet transform and processes the coefficients in a specific way as WD does. It increases the coefficients corresponding to the image details (i.e., HH, HL, and LH subbands) and does histogram equalization with the LL subband. These operations imply the accentuation of R and suppression of L, which are also considered capable for illumination insensitive representation. The processing can be shown by

$$I'(x,y) = IW(\mu\cdot(WI_i), HE(WI_{LL})), \tag{6.20}$$

where $i \in \{HH, HL, LH\}$, WI denotes the wavelet coefficient and its subscript

denotes the corresponding subband, and μ is a scalar factor (>1) with which the multiplication is applied when WI is the coefficient of the detail subbands. $HE(\cdot)$ denotes the operation of histogram equalization and $\mathit{IW}(\cdot)$ denotes the inverse transform with the information of wavelet coefficients.

6.3.4 Summary

From the above discussion of various approaches, it is more clear that the key goal of the illumination processing in face recognition is to maintain the intrinsic-feature-dependant component R and to attenuate the illumination component L. It can be easily drawn that L is commonly associated to the following terms:

- Low resolution,
- Slow spatial changes,
- Low frequency,
- Large scale.

By contrast, R is associated to

- High resolution,
- Rapid spatial changes,
- High frequency,
- Small scale.

Under these considerations, R can be effectively approximated and L can be largely suppressed by means of filtering, denoising, decomposing, etc. These techniques can be further promoted via multi-scale analysis, addition of variable weights, or integration with other approaches to make it more sophisticated. This is a potential direction for future research.

One more notable thing is that the idea of local operation is shared by many approaches. Besides ASR, GRF, WF, and LoN mentioned above, the famous LBP [27, 221], its improved versions [222, 26], the recently proposed LRM [223], NLM [2], and OLHE [196] also adopt the local operation. They all describe the relations between neighboring pixels for illumination insensitive representation. There may be two explanations for the wide usage of the local operation: a) illumination has different effects in different regions, and b) component L varies

slowly within local small region. Therefore, one can take advantage of the local operation which tends to be robust to the change of L and achieve reliable face representation.

6.4 Considerations for illumination processing

In this section, apart from the reflectance model, we analyze some other aspects from the viewpoints of contribution of fundamental illumination correction, processing domain, feature scale, and a common problem. By these analyses, we hope to show constructive suggestions.

6.4.1 Fundamental illumination correction

This subsection presents four widely used illumination correction methods. They were proposed for denoising or image enhancement. They can correct the illumination to some extent and are taken as one part or preprocessing step in many approaches. In general, they are basic processings commonly shared by other approaches.

Logarithm transform (LT). LT is often used in image enhancement to expand the values of dark pixels while compressing the values of bright pixels [224]. In this sense, LT itself can partially reduce the illumination effect. It can be deemed as a preprocessing step, which would facilitate the subsequent illumination normalization. Under this transformation, the reflectance model can be rewritten as

$$\ln(I(x,y)) = \ln(R(x,y)) + \ln(L(x,y)). \tag{6.21}$$

It changes the multiplicative form of the reflectance model to additive one, which is likely to remove noise from the image and make it possible to promote useful signals.

We have already given examples of approaches using this transformation in Sect. 6.2. Besides, other approaches making use of LT can be referred to the recently proposed methods using neighboring wavelet coefficients [225], nonsubsampled contourlet transform [192], total variation for preprocessed decomposi-

tion [226, 227], etc. As seen from (6.5), the final representation sometimes skips the inverse transformation from the logarithmic form, which actually justifies the significance of LT.

Histogram equalization (HE). HE [166] maps the histogram of an input image to a new uniform distribution, so that the intensity probabilities are equal and the image is fairly enhanced. Consider an input image $I(x, y)$ with n pixels and a total number of L grey levels. The probability of the occurrence of one pixel with r_q grey level is denoted by $p_r(r_q) = n_q/n$, where n_q $(q = 0, 1, \cdots, L-1)$ is the number of pixels with r_q grey level. HE transforms a given intensity value r_q to a new one r'_q by

$$r'_q = \sum_{i=0}^{q} \frac{n_i}{n} = \sum_{i=0}^{q} p_r(r_i) = \mathrm{CDF}(r_q), \tag{6.22}$$

where $\mathrm{CDF}(\cdot)$ represents the cumulative distribution function. This equation redistributes the intensity to the domain of $[0, 1]$. The target values r'_q need to be rescaled to obtain pixel values in the original domain, e.g., $[0, 255]$.

During the implementation of WA [175], HE is applied. The reason is that only using most of the available dynamic range of the original image, the wavelet coefficients can possibly use that range sufficiently. LoN [197] combines HE with the proposed method to get better result than only using their method. OLHE [196] also takes HE as a useful instrument by applying oriented HE locally.

Gamma correction (GC). GC nonlinearly transforms the original gray-level image I to I^γ, where $\gamma \in [0, 1]$ is a user-defined parameter. By setting different values of γ, it can accordingly enlarge the local dynamic intensity range in dark regions while compressing that in bright regions, which meets the requirement of human visual conception under common illumination conditions (not pitch black nor blindingly bright) [167]. This capability in illumination balance of GC allows for its application to illumination processing in face recognition. It is applied in Tan and Triggs' method (TT) [191] as one of the three steps of the proposed pipeline. In [228, 229], GC was used for illumination correction by solving an optimization problem.

Gaussian filtering (GF). Gaussian kernel function with standard deviation σ is defined as

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right). \tag{6.23}$$

It is also frequently used as a preprocessing technique for the purpose of excluding noises and specific illumination from an input image. Especially, the techniques which extract gradient-related features prefer GF for computational ease with derivative of Gaussian kernel function such as GRF [177]. In [191], Difference of Gaussian (DoG) was used to discard high spatial frequencies that were present in the given face image such as shading effects. A 2D Gaussian illumination model was also used in [230] to compensate the illumination variances caused by shadows.

6.4.2 Processing domain

It is important to articulate how the face recognition is affected by taking into account the processing domain. Some classic face recognition algorithms are only implemented within the spatial domain, which ignore the information concealed in the relationship between neighboring pixels. More sophisticated facial feature representation methods considering frequency information had been proposed later such as Garbor feature based method [231] and Wavelet feature based method [232]. The features extracted from the frequency domain seem more discriminant than those from the spatial domain.

It can be found that some illumination processing approaches explicitly transform an original image into frequency domain (e.g., DCT, WD, WA, and NWC [225]). Thus we also suppose that frequency domain provides more useful information which can aid to distinguish between facial features and illumination. This is easy to understand if one recalls the reflectance model and summary in Sect. 6.3.4. Others utilize the local operation to obtain neighborhood relations, which also contain frequency information of faces. By contrast, early work such as LT or HE handles the illumination problem with the original image. It spatially normalizes the whole given image without distinguishing between the illumination and facial features. Hence, methods of this kind may result in unreliable performance especially in uneven illumination.

6.4.3 Feature scale

We have mentioned that some terms can be usually associated to the intrinsic facial feature component R and illumination component L. The small scale (high resolution or high frequency) features refer to R, and the large scale (low resolution or low frequency) features refer to L. Therefore the concept of feature scale is important for extracting R from illuminated face images $I(= R \cdot L)$.

To our observation, the contribution and essentiality of large scale features and small scale features had been rarely evaluated and analyzed before [226, 233]. In [226], Xie et al. claimed that, for a face image, there is no doubt that the small intrinsic facial details are small scale features, but the large scale features not only contain the illumination but also the large intrinsic facial features. However, many existing approaches either correct the illumination directly (e.g., LT, HE) or regard the whole large scale features as L and discard them completely (e.g., LTV, DCT). As a result, they neglect the useful large intrinsic facial features. In order to utilize both of the small and large intrinsic features of a face, the method developed in [226] first decomposes the given face image into two images of small and large scale and then executes smoothing on the small scale image and normalizing on the large scale image, respectively. Final reconstruction is accomplished by fusing the smoothed small scale image and normalized large scale image. It makes use of the normalized large scale image rather than completely discarding it. The procedure can be illustrated in Fig. 6.3. It was proved that this method significantly outperformed LTV which also adopts the decomposition but only reserves the small scale image while discarding the whole large scale image.

More recently, in [227], Matsukawa et al. also kept the large intrinsic features by normalizing the large scale features rather than discarding them. Only the normalization method is different from that in [226]. They experimentally confirmed that their method could lead to better visual quality and more robust performance under varying illumination compared with other approaches.

In fact, previously analyzed WA also complies with the rule of keeping the large intrinsic features. But it just does HE on the large scale features (LL subband) and enlarges the coefficient of small scale features (HH, HL, LH subbands). The large scale features with HE are wholly kept. In other words, it does not

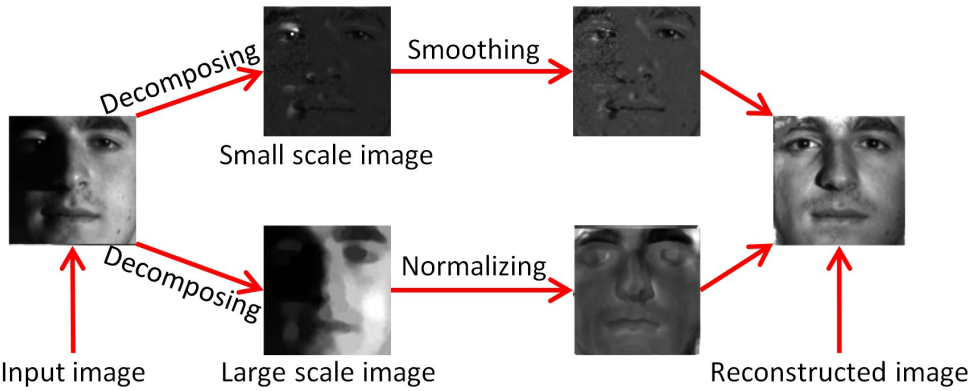


Figure 6.3: Illustration of illumination processing based on large and small scale features.

manage to erase the illumination enough. This early work is relatively not so comparable.

