2016

Cross-Spectral Full and Partial Face Recognition: Preprocessing, Feature Extraction and Matching

Zhicheng Cao

Recommended Citation

Cao, Zhicheng, "Cross-Spectral Full and Partial Face Recognition: Preprocessing, Feature Extraction and Matching" (2016). *Graduate Theses, Dissertations, and Problem Reports*. 5311.
https://researchrepository.wvu.edu/etd/5311

Cross-Spectral Full and Partial Face Recognition: Preprocessing, Feature Extraction and Matching

Zhicheng Cao

Dissertation submitted to the
College of Engineering and Mineral Resources
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Electrical Engineering

Natalia A. Schmid, Ph.D., Chair
Thirimachos Bourlai, Ph.D.
Xin Li, Ph.D.
Keith Morris, Ph.D.
Matthew C. Valenti, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia
2016

Keywords: face recognition, cross-spectral, periocular, partial face, heterogeneous, feature extraction, composite operators, fusion, levels of measure, quality disparity, varying ranges

Abstract

Cross-Spectral Full and Partial Face
Recognition: Preprocessing, Feature Extraction and Matching

Zhicheng Cao

Cross-spectral face recognition remains a challenge in the area of biometrics. The problem arises from some real-world application scenarios such as surveillance at night time or in harsh environments, where traditional face recognition techniques are not suitable or limited due to usage of imagery obtained in the visible light spectrum. This motivates the study conducted in the dissertation which focuses on matching infrared facial images against visible light images. The study outspreads from aspects of face recognition such as preprocessing to feature extraction and to matching.

We address the problem of cross-spectral face recognition by proposing several new operators and algorithms based on advanced concepts such as composite operators, multi-level data fusion, image quality parity, and levels of measurement. To be specific, we experiment and fuse several popular individual operators to construct a higher-performed compound operator named GWLH which exhibits complementary advantages of involved individual operators. We also combine a Gaussian function with LBP, generalized LBP, WLD and/or HOG and modify them into multi-lobe operators with smoothed neighborhood to have a new type of operators named Composite Multi-Lobe Descriptors. We further design a novel operator termed Gabor Multi-Levels of Measurement based on the theory of levels of measurements, which benefits from taking into consideration the complementary edge and feature information at different levels of measurements.

The issue of image quality disparity is also studied in the dissertation due to its common occurrence in cross-spectral face recognition tasks. By bringing the quality of heterogeneous imagery closer to each other, we successfully achieve an improvement in the recognition performance. We further study the problem of cross-spectral recognition using partial face since it is also a common problem in practical usage. We begin with matching heterogeneous periocular regions and generalize the topic by considering all three facial regions defined in both a characteristic way and a mixture way.

In the experiments we employ datasets which include all the sub-bands within the infrared spectrum: near-infrared, short-wave infrared, mid-wave infrared, and long-wave infrared. Different standoff distances varying from short to intermediate and long are considered too. Our methods are compared with other popular or state-of-the-art methods and are proven to be advantageous.

Acknowledgements

My deepest gratitude goes to my committee chair and advisor Dr. Natalia A. Schmid, who consistently and generously supported my research, gave me invaluable advice, encouraged and inspired me during the past six years. This dissertation would not be possible without her enormous work and help. I am deeply impressed by her passionate, rigorous and diligent attitude toward both teaching and research, as well as her kindness and caring. It was a great pleasure to learn in her classes, to carry out research work with her, and to know her as a mentor and friend.

I would also like to thank Dr. Thirimachos Bourlai, Dr. Xin Li, Dr. Keith Morris and Dr. Matthew C. Valenti for serving as my committee members. They spent precious time and effort on giving suggestions to my research and helping to improve and refine this dissertation.

I am grateful to Dr. Brian E. Lemoff from West Virginia High Technology Consortium Foundation, Dr. Stephanie Schuckers from Clarkson University, and the Pinellas County Sheriff's Office for providing the datasets employed in this work.

Hereby I also thank all my lab colleagues and friends. My special thanks go to Francesco and Jinyu who provided great help when I first started my doctoral study here at WVU.

Finally I would like to thank my whole family, especially my parents who gave me life, nurtured me, and love and support me unconditionally.

Contents

List of Figures

List of Tables

List of Abbreviations

ASM Adaptive Sharpness Measure

CCA Canonical-Correlation Analysis

CCD charge-coupled device

CDFE Common Discriminant Feature Extraction

CMC Cumulative Match Curve

CMLD Composite Multi-Lobe Descriptor

DOG Difference of Gaussian

EBGM Elastic Bunch Graph Matching

EM Expectation Maximization

FAR false acceptance rate

FRR false rejection rate

FERET Face Recognition Technology Evaluation

FIR far-infrared

FRGC Face Recognition Grand Challenge

FRVT Face Recognition Vendor Tests

GAR genuine acceptance rate

GIM Gabor Interval Measure

GLBP Generalized Local Binary Patterns

GOM Gabor Ordinal Measures

GWLH Gabor-WLD-LBP-HOG

GMLM Gabor Multi-Levels of Measurement

HCI human-computer interaction

HOG Histograms of Gradients

HMM hidden Markov model

ICA Independent Component Analysis

IM Interval Measure

IMD Interval Measure Descriptor

IR infrared

LATP Local Adaptive Ternary Pattern

LBP Local Binary Patterns

LDA Linear Discriminant Analysis

LTP Local Ternary Pattern

LWIR long-wave infrared

MB-LBP Multiscale Block Local Binary Patterns

MSL Multilinear Subspace Learning

MWIR mid-wave infrared

NFIQ NIST Fingerprint Image Quality

NIR near-infrared

PCA Principal Component Analysis

PCSO Pinellas County Sheriff's Office

Pre-TINDERS Pre-Tactical Imager for Night/Day Extended-Range Surveillance

Q-FIRE Quality-Face/Iris Research Ensemble

ROC receiver operating characteristic

SIFT Scale-Invariant Feature Transform

SWIR short-wave infrared

SWLD simplified Weber Local Descriptor

SVM Support Vector Machines

SNR signal-to-noise ratio

TINDERS Tactical Imager for Night/Day Extended-Range Surveillance

WLD Weber Local Descriptor

Chapter 1