In summary, to create a more discriminating feature representation, one can consider along this direction of maintaining small scale intrinsic features together with large scale intrinsic ones. Nevertheless, to our knowledge, it is not easy to distinguish which is useful (intrinsic features) or not (illumination) in the large scale features. It is a familiar phenomenon that illumination and facial features tend to be removed or maintained simultaneously. This is what we want to talk about in the following analysis.

6.4.4 Common problem

When surveying the majority of existing illumination processing approaches, some specific problems remain difficult to overcome.

One obvious problem is the negative effect of shadow boundaries. Figure 6.4 shows several face images of the same subject under varying illumination and their corresponding face representations using different illumination processing methods. The red blocks demonstrate the shadow boundaries. We can see that these methods are able to remove most effects caused by the illumination even within shadows. However, the artifacts of shadow boundaries are still extrusive in the processed images.

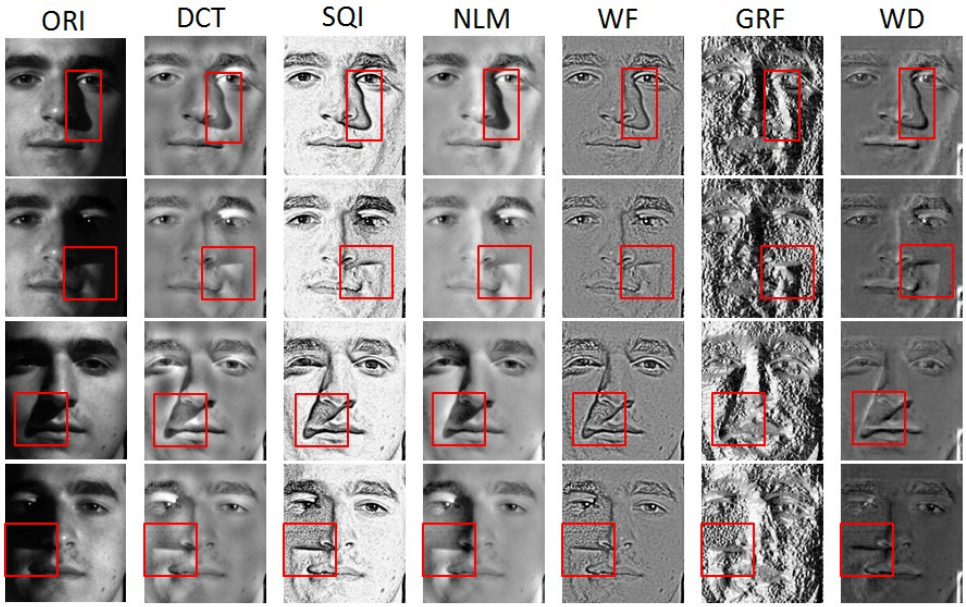


Figure 6.4: Original face images under varying illumination (1st column) and images processed by DCT, SQI, NLM [2], WF, GRF, and WD, respectively (from the 2nd to the last column).

The reason is that the luminance component L is assumed to vary very slowly, which contradicts with the fact that shadow boundaries are of fast spatial changes. This negative effect could be very strong to impair the recognition effectiveness. In the literature, we found two famous approaches concerning about the processing of shadow boundaries. They are ASR and LTV. As introduced in Section 6.2, ASR [171] takes the illumination discontinuity into account and measures them during each iterated smoothing. LTV [174] uses TV-L^1 model for decomposition, which is able to preserve the strong edges in the illumination component. Figure 6.5 displays the face representations by ASR and LTV. It can be noticed that the shadow boundaries are largely suppressed by these two approaches.

Despite their ability to counter the influence of shadow boundaries, they tend to lose more facial information at the same time, which is also notable in Fig. 6.5. This is one case of the more general problem which can be stated as: simultaneous increase or decrease of the facial features and the illumination. This so called "synchronization" is usually controlled by specific parameters. In LTV,

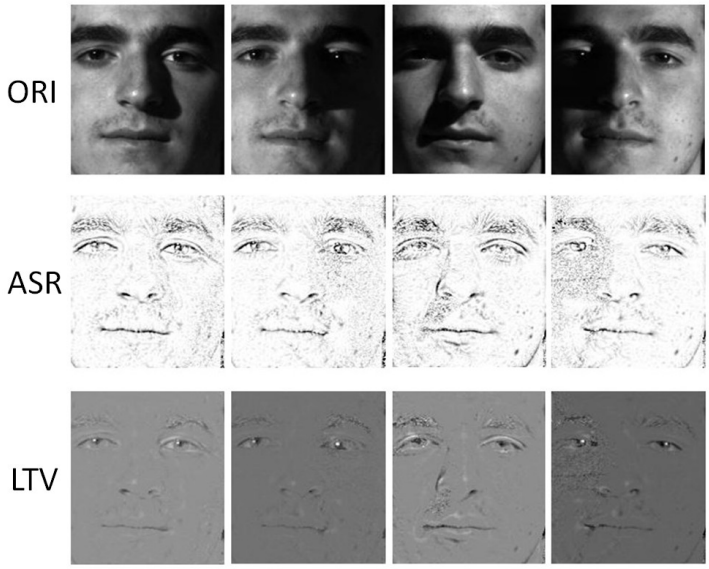


Figure 6.5: Original face images (1st row) and their corresponding processed images using ASR (2nd row) and LTV (3rd row).

the parameter λ in (6.8) is used to adjust the decomposition scale; when λ becomes large, illumination and facial features will be removed to a large extent simultaneously. In ASR, the maximal iteration number for termination denoted by T needs to be set in (6.7); if T is too small, facial features will be hardly distinguished although shadows will be removed effectively. In LN, the block size N in (6.18) is proportional to the extent of both facial feature maintenance and illumination redundancy. In WF, the parameter α in (6.15) controls the contrast in the neighborhood, affecting facial features and illumination at the same time. Mostly, compromised parameter values are selected based on the recognition performance with little theoretical analysis. This seems an open problem which is worthy of further study.

6.5 What to normalize

In some researches, it is called "illumination normalization" to cover the whole passive approaches of illumination processing. However, this term may bring about confusion in this paper since we would like to distinguish between objects

to be normalized. Some of the passive approaches aim to normalize the illumination of given images and others try to normalize the intra-class difference between images of one subject. In the following, we prefer to denote these two normalizations as "illumination normalization" and "intra-class difference normalization", separately. It seems constructive to discuss this point because different viewpoints will lead to completely different research directions.

6.5.1 Illumination normalization

Almost all of the existing approaches have tried to compensate the illumination in each of given images separately. They fundamentally correct the illumination or represent the faces with extracted illumination insensitive features, by focusing on one target image. They need to analyze the property of illumination to discriminate the illumination from faces and finally construct a new face (maybe not a face, e.g., GRF) for each given image. These methods have been exploited extensively for recognition applications. However, illumination and faces share much similarity and they still remain unsolved to be thoroughly separated. For instance, the shadow boundaries are difficult to remove, even with the commonly adopted assumption. One more negative example is the illumination normalization using HE. HE copes with the illumination in different images separately and suppresses the illumination largely with regard to each given image. However, since input images are illuminated in different ways, the resultant images will be far different from each other even they are the same subject. Thus, images processed by HE may not be suitable for the subsequent recognition. Figure 6.6 shows some visual examples of illumination normalization by HE.

Consider that the ultimate goal is to accomplish face recognition (classification) under varying illumination with high accuracy. One can figure out that producing more similar cues for the same subject seems more important than trying to get rid of the illumination. The illumination only needs to be suppressed under the condition of not disturbing the classification. Actually, GRF is not fully removing illumination. It sacrifices the visual quality for closer representations of same subjects and obtains relatively better recognition performance.

Figure 6.6: Input face images of one subject under varying illumination (1st row) and their corresponding results using HE (2nd row).

6.5.2 Intra-class difference normalization

As mentioned above, it is difficult to result in uniform face images, and thus "illumination processing in face recognition" can be considered in another way. That is not to normalize the illumination in each face image, but to obtain approximately identical face representation for same subjects. This can be thought as intra-class difference normalization.

One example embracing the idea of normalizing the intra-class difference is *maximizing a correlation* (MAC) [229]. It attempts to maximize the intra-individual correlations rather than separately processing the illumination for each face image. With a gallery face image I_G and a probe I_P, it represents the optimized results by

$$\{I'_G, I'_P\} = \{f(I_G, \xi^*_G), f(I_P, \xi^*_P)\}, \tag{6.24}$$

where $f(\cdot)$ is a transform function for illumination processing. ξ^*_G and ξ^*_P are parameters of f, which are estimated by solving the optimization problem of

$$\{\xi^*_G, \xi^*_P\} = \underset{\xi_G, \xi_P}{\text{argmax}}\ corr\{f(I_G, \xi_G), f(I_P, \xi_P)\}, \tag{6.25}$$

where $corr\{\cdot\}$ calculates the correlation between two face images. MAC processes images pair by pair between galleries and probes. It adjusts the illumination rather than normalizing it and aims to generate the illumination as similar as possible for the same subject. In other words, it normalizes the difference within

the same class. MAC actually provides a nice scheme which needs to incorporate a proper existing or new illumination processing method as denoted by $f(\cdot)$ in (6.24). In [229], GC and LTV were tried in this scheme. Other methods can be also properly implemented in this scheme in the future work.

The intra-class difference normalization is likely to be more suitable for the face recognition task than the separate illumination normalization. The reason is that the former one can assist the final classification by diminishing the intra-class difference while prolonging the inter-class distance. For example, MAC helped achieve much better performance than only using an illumination processing method which focuses on each single input image in [229]. In this sense, a more potential way to inherit this idea may be a joint normalization of intra-class difference, which processes the illumination in all face images jointly.

6.6 Experiments

In this section, we do summative evaluations from two perspectives. One is the contribution of fundamental illumination correction methods to illumination insensitive face recognition. The other is the comparative performance of various illumination processing methods. Two databases were used for each assessment.

The first database, cropped Extended Yale B [168], consists of 38 subjects with 9 poses and 64 illumination conditions. We choose the frontal faces in our experiments to focus only on the illumination problem and there are totally 2432 face images. The size of each image is 192×168. We divide this database into five subsets according to the lighting angles. The five subsets are: subset 1 ($0° \sim 12°$, 266 images), subset 2 ($13° \sim 25°$, 456 images), subset 3 ($26° \sim 50°$, 456 images), subset 4 ($51° \sim 77°$, 532 images), and subset 5 (above $78°$, 722 images). On this database, two experiments are conducted. Experiment 1 uses relatively more gallery images, that is, all the images from subset 1 are used as the gallery images (266 images), and the other images from subset 2 to 5 as the probe images. Experiment 2 uses fewer gallery images, only 1 image for each subject from subset 1 as the galleries (38 images), the rest images in subset 1 together with subsets from 2 to 5 as the probes. These two situations test the effectiveness of various methods with different amounts of reference information

during recognition. Figure 6.7 displays 5 sample images of one subject from different subsets from Extended Yale B database.

The second database is CMU-PIE database [199], which contains 68 subjects under large variations in illumination, pose and expression, totally 41368 images. The illumination subset including 68 subjects under 21 different illumination directions (21 images per subject) is used in our experiments. All the images are properly cropped to the size of 161×161 with only the face region. To get fair evaluation, one image per subject (totally 68 images) is chosen as the galleries each time (totally 21 times) and the others are used as the probes. The results are averaged finally for each method. Figure 6.8 shows 21 face images with various illumination conditions of one subject from CMU-PIE database.

With regard to the recognition protocol, we use the raw pixel values of the resultant face representations of different methods to do template matching and evaluate each method by using the one nearest neighbor rule with three distance measures (L_1 norm, L_2 norm, and χ^2 distance measure).

Figure 6.7: Face images of the same subject from 5 subsets.

Figure 6.8: Face images of the same subject with different illumination conditions.

Since different methods achieved the highest recognition rate using different distance measures, we choose the best result among the three distance measures for each method as its result for comparison.

6.6.1 Contribution of fundamental illumination correction

We first explore the contribution of fundamental illumination correction methods to the illumination insensitive face recognition. For this purpose, we select four methods DCT [190], WA [175], TT [191] and GRF [177], each of which were originally developed by using one corresponding fundamental illumination correction method (i.e., LT, HE, GC, and GF, respectively). We test these methods before and after removing the corresponding fundamental illumination corrections, that is to compare the performance of DCT vs DCT without LT, WA vs WA without HE, TT vs TT without GC, and GRF vs GRF without GF.

Tables 6.1 and 6.2 show the comparative results of the two experiments on Extended Yale B database using different numbers of gallery images, respectively. The distance measure used by each method to achieve the best performance is denoted as well. The face recognition rates achieved by using only the fundamental illumination correction and the raw image with no processing are also presented. From these tables, it is obvious that all the methods with corresponding fundamental illumination correction techniques perform better than those without them. In the first experiment, DCT, WA, TT, and GRF outperform their counterparts without the corresponding fundamental illumination corrections by 1.23%, 1.78%, 5.19%, and 1.47%, respectively. Similarly, in the second experiment, the corresponding performance differences reach 0.88%, 1.46%, 9.90%, and 2.50%, respectively. Table 6.3 gives the average recognition rates of different methods on CMU-PIE database, also demonstrating the usefulness of fundamental illumination corrections. These results confirm that the fundamental illumination corrections are important for illumination processing in face recognition.

6.6.2 Comparative performance of various approaches

In this part, we present recognition results of comparative study on eight illumination insensitive approaches including DCT [190], WA [175], SQI [172], ASR

Table 6.1: Recognition rates (%) on Extended Yale B face database with one image per subject in subset 1 as the galleries.

Method		Subset2	Subset3	Subset4	Subset5	Avg.
ORI	L_2	90.13	41.89	5.45	2.63	30.01
LT	χ^2	96.05	74.34	51.45	7.62	51.05
DCT_no	χ^2	98.90	94.85	70.50	4.85	59.72
DCT	χ^2	94.74	74.01	58.59	33.10	60.95
HE	χ^2	99.12	91.01	65.03	34.90	67.63
WA_no	L_2	92.11	72.04	47.37	2.49	47.02
WA	χ^2	88.82	71.05	48.13	9.97	48.80
GC	χ^2	96.05	76.54	53.39	9.97	52.77
TT_no	L_1	99.56	99.67	97.58	83.10	93.61
TT	L_1	99.56	99.78	98.50	97.92	98.80
GF	L_2	96.93	74.56	49.17	3.74	49.43
GRF_no	L_1	98.90	99.34	90.98	51.52	81.26
GRF	L_1	99.34	99.78	85.90	59.14	82.73

[171], TT [191], GRF [177], WF [178], and LTV [174]. Two experiments were devised. The first one used all the images from subset 1 as the gallery images (70 images), the other images from subset 2 to 5 as probe images. The second one used 1 image for each subject from subset 1 as the galleries (10 images), the rest images in subset 1 together with subsets from 2 to 5 as the probes. We also selected the best result among the three distance measures (L_1 norm, L_2 norm, χ^2 distance measure) for each approach as its result for comparison. Figure 6.8 shows the average comparative results of these two experiments. We can see that the highest two recognition rates are achieved by TT and ASR in both of these two experiments. WF ranks the third, also accomplishing recognition rates more than 90% in both experiments. In addition, it can be found that only DCT get better performance in experiment 2 than that in experiment 1. Unfortunately, SQI drops the most from experiment 1 to 2 with the decrease of almost 35%. This unstable result demonstrates that SQI is less reliable for illumination insensitive face recognition with small number of gallery images.

In this part, we present recognition results of comparative study on nine illumination insensitive approaches including DCT [190], WA [175], SQI [172], ASR

Table 6.2: Recognition rates (%) on Extended Yale B face database with one image per subject in subset 1 as the galleries.

Method		Subset1	Subset2	Subset3	Subset4	Subset5	Avg.
ORI	χ^2	95.18	91.45	21.49	4.70	2.77	32.46
LT	χ^2	96.49	96.05	36.84	9.77	5.68	38.39
DCT_no	χ^2	90.79	93.20	57.46	22.93	4.02	43.65
DCT	χ^2	86.84	75.66	45.83	31.20	20.50	44.53
HE	χ^2	98.68	94.30	44.74	17.11	28.53	48.29
WA_no	χ^2	86.40	87.06	33.33	5.26	3.05	33.25
WA	χ^2	81.58	84.43	35.31	10.34	6.09	34.71
GC	χ^2	96.49	95.61	36.84	11.28	6.09	38.76
TT_no	L_1	82.89	100.00	82.24	82.52	61.50	79.49
TT	L_1	84.65	100.00	83.55	90.98	86.70	89.39
GF	χ^2	96.93	95.39	28.95	5.08	3.19	35.01
GRF_no	L_1	85.09	98.68	82.46	64.10	26.32	64.79
GRF	L_1	85.96	99.34	84.43	67.67	30.06	67.29

[171], TT [191], GRF [177], WF [178], LTV [174], and NLM [2]. Figure 6.9 shows the average comparative results of the two experiments on Extended Yale B database. We can see that the highest three recognition rates are achieved by TT, ASR, and WF in both of these two experiments. NLM ranks the third and GRF ranks the fourth. It should be pointed out that a high recognition rate is often achieved by using a specific distance measure according to different methods. For example, TT using L_1 norm yields better result than ASR using L_1 norm; but TT using L_2 norm performs worse than ASR using L_2 norm. The distance measure used by each method to achieve the best performance is denoted. In addition, it can be found that SQI drops the most from experiment 1 to 2. This unstable result demonstrates that SQI is less reliable for illumination insensitive face recognition with small number of gallery images. This is because SQI does not consider about tackling shadow boundaries and thus its discriminative ability between facial features and illumination effects is not satisfactory enough and is sensitive to reference information. When less reference information is given, the illumination tends to be confused with the inter-class difference and thus more false recognition will be yielded. Figure 6.10 gives the average recognition rates

Table 6.3: Recongition rates (%) on CMU-PIE database.

Method		Avg.
ORI	L_2	28.47
LT	χ^2	34.48
DCT_no	χ^2	40.70
DCT	χ^2	44.50
HE	χ^2	54.77
WA_no	L_2	31.04
WA	χ^2	35.59
GC	χ^2	34.39
TT_no	L_1	88.32
TT	L_1	98.39
GF	L_2	38.47
GRF_no	L_1	84.17
GRF	L_1	85.18

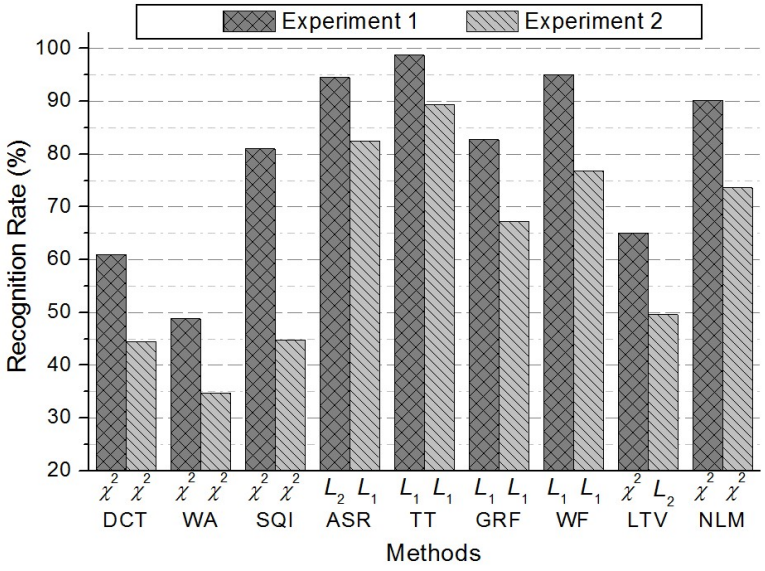


Figure 6.9: Comparative results of various methods on Extended Yale B database.

of different methods on CMU-PIE database. It shows similar rank of the tested methods, in which TT, ASR, and WF generate the best three recognition results using L_1, L_2, and L_1 norm, respectively.

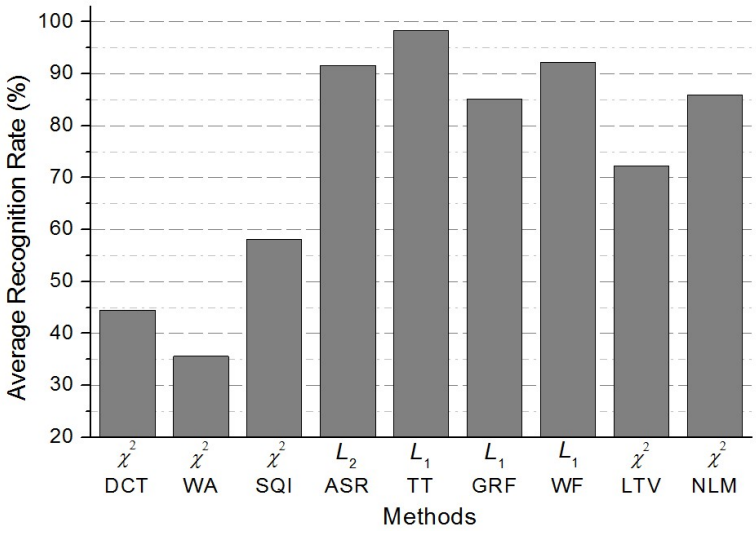


Figure 6.10: Comparative results of various methods on CMU-PIE database.

6.7 Summary

In this study, some noteworthy issues in illumination processing for face recognition have been analyzed. Focusing on the special topic of face illumination processing, we explained the correspondence of several well-known approaches to the reflectance model. We then discussed various considerations including contribution of fundamental illumination correction methods, processing domain, feature scale, and a common problem. Furthermore, we discussed an essential question of what to actually normalize. From these explanations and discussions, potential directions for future work were also analyzed. Finally, the experimental results on the contribution of fundamental illumination correction to the illumination insensitive face recognition and comparative performance of various methods were given as well. It has been shown that the fundamental illumination correction is often important in increasing the recognition performance and Tan and Triggs' method (TT) using L_1 norm achieved the highest recognition rates among nine tested methods.

Chapter 7

Conclusion

7.1 Observation

Face detection and recognition are important research topics nowadays and have various applications in the real world. Among the factors that make the detection & recognition tasks difficult to be accurate or efficient, severe/uneven illumination conditions is one of the key problems. This dissertation deeply investigated the illumination processing in face detection and recognition. In this correspondence, by analyzing related works in the literature, we presented some novel proposals which have specific advantages themselves.

In Chapter 3, we studied how much the prepended normalization step could affect the face detection results by combining the normalization and two famous learning-based face detectors. We applied a number of normalization techniques under this framework. They were initially proposed for different purposes rather than face detection. We also proposed a novel image normalization method SH for the purpose of removing the negative effects from non-uniform illumination for face detection. We observed that several used normalization techniques largely improved the correct detection of the original face detectors, which demonstrated the effectiveness of using specific normalization methods. Our method yielded the best performance in terms of F-measure, which verified its advantages in suppressing the illumination and preserving intrinsic facial features. We also reached an interesting conclusion that useful normalization techniques for face recognition or image enhancement are not necessarily effective for face detection.

7. Conclusion

In Chapter 4, in order to address the limitations of a recently proposed face representation method Weberface (WF) for illumination insensitive face recognition, we proposed to improve it in two ways. One was to incorporate the oriented information into the final face representation by concatenating eight directional face images rather than simply summing them together. Thus we named this method oriented Weberface (OWF). These images were computed based on the Weber's law, similar to the conventional WF. Another improvement was achieved by characterizing the facial features in larger granularities and was called largely-scaled Weberfaces (LSWFs). This was based on the fact that the local changes may not mean those in the nearest neighborhood. Both of these improvements could result in representations related to only the reflectance component R according to the Lambertian reflectance model, which theoretically proved their insensitivity to varying illumination. We compared the recognition performance using our methods with several other well-known methods. The benchmarked results illustrated that our methods significantly outperformed the others.

In Chapter 5, we proposed another illumination insensitive face representation method using multifractal analysis, named Mulrifractal-Face (MFF). Multifractal analysis is a popular tool to derive feature descriptors for a variety of image processing purposes. The $f(\alpha)$ features were computed from the given image as the representation for the subsequent face recognition. We analyzed its discrimination capacity and insensitivity to illumination variances and its core merit is the ability of coding the local and global image regularity. The usage of this ability was inspired from the illumination property. The experimental results showed that MFF was more powerful in illumination insensitive and discriminative face classification than several other state-of-the-art methods.

Apart from the above particular methods proposed for tackling the illumination problem, we also did a survey on the illumination processing in face recognition. We found that most of the existing surveys of this topic lack the emphasis on important aspects which are noteworthy for the upcoming researchers to construct new algorithms. To fill this gap, we summarized and discussed some considerations from several perspectives. Studying along these considerations may assist to improve the face recognition performance under varying illumination.

Our survey played just such an enlightening role. The summative evaluations through some experiments were given as well.

7.2 Limitation

Although we argued that a prior normalization step may be fairly helpful in raising the face detection results, one needs to carefully select the normalization technique. With regard to the proposed method SH, it may sacrifice some details of the image for more local enhancement and this led to false negatives in some cases. The parameters in SH could not be fixed for different situations.

In OWF, the oriented information was gained by concatenating eight face images, which means that the size of each resultant face image is eight times larger than that of an ordinary method. Its high computational cost and memory cost are big problems.

MFF is also not a perfect method. Actually, as mentioned in our survey in Chapter 6, there is some simultaneous increase or decrease of the facial features and the illumination, relatively speaking. MFF just achieved satisfactory balance between the facial features maintenance and illumination redundancy. But it is not as insensitive to illumination as ASR. For face recognition under rather difficult lighting conditions, MFF may not be so effective.

7.3 Further work

First, considering the limitations of our proposed methods, we may find the corresponding way to address each problem in the future. With SH, we shall study on better trade-off between the local contrast enhancement/illumination removement and the loss of useful information.

With OWF, in order to make it more efficient, PCA, LDA or other principle features extraction methods shall be used to reduce the feature dimension. On the other hand, OWF and LSWFs possess the oriented and scaled information, separately. We intend to investigate more advanced face representation method incorporating both of their merits, aiming at preserving more discriminant facial features.

With MFF, we may find some way to combine it with other approaches to make it more insensitive to large illumination variances. Applying primitive image preprocessing is one option since MFF is likely to have the irregular illumination (e.g., shadow boundaries) left as much as the intrinsic facial patterns. Another option may be to combine MFF with other representation approach which is more insensitive to illumination but less discriminative such as ASR.

Besides the improvement of our methods, we would also like to settle some theoretical issues or explore novel methods which may be more proper to handle the illumination problem. Actually, our survey in this dissertation has stated several directions of further work on illumination processing in face recognition. We intend to study on statistical measures of selecting optimal parameters for different approaches. Also, we intend to apply joint alignment to all the given images for intra-class normalization. As we know, "congealing" [234] might be possible to be used in this research.

Bibliography

[1] R. Brunelli and T. Poggio. Face recognition: features versus templates. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15. ix, 24, 25

[2] V. Struc and N. Pavesi. Illumination invariant face recognition by non-local smoothing. *BIOID Multicomm September*, 2009. xi, 94, 100, 108

[3] C. Zhang and Z. Zhang. A survey of recent advances in face detection. *Technical Report, Microsoft Research*, 2010. xiii, 1, 14, 19, 20, 21, 31

[4] Y. Lu, J. Zhou, and S. Yu. A survey of face detection, extraction and recognition. *Computing and Informatics*, 22, 2003. 1

[5] M. H. Yang, D. J. Kriegman, and N. Ahuja. Detecting faces in images: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(1):34–58, 2002. 1, 11, 12, 13, 14

[6] R. Jafri and H. R. Arabnia. A survey of face recognition techniques. *J. Information Processing Systems*, 5(2):41–68, June 2009. 2

[7] A. K. Jain, R. Bolle, and S. Pankanti. Biometrics: Personal identification in networked security. *A. K. Jain, R. Bolle, and S. Pankanti, Eds.: Kluwer Academic Publishers*, 1999. 2

[8] K. Kim. Intelligent immigration control system by using passport recognition and face verification. *Int. Symp. Neural Networks*, pages 147–156, 2005. 3

[9] J. N. K. Liu, M. Wang, and B. Feng. ibotguard: an internet-based intelligent robot security system using invariant face recognition against intruder. *IEEE Trans. Systems Man And Cybernetics Part C-Applications And Reviews*, 35(2):97–105, 2005. 3

[10] H. Moon. Biometrics person authentication using projection-based face recognition system in verification scenario. *Int. Conf. Bioinformatics and its Applications*, pages 207–213, 2004. 3

[11] D. McCullagh. Call it super bowl face scan 1. *Wired Magazine*, 2001. 3

[12] CNN. Education school face scanner to search for sex offenders. *Phoenix, Arizona: The Associated Press*, 2003. 3

[13] P. J. Phillips, H. Moon, P. J. Rauss, and S. A. Rizvi. The feret evaluation methodology for face recognition algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22:1090–1104, 2000. 3

[14] T. Choudhry, B. Clarkson, T. Jebara, and A. Pentland. Multimodal person recognition using unconstrained audio and video. *Int. Conf. Audio and Video-Based Person Authentication*, pages 176–181, 1999. 3

[15] S. L. Wijaya, M. Savvides, and B. V. K. V. Kumar. Illumination-tolerant face verification of low-bitrate jpeg2000 wavelet images with advanced correlation filters for handheld devices. *Applied Optics*, 44:655–665, 2005. 3

[16] E. Acosta, L. Torres, A. Albiol, and E. J. Delp. An automatic face detection and recognition system for video indexing applications. *IEEE Int. Conf. Acoustics, Speech and Signal Processing*, 4:3644–3647, 2002. 3

[17] J. H. Lee and W. Y. Kim. Video summarization and retrieval system using face recognition and mpeg-7 descriptors. *Image and Video Retrieval, Lecture Notes in Computer Science*, 3115:179–188, Springer Berlin Heidelberg, 2004. 3

[18] C. G. Tredoux, Y. Rosenthal, L. d. Costa, and D. Nunez. Face reconstruction using a configural, eigenface-based composite system. *3rd Biennial Meeting of the Society for Applied Research in Memory and Cognition (SARMAC)*, 1999. 3

[19] K. Balci and V. Atalay. Pca for gender estimation: Which eigenvectors contribute. *16th Int. Conf. Pattern Recognition*, 3:363–366, 2002. 4

[20] B. Moghaddam and M. H. Yang. Learning gender with support faces. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24:707–711, 2002. 4

[21] Y. Shinohara and N. Otsu. Facial expression recognition using fisher weight maps. *IEEE Int. Conf. Automatic Face and Gesture Recognition*, 100:499–504, 2004. 4

[22] F. Bourel, C. C. Chibelushi, and A. A. Low. Robust facial feature tracking. *British Machine Vision Conference*, pages 232–241, 2000. 4

[23] H. Schneiderman and T. Kanade. A statistical method for 3d object detection applied to faces and cars. *IEEE Conf. Computer Vision and Pattern Recognition*, 2000. 4

[24] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proc. the IEEE*, 94(11):1948–1962, Nov. 2006. 4

[25] B. Froba and A. Ernst. Face detection with the modified census transform. *Proc. IEEE Intl. Conf. Automatic Face and Gesture Recognition*, pages 91–96, 2004. 5, 15, 19, 23, 29, 31

[26] H. Jin, Q. Liu, H. Lu, and X. Tong. Face detection using improved lbp under bayesian framework. *Proc. Intl. Conf. Image and Graphics*, pages 306–309, 2004. 15, 19, 23, 29, 94

[27] L. Zhang, R. Chu, S. Xiang, S. Liao, and S.Z. Li. Face detection based on multi-block lbp representation. *Proc. Intl. Conf. Biometrics*, 2007. 15, 19, 23, 31, 94

[28] A. Roy and S. Marcel. Haar local binary pattern feature for fast illumination invariant face detection. *British Machine Vision Conference*, 2009. 5, 23, 29

[29] H.F. Chen, P.N. Belhumeur, and D.W. Jacobs. In research of illumination invariants. *Proc. CVPR*, pages 254–261, 2000. 5

[30] P. J. Phillips, H. Wechsler, J.Huang, and P. J. Rauss. The feret database and evaluation procedure for face-recognition algorithm. *Image and Vision Computing*, 16:295–306, 1998. 6

[31] D. Blackburn, J. Bone, and P. J. Phillips. Face recognition vendor test 2000. Defense Advanced Research Projects Agency, Arlington, VA, Technical report A269514, February 16, 2001. 6

[32] P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and J. M. Bone. Face recognition vendor test (frvt 2002). National Institute of Standards and Technology, Evaluation report IR 6965, March, 2003. 6

[33] K. Messer, J. Kittler, M. Sadeghi, M. Hamouz, A. Kostin, F. Cardinaux, S. Marcel, S. Bengio, C. Sanderson, J. Czyz, L. Vandendorpe, C. McCool, S. Lowther, S. Sridharan, V. Chandran, R. P. Palacios, E. Vidal, L. Bai, L. Shen, Y. Wang, Y. H. Chiang, H. C. Liu, Y. P. Hung, A. Heinrichs, M. Mller, A. Tewes, C. V. D. Malsburg, R. P. Wrtz, Z. Wang, F. Xue, Y. Ma, Q. Yang, C. Fang, X. Ding, S. Lucey, R. Goss, H. Schneiderman, N. Poh, and Y. Rodriguez. Face authentication test on the banca database. *17th Int. Conf. Pattern Recognition*, pages 523–532, 2004. 6

[34] Y. Adini, Y. Moses, and S. Ullman. Face recognition: The problem of compensating for changes in illumination direction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):721–732, 2003. 6, 27

[35] W. Zhao and R. Chellappa. Robust face recognition using symmetric shape-from-shading. *technical report*, Center for Automation Research, Univ. of Maryland, 1999. 6, 27

[36] X. Zou, J. Kittler, and K. Messer. Illumination invariant face recognition: A survey. *IEEE Int. Conf. Biometrics: Theory, Applications, and Systems*, pages 1–8, 2007. 6

[37] Y. Li, C. Wang, and X. Ao. Illumination processing in face recognition. *Face Recognition*, Milos Oravec (Ed.), ISBN: 978-953-307-060-5, InTech, 2010.

[38] M. Yao and H. Nagahashi. Analysis of noteworthy issues in illumination processing for face recognition. *Submitted*. 6, 51

[39] E. Hjelmas and B.K. Low. Face detection: A survey. *Comput. Vis. Image Und.*, 83:236–274, 2001. 11, 14

[40] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(1):34–58, 2002. 14, 16, 17, 18, 19, 20, 23, 30, 31

[41] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision (IJCV)*, pages 137–154, 2004. 14

[42] M. Jones and P. Viola. Fast multi-view face detection. *Technical Report, Mitsubishi Electric Research Laboratories, TR2003-96*, 2003. 14, 15, 18, 19, 20

[43] K-K. Sung and T. Poggio. Example-based learning for view-based human face detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20 (1):39–51, 1998. 14

[44] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(1):23–38, 1998. 14

[45] H. Rowley, S. Baluja, and T. Kanade. Rotation invariant neural network-based face detection. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 38–44, 1998. 14

[46] R. Lienhart and J. Maydt. An extended set of haar-like features for rapid object detection. *Proc. of ICIP*, 2002. 14, 16, 19, 20

[47] T. Mita, T. Kaneko, and O. Hori. Joint haar-like features for face detection. *Proc. ICCV*, 2:1619–1626, 2005. 14, 16, 19, 20

[48] S. C. Brubaker, J.Wu, J. Sun, M. D. Mullin, and J. M. Rehg. On the design of cascades of boosted ensembles for face detection. Technical report, Georgia Institute of Technology, GIT-GVU-05-28, 2005. 15, 16, 17, 20

[49] S. Li, L. Zhu, Z. Zhang, A. Blake, H. Zhang, and H. Shum. Statistical learning of multi-view face detection. *Proc. ECCV*, pages 67–81, 2002. 15, 16, 17, 18, 19, 20

[50] M. Jones, P. Viola, and D. Snow. Detecting pedestrians using patterns of motion and appearance. *Technical report, Mitsubishi Electric Research Laboratories, TR2003-90*, 2003. 15, 19

[51] C. Liu and H.-Y. Shum. Kullback-leibler boosting. *Proc. Of CVPR*, 2003. 15, 19

[52] J. Meynet, V. Popovici, and J.-P. Thiran. Face detection with boosted gaussian features. *Pattern Recognition*, (8):2283–2291, 2007. 15, 19

[53] X. Chen, L. Gu, S. Z. Li, and H.-J. Zhang. Learning representative local features for face detection. *Proc. of CVPR*, 2001. 15, 19

[54] P. Wang and Q. Ji. Learning discriminant features for multiview face and eye detection. *Proc. of CVPR*, 2005. 15, 19, 20

[55] K. Levi and Y. Weiss. Learning object detection from a small number of examples: The importance of good features. *Proc. of CVPR*, 2004. 16, 19

[56] N. Dalal and B. Triggs. Histogram of oriented gradients for human detection. *Proc. of CVPR*, 2005. 16, 19

[57] M. Enzweiler and D. M. Gavrila. Monocular pedestrian detection: Survey and experiments. *IEEE Trans. on PAMI*, 31(12):2178–2195, 2009. 16

[58] H. Grabner and H. Bischof. On-line boosting and vision. *Proc. of CVPR*, 2006.

[59] I. Laptev. Improvements of object detection using boosted histograms. *British Machine Vision Conference*, 2006.

[60] F. Suard, A. Rakotomamonjy, A. Bensrhair, and A. Broggi. Pedestrian detection using infrared images and histograms of oriented gradients. *IEEE Intelligent Vehicles Symposium*, 2006.

[61] Q. Zhu, S. Avidan, M.-C. Yeh, and K.-T. Cheng. Fast human detection using a cascade of histograms of oriented gradients. *Proc. of CVPR*, 2006. 16

[62] H. Zhang, W. Gao, X. Chen, and D. Zhao. Object detection using spatial histogram features. *Image and Vision Computing*, 24(4):327–341, 2006. 16, 19

[63] X. Wang, T. X. Han, and S. Yan. An hog-lbp human detector with partial occlusion handling. *Proc. of ICCV*, 2009. 16, 19

[64] C. Huang, H. Ai, Y. Li, and S. Lao. Learning sparse features in granular space for multi-view face detection. *Intl. Conf. on Automatic Face and Gesture Recognition*, 2006. 16, 18, 19, 20

[65] A. Opelt, A. Pinz, and A. Zisserman. A boundary fragment-model for object detection. *Proc. of CVPR*, 2006. 16, 19

[66] J. Shotton, A. Blake, and R. Cipolla. Contour-based learning for object detection. *Proc. of ICCV*, 2005. 16, 19

[67] B. Wu and R. Nevatia. Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors. *Proc. of ICCV*, 2005. 16, 19

[68] P. Sabzmeydani and G. Mori. Detecting pedestrians by learning shapelet features. *Proc. of CVPR*, 2007. 16, 19

[69] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *European Conf. on Computational Learning Theory*, 1995. 16, 27

[70] C. Bishop and P. Viola. Learning and vision: Discriminative methods. *ICCV Course on Learning and Vision*, 2003. 16, 20

[71] B.Wu, H. Ai, C. Huang, and S. Lao. Fast rotation invariant multi-view face detection based on real adaboost. *Proc. of IEEE Automatic Face and Gesture Recognition*, 2004. 16, 18, 20

[72] R. Xiao, L. Zhu, and H. Zhang. Boosting chain learning for object detection. *Proc. of ICCV*, 2003. 17, 20

[73] P. Viola and M. Jones. Fast and robust classification using asymmetric adaboost and a detector cascade. *Proc. of NIPS*, 2002. 17

[74] H. Masnadi-Shirazi and N. Vasconcelos. Asymmetric boosting. *Proc. of ICML*, 2007. 17, 20

[75] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. Technical report, Dept. of Statistics, Stanford University, 1998. 17

[76] J. Wu, S. C. Brubaker, M. D. Mullin, and J. M. Rehg. Fast asymmetric learning for cascade face detection. Technical report, Georgia Institute of Technology, GIT-GVU-05-27, 2005. 17, 20

[77] J. Sochman and J. Matas. Waldboost-learning for time constrained sequential detection. *Proc. CVPR*, 2:150–156, 2005. 17, 20, 31

[78] R. Xiao, H. Zhu, H. Sun, and X. Tang. Dynamic cascades for face detection. *Proc. of ICCV*, 2007. 17, 20

[79] L. Bourdev and J. Brandt. Robust object detection via soft cascade. *Proc. of CVPR*, 2005. 17, 20

[80] C. Zhang and P. Viola. Multiple-instance pruning for learning efficient cascade detectors. *Proc. of NIPS*, 2007. 17, 20

[81] B. McCane and K. Novins. On training cascade face detectors. *Image and Vision Computing*, 2003. 17, 20

[82] H. Schneiderman. Feature-centric evaluation for efficient cascaded object detection. *Proc. of CVPR*, 2004. 18, 20

[83] S. Yan, S. Shan, X. Chen, and W. Gao. Locally assembled binary (lab) feature with feature-centric cascade for fast and accurate face detection. *Proc. of CVPR*, 2008. 18, 19, 20

[84] B. Froba and A. Ernst. Fast frontal-view face detection using a multi-path decision tree. *Proc. of Audio- and Video-based Biometric Person Authentication*, 2003. 18, 20

[85] C. Huang, H. Ai, Y. Li, and S. Lao. Vector boosting for rotation invariant multi-view face detection. *Proc. of ICCV*, 2005.

[86] Y.-Y. Lin and T.-L. Liu. Robust face detection with multiclass boosting. *Proc. of CVPR*, 2005. 18, 20

[87] T.-K. Kim and R. Cipolla. Mcboost: Multiple classifier boosting for perceptual co-clustering of images and visual features. *Proc. of NIPS*, 2008. 18, 20

[88] P. Viola, J. C. Platt, and C. Zhang. Multiple instance boosting for object detection. *Proc. of NIPS*, 2005. 18

[89] S. Baluja, M. Sahami, and H. A. Rowley. Efficient face orientation discrimination. *Proc. of ICIP*, 2004. 19

[90] Y. Abramson and B. Steux. Yef real-time object detection. *International Workshop on Automatic Learning and Real-Time*, 2005. 19

[91] C. A. Waring and X. Liu. Face detection using spectral histograms and svms. *IEEE Trans. on Systems, Man, and CyberneticsPart B: Cybernetics*, (3). 19

[92] O. Tuzel, F. Porikli, and P. Meer. Region covariance: A fast descriptor for detection and classification. *Proc. of ECCV*, 2006. 19

[93] M.-T. Pham and T.-J. Cham. Online learning asymmetric boosted classifiers for object detection. *Proc. of CVPR*, 2007. 20

[94] H. T. Luo. Optimization design of cascaded classifiers. *Proc. CVPR*, 1: 480–485, 2005. 20, 31

[95] J.-S. Jang and J.-H. Kim. Fast and robust face detection using evolutionary pruning. *IEEE Trans. on Evolutionary Computation*, (25):562–571, 2008. 20

[96] M.-T. Pham and T.-J. Cham. Detection caching for faster object detection. *Proc. of CVPR*, 2005. 20

[97] E. Seemann, B. Leibe, and B. Schiele. Multi-aspect detection of articulated objects. *Proc. of CVPR*, 2006. 20

[98] Y. Shan, F. Han, H. S. Sawhney, and R. Kumar. Learning exemplar-based categorization for the detection of multi-view multi-pose objects. *Proc. of CVPR*, 2006. 20

[99] Z. Tu. Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. *Proc. of ICCV*, 2005. 20

[100] B. Wu and R. Nevatia. Cluster boosted tree classifier for multi-view, multi-pose object detection. *Proc. of ICCV*, 2007. 20

[101] C. Zhang and Z. Zhang. Winner-take-all multiple category boosting for multi-view face detection. Technical report, Microsoft Research MSR-TR-2009-190, 2009. 20

[102] D. Keren, M. Osadchy, and C. Gotsman. Antifaces: A novel fast method for image detection. *IEEE Trans. on PAMI*, (7):747–761, 2001. 21

[103] C. Liu. A bayesian discriminating features method for face detection. *IEEE Trans. on PAMI*, (6):725–740, 2003. 21

[104] M. Ratsch, S. Romdhani, and T. Vetter. Efficient face detection by a cascaded support vector machine using haar-like features. *Pattern Recognition Symposium*, 2004. 21

[105] S. Romdhani, P. Torr, B. Scholkopf, and A. Blake. Computationally efficient face detection. *Proc. of ICCV*, 2001. 21

[106] B. Heisele, T. Serre, S. Prentice, and T. Poggio. Hierarchical classification and feature reduction for fast face detection with support vector machines. *Pattern Recognition*, pages 2007–2017, 2003. 21

[107] Y. Li, S. Gong, and H. Liddell. Support vector regression and classification based multi-view face detection and recognition. *International Conference on Automatic Face and Gesture Recognition*, 2000. 21

[108] J. Yan, S. Li, S. Zhu, and H. Zhang. Ensemble svm regression based multi-view face detection system. *International Conference on Automatic Face and Gesture Recognition*, Technical report, Microsoft Research, MSR-TR-2001-09, 2001. 21

[109] P. Wang and Q. Ji. Multi-view face detection under complex scene based on combined svms. *Proc. of ICPR*, 2004. 21

[110] K. Hotta. View independent face detection based on combination of local and global kernels. *International Conference on Computer Vision Systems*, 2007. 21

[111] R. Feraud, O. J. Bernier, J.-E. Viallet, and M. Collobert. A fast and accurate face detector based on neural networks. *IEEE Trans. on PAMI*, (1):42–53, 2001. 21

[112] C. Garcia and M. Delakis. Convolutional face finder: A neural architecture for fast and robust face detection. *IEEE Trans. on PAMI*, (11):1408–1423, 2004. 21

[113] M. Osadchy, M. L. Miller, and Y. L. Cun. Synergistic face detection and pose estimation with energy-based models. *Proc. of NIPS*, 2004. 21

[114] H. Schneiderman. Learning a restricted bayesian network for object detection. *Proc. of CVPR*, 2004. 21

[115] H. Schneiderman and T. Kanade. Object detection using the statistics of parts. *International Journal of Computer Vision*, (3):151–177, 2004. 21

[116] B. Heisele, T. Serre, and T. Poggio. A component-based framework for face detection and identification. *International Journal of Computer Vision*, (2):167–181, 2007. 21

[117] K. Mikolajczyk, C. Schmid, and A. Zisserman. Human detection based on a probabilistic assembly of robust part detectors. *Proc. of ECCV*, 2004. 21

[118] V. Vezhnevets, V. Sazonav, and A. Andreeva. A survey on pixel-based skin color detection technique. *Proc. of GraphiCon*, pages 85–92, 2003. 18

[119] K. Sandeep and A. N. Rajagopalan. Human face detection in cluttered color images using skin color and edge information. *Proc. of Indian Conference on Computer Vision, Graphics and Image Processing*, 2002. 18

[120] K. Lim, L. Ang, and K. Seng. New face segmentation technique insusceptible to illumination. *Proc. ECTI-CON*, pages 449–452, 2008. 21, 22

[121] D. Ghimire and J. Lee. A lighting insensitive face detection method on color images. *Spring Congress on Engineering and Technology*, pages 1–4, 2012. 21, 22

[122] Y. Wang, X. Ning, C. Yang, and Q. Wang. A novel method for face detection across illumination changes. *Global Congress on Intelligent Systems*, pages 374–378, 2009. 21, 22

[123] C. Lin. Face detection by color and multilayer feedforward neural network. *Proc. of Indian Conference on Information Acquisition*, pages 518–523, 2005. 21

[124] C.E. Erdem, S. Ulukaya, A. Karaali, and A.T. Erdem. Combining haar feature and skin color based classifiers for face detection. *Proc. ICASSP*, pages 1497–1500, 2011. 22, 29

[125] Z.S. Tabatabaie, R.W. Rahmat, N.I.B. Udzir, and E. Kheirkhah. A hibrid face detection system using combination of appearance-based and feature-based methods. *Intl. J. Computer Science and Network Security*, 9(5), 2009. 29

[126] L. Shobana, A. K. Yekkala, and S. Eajaz. Face detection using skin segmentation as pre-filter. *Int. Conf. on Advances in Pattern Recognition*, 2007.

[127] H. Y. Chen, C. L. Huang, and C. M. Fu. Hybrid-boost learning for multi-pose face detection and facial expression recognition. *Pattern Recognition*, 41(3).

[128] K.-B. Ge, J. Wen, and B. Fang. Adaboost algorithm based on mb-lbp features with skin color segmentation for face detection. *Proc. of Int. Conf. Wavelet Analysis and Pattern Recognition*, 2011. 22

[129] J. Cai and A. Goshtasby. Detection human faces in color images. *Image and Vision Computing*, 18(1). 22

[130] T. Ojala, M. Pietikinen, and T. Menp. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24:971–987, 2002. 23, 26, 31

[131] Jianguo Wang and Tieniu Tan. A new face detection method based on shape information. *Pattern Recognition Letters*, 21:463–471, 2000. 23, 30

[132] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4):399–458, 2003. 24

[133] X. Tan, S. Chen, Z. H. Zhou, and F. Zhang. Face recognition from a single image per person: A survey. *Computer Vision and Image Understanding*, 39(9):1725–1745, 2006.

[134] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino. 2D and 3D face recognition: A survey. *Pattern Recognition Letters*, 28(14):1885–1906, 2007. 24, 27

[135] R. Jafri and H. R. Arabnia. A survey of face recognition techniques. *Journal of Information Processing Systems*, 5(2). 24

[136] M. A. Grudin. On internal representations in face recognition systems. *Pattern Recognition*, 33. 24

[137] B. Heisele, P. Ho, J. Wu, and T. Poggio. Face recognition: component-based versus global approaches. *Computer Vision and Image Understanding*, 91. 24

[138] T. Kanade. Picture processing system by computer complex and recognition of human faces. Kyoto University, Japan, PhD. Thesis 1973.D. 24

[139] I. Craw, D. Tock, and A. Bennett. Finding face features. *Proc. of ECCV*, pages 92–96, 1992. 25

[140] L. Wiskott, J.-M. Fellous, N. Kuiger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19. 25

[141] M. Lades, J. C. Vorbrggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Computers*, 42. 25

[142] H. Shin, S. D. Kim, and H. C. Choi. Generalized elastic graph matching for face recognition. *Pattern Recognition Letters*, 28. 25

[143] A. Albiol, D. Monzo, A. Martin, J. Sastre, and A. Albiol. Face recognition using hogebgm. *Pattern Recognition Letters*, 29. 25

[144] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal Of Cognitive Neuroscience*, 3, . 26

[145] M. Turk and A. Pentland. Face recognition using eigenfaces. *Proc. of CVPR*, . 26

[146] Y. Moses, Y. Adini, and S. Ullman. Face recognition: the problem of compensating for changes in illumination direction. *Proc. of ECCV*. 26

[147] R. A. Fisher. The use of multiple measures in taxonomic problems. *Annals of Eugenics*, 7. 26

[148] J. Weng, N. Ahuja, and T. S. Huang. Learning recognition and segmentation of 3-d objects from 3-d images. *Proc. of ICCV*. 26

[149] G. C. Zhang, X. S. Huang, S. Z. Li, Y. S. Wang, and X. H. Wu. Boosting local binary pattern (lbp)-based face recognition. *Proc. of Advances In Biometric Person Authentication*, 3338. 26

[150] H. Q. Li, S. Y. Wang, and F. H. Qi. Automatic face recognition by support vector machines. *Proc. of Combinatorial Image Analysis*, 3322. 27

[151] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Readings in Speech Recognition*, 77. 27

[152] X. Zou, J. Kittler, and K. Messer. Illumination invariant face recognition: A survey. *IEEE Intl. Conf. Biometrics: Theory, Applications, and Systems*, pages 1–8, 2007. 27, 33, 85

[153] Y. Li, C. Wang, and X. Ao. Illumination processing in face recognition. *Face Recognition, Milos Oravec (Ed.), ISBN: 978-953-307-060-5, InTech*, 2010. 27, 33, 85

[154] C. Hesher, A. Srivastava, and G. Erlebacher. A novel technique for face recognition using range imaging. *Intl. Symp. Signal Processing and Its Applications*, 2:1–4, 2003. 27

[155] J. Kittler, A. Hilton, M. Hamouz, and J. Illingworth. 3D assisted face recognition: a survey of 3D imaging, modeling and recognition approaches. *IEEE CVPR workshops*, pages 114–120, 2005.

[156] K. W. Bowyer, K. Chang, and P. Flynn. 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. 27

[157] S. G. Kong, J. Heo, B. R. Abidi, J. Paik, and M. A. Abidi. Recent advances in visual and infrared face recognition: A review. *Computer Vision and Image Understanding*, 97(1):103–135, 2005. 27

[158] A. Gyaourova, G. Bebis, and I. Pavlidis. Fusion of infrared and visible images for face recognition. *ECCV*, pages 456–468, 2004.

[159] D. A. Socolinsky and A. Selinger. Thermal face recognition over time. *ICPR*, pages 189–190, 2004.

[160] D. A. Socolinsky and A. Selinger. Thermal face recognition in an operational scenario. *CVPR*, 2:1012–1019, 2004. 27

[161] X. Zou, J. Kittler, and K. Messer. Face recognition using active near-IR illumination. *BMVC*, 2005. 27

[162] S. Z. Li, R. Chu, S. Liao, and L. Zhang. Illumination invariant face recognition using near-infrared images. *IEEE Trans. PAMI*, 29(4):627–639, 2007.

[163] J. Chen, D. Yi, J. Yang, G. Zhao, S. Z. Li, and M. Pietikainen. Learning mappings for face synthesis from near infrared to visual light images. *CVPR*, pages 156–163, 2009. 27

[164] R. M. Makwana. Illumination invariant face recognition: A survey of passive methods. *Procedia Computer Science*, 2:101–110, 2010. 27

[165] H. Han, S. Shan, X. Chen, and W. Gao. A comparative study on illumination preprocessing in face recognition. *Pattern Recognition*, 46(6):1691–1699, 2013. 27

[166] *Available: urlhttp://en.wikipedia.org/wiki/Histogram_equalization.* 27, 62, 96

[167] *Available:* `http://en.wikipedia.org/wiki/Gamma_correction`. 27, 96

[168] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(6):643–660, 2001. 28, 37, 61, 80, 104

[169] R. Basri and D. W. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(2):218–233, 2003. 28

[170] D. J. Jobson, Z. Rahman, and A. Glenn. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans. Image Process.*, 6(7):965–976, 1997. 28

[171] Y. K. Park, S. L. Park, and J. K. Kim. Retinex method based on adapative smoothing for illumination invariant face recognition. *Signal Process.*, 88 (8):1929–1945, 2008. 28, 33, 62, 77, 88, 100, 107, 108

[172] H. Wang, S. Z. Li, and Y. Wang. Face recognition under varying lighting conditions using self quotient image. *IEEE Int. Conf. Automatic Face and Gesture Recognition*, pages 819–824, 2004. 28, 62, 88, 106, 107

[173] X. Xie, J. Lai, and W. S. Zheng. Extraction of illumination invariant facial features from a single image using nonsubsampled contourlet transform. *Pattern Recognition*, 43(12):4177–4189, 2010. 28

[174] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. S. Huang. Total variation models for variable lighting face recognition. *IEEE Trans Pattern Anal. Mach. Intell.*, 28(9):1519–1524, 2006. 28, 89, 100, 107, 108

[175] S. Du and R. Ward. Wavelet-based illumination normalization for face recognition. *IEEE Intl. Conf. Image Process.*, 2, 2005. 28, 33, 62, 93, 96, 106, 107

[176] T. Zhang, B. Fang, Y. Yuan, Y. Y. Tang, Z. Shang, D. Li, and F. Lang. Multiscale facial structure representation for face recognition under varying illumination. *Pattern Recognition*, 42(2):251–258, 2009. 28, 90

[177] T. Zhang, Y. Y. Tang, B. Fang, Z. Shang, and X. Liu. Face recognition under varying illumination using gradientfaces. *IEEE Trans. Image Process.*, 18(11):2599–2606, 2009. 28, 33, 53, 62, 77, 91, 97, 106, 107, 108

[178] B. Wang, W. Li, W. Yang, and Q. Liao. Illumination normalization based on weber's law with application to face recognition. *IEEE Signal Process. Lett.*, 18(8):462–465, 2011. 28, 33, 53, 54, 62, 77, 92, 107, 108

[179] Rein-Lien Hsu, M. Abdel-Mottaleb, and A.K. Jain. Face detection in color images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24:696–706, 2002. 29

[180] S. W. Lyu. Infomax boosting. *Proc. CVPR*, 1:533–538, 2005. 31

[181] J. Sun, J. M. Rehg, and A. F. Bobick. Automatic cascade training with perturbation bias. *Proc. CVPR*, 2:276–283, 2004. 31

[182] *OpenCV documentation. Available:* `http://docs.opencv.org/`. 31, 37

[183] C. Chao, Q. Chen, and X. Sui. Range limited bi-histogram equalization for image contrast enhancement. *Optik*, 124:425–431, 2013. 32, 34

[184] R.C. Gonzalez and R.E. Woods. Digital image processing. *Second Edition, Prentice-Hall, Upper Saddle River, New Jersey*, 2002. 32, 93

[185] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. T. H. Romeny, and J. B. Zimmerman. Adaptive histogram equalization and its variations. *Comput. Vis. Graph. Image Process.*, 39(3): 355–368, 1987. 32

[186] J. B. Zimmerman, S. B. Cousins, K. M. Hartzell, M. E. Frisse, , M. G. Kahn, and H. Nagahashi. A psychophysical comparison of two methods for adaptive histogram equalization. *J. Digital Imaging*, 2:82–91, 1989. 32

[187] Y.-T. Kim. Contrast enhancement using brightmess preserving bi-histogram equalization. *IEEE Tran. Consum. Electron.*, 43:1–8, 1997. 32

[188] V. Struc, J. Zibert, and N. Pavesic. Histogram remapping as a preprocessing step for robust face recognition. *WSEAS Trans. Information Science and Applications*, 6:520–529, 2009. 33, 85

[189] H. Han, S. Shan, X. Chen, and W. Gao. A comparative study on illumination preprocessing in face recognition. *Pattern Recogn.*, 46:1691–1699, 2013. 33

[190] W. Chen, M. J. Er, and S. Wu. Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain. *IEEE Trans. Syst., Man, Cybern., B*, 36(2):458–466, 2006. 33, 62, 89, 106, 107

[191] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans. Image Process.*, 19 (6):1635–1650, 2010. 33, 96, 97, 106, 107, 108

[192] G.Y. An and J.Y. Wuand Q.Q. Ruan. An illumination normalization model for face recognition under varied lighting conditions. *Pattern Recogn. Lett.*, 31(9):1056–1067, 2010. 34, 95

[193] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Trans. Syst., Man, Cybern.*, 9(1):62–66, 1979. 34

[194] M. Leszczynski. Image preprocessing for illumination invariant face verification. *J. Telecommunications and Information Technology*, pages 19–25, 2010. 35

[195] *Face detection homepage. Available:* `http://www.facedetection.com`. 47

[196] P. H. Lee, S. W. Wu, and Y. P. Hung. Illumination compensation using oriented local histogram equalization and its application to face recognition. *IEEE Trans. Image Process.*, 21(9):4280–4289, 2012. 56, 94, 96

[197] X. Xie and K. M. Lam. An efficient illumination normalization method for face recognition. *Pattern Recognition Letters*, 27(6):609–617, 2006. 59, 92, 96

[198] J. Chen, S. Shan, C. He adn G. Zhao, M. Pietikainen, X. Chen, and W. Gao. Wld: A robust local image descriptor. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(9):1705–1720, 2010. 59

[199] Terence Sim, Simon Baker, and Maan Bsat. The cmu pose, illumination, and expression database. *IEEE Trans Pattern Anal. Mach. Intell.*, 25(12): 1615–1618, 2003. 61, 77, 105

[200] K. C. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans Pattern Anal. Mach. Intell.*, 27(5):684–698, 2005. 61, 64, 77

[201] *Extended Yale B Face Database. Available:* ***http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html***. 61, 62, 64, 77

[202] V. Struc and N. Pavesi. Illumination invariant face recognition by non-local smoothing. *BIOID Multicomm September*, 2009. 69, 77

[203] B. Mandelbrot. *Les objets fiactals: forme, hasard et dimension*, 1st ed., 1975. 70

[204] R. Lopes and N. Betrouni. Fractal and multifractal analysis: a review. *Med. Image Anal.*, 13(4):634–649, 2009. 71, 73

[205] C. Atupelage. Multifractal feature descriptor for cancer grading of histological images. Doctoral Thesis, 2013. 71

[206] C. Atupelage, H. Nagahashi, M. Yamaguchi, T. Abe, A. Hashiguchi, and M. Sakamoto. Classification of prostate histopathology images based on multifractal analysis. *IEICE Trans. Inf. & Syst.*, E95–D(12):3037–3045, 2012. 73, 75

[207] A. Hemsley and R. Mukundan. Multifractal measures for tissue image classification and retrieval. *IEEE Int. Symp. Multimedia*, pages 618–623, 2009. 73, 80

[208] J. Zhong and R. Ning. Image denoising based on wavelets and multifractals for singularity detection. *IEEE Trans. Image Process.*, 14(10):1435–1447, 2005. 73

[209] D. Russell, J. Hanson, and E. Ott. Dimension of strange attractors. *Phys. Rev. Lett.*, 45(14):1175–1178, 1980. 74

[210] I. Reljin, B. Reljin, I. Pavlovic, and I. Rakocevic. Multifractal analysis of gray-scale images. *IEEE Electrotechnical Conf., MELECON 2000*, 2: 490–493, 2000. 75

[211] S. H. Abdul Jauwad and R. Ullah. On robustness of multi fractal spectrum to geometric transformations and illumination. *ICIEIS 2011, Part II, CCIS 252*, pages 76–95, 2011. 75

[212] J. Levy Vehel and P. Mignot. Multifractal segmentation of images. *Fractals*, 2(3):371–378, 1994. 75

[213] M. Leszczynski. Image preprocessing for illumination invariant face verification. *J. Telecommunication and Information Technology*, 4:19–25, 2010. 85, 88

[214] R. M. Makwana, V. K. Thakar, and N. C. Chauhan. Evaluation and analysis of illumination normalization methods for face recognition. *IEEE Intl. Conf. Image Information Processing*, pages 1–6, 2011. 85

[215] B.K.P. Horn. Robot vision. *MIT Press, Cambridge, MA, ISBN: 978-026-208-159-7*, 1986. 86

[216] G. C. Feng and P. C. Yuen. Recognition of head-and-shoulder face image using virtual frontal-view image. *IEEE Trans. Syst., Man, Cybern., A*, 30 (6):871–882, 2000. 86

[217] B. W. Hwang and S. W. Lee. Reconstruction of partially damaged face images based on morphable face model. *IEEE Trans. PAMI*, 25(3):365–372, 2003. 86

[218] J. Ahlberg. CANDIDE-3–an updated parameterized face. *Report No. LiTH-ISY-R-2326*, 2001. 87

[219] D. Strong and T. Chan. Edge-preserving and scale-dependent properties of total variation regularization. *Inverse Problems*, 19(6):S165–S187, 2003. 89

[220] K. Delac, M. Grgic, and T. Kos. Sub-image homomorphic filtering technique for improving facial idenrification under difficult illumination conditions. *Intl. Conf. Systems, Signals and Image Processing*, 1:21–23, 2006. 93

[221] G. Heusch, Y. Rodriguez, and S. Marcel. Local binary patterns as an image preprocessing for face authentication. *IEEE Intl. Conf. Automatic Face and Gesture Recognition*, pages 9–14, 2006. 94

[222] Q. Tao and R. Veldhuis. Illumination normalization based on simplified local binary patterns for a face verification system. *IEEE Biometrics Symp.*, pages 1–6, 2007. 94

[223] Z. Lian, M. J. Er, and J. Li. A novel illumination normalization method based on local relation map. *IEEE Intl. Conf. Industrial Electronics and Applications*, pages 250–253, 2012. 94

[224] Y. Aini, Y. Moses, and S. Ullman. Face recognition: the problem of compensating for changes in illumination direction. *IEEE Trans. PAMI*, 19(7): 721–732, 1997. 95

[225] X. Cao, W. Shen, L. G. Yu, Y. L. Wang, J. Y. Yang, and Z. W. Zhang. Illumination invariant extraction for face recognition using neighboring wavelet coefficients. *Pattern Recognition*, 45(5):1299–1305, 2012. 95, 97

[226] X. Xie, W. S. Zheng, J. Lai, P. C. Yuen, and C. Y. Suen. Face illumination normalization on large and small scale features. *CVPR*, pages 1–8, 2008. 96, 98

[227] T. Matsukawa, T. Okabe, and Y. Sato. Illumination normalization of faces images with cast shadows. *ICPR*, pages 1848–1851, 2012. 96, 98

[228] S. Shan, W. Gao, B. Cao, and D. Zhao. Illumination normalization for robust face recognition against varying lighting conditions. *IEEE Intl. Workshop Analysis and Modeling of Faces and Gestures*, pages 157–164, 2003. 96

[229] H. Han, S. Shan, X. Chen, and W. Gao. Maximizing intra-individual correlations for illumination-insensitive face recognition. *IEEE Intl. Conf. Image Processing*, pages 3833–3836, 2010. 96, 103, 104

[230] Y. Cheng, Z. Jin, and C. Hao. Illumination normalization based on 2D gaussian illumination model. *Pattern Recognition Letters*, 27(6):609–617, 2006. 97

[231] Z. Zheng, F. Yang, W. Tan, J. J, and J. Yang. Gabor feature-based face recognition using supervised locality preserving projection. *Signal Processing*, 87(10):2473–2483, 2007. 97

[232] C.C. Liu and D.Q. Dai. Gabor feature-based face recognition using supervised locality preserving projection. *IEEE Trans. Image Process.*, 18(11): 2593–2599, 2009. 97

[233] X. Xie, W. S. Zheng, J. Lai, P. C. Yuen, and C. Y. Suen. Normalization of face illumination based on large-and small- scale features. *IEEE Trans. Image Process.*, 20(7):1807–1821, 2011. 98

[234] *Available:* `https://people.cs.umass.edu/~elm/congealing/`. 114

Publication List

Journal Articles

1. **Min Yao**, Hiroshi Nagahashi. Illumination Insensitive Face Representation for Face Recognition based on Modified Weberface. *International Journal of Advances in Engineering & Technology*, vol. 6, no. 5, pp. 1995–2005, 2013.

2. **Min Yao**, Hiroshi Nagahashi, Kota Aoki. Illumination Normalization-based Face Detection under Varying Illumination. *IEICE Transactions on Information and Systems*, vol. E97-D, no. 6, pp. 1590–1598, 2014.

3. **Min Yao**, Hiroshi Nagahashi. Analysis of Noteworthy Issues in Illumination Processing for Face Recognition. *IEICE Transactions on Information and Systems*. (*In review*).

International Conferences

1. **Min Yao**, Kota Aoki, Hiroshi Nagahashi. Photograph Based Pair-matching Recognition of Human Faces. *WASET International Conference on Machine Learning and Pattern Recognition (ICMLPR 2011)*, pp. 791–798, Dec. 2011.

2. **Min Yao**, Kota Aoki, Hiroshi Nagahashi. Segmentation-based Illumination Normalization for Face Detection. *IEEE 6th International Workshop on Computational Intelligence & Applications (IWCIA 2013)*, pp. 95–100, Jul. 2013.

3. **Min Yao**, Hiroshi Nagahashi. Oriented Weberface for Illumination Insensitive Face Representation in Face Recognition. *IEEE 3rd International Conference on Computer Science and Automation Engineering (CSAE 2013)*, pp. 1623–1627, Nov. 2013.

Domestic Conferences

1. **Min Yao**, Hiroshi Nagahashi. Extraction of Illumination Insensitive Facial Features Using Multifractal Analysis for Face Recognition. *Pattern Recognition and Media Understanding (PRMU 2014)*, IEICE Technical Report, June 2014.

2. **Min Yao**, Hiroshi Nagahashi. Study on Illumination Insensitive Face Detection Based on Normalization Techniques. *Pattern Recognition and Media Understanding (PRMU 2013)*, IEICE Technical Report, vol. 113, no. 196, pp. 1–6, Sep. 2013.

3. **Min Yao**, Kota Aoki, Hiroshi Nagahashi. A simple approach tackling illumination problem of face detection. In *Proceedings of IEICE General Conference 2013*, p. 115, Japan, March 2013.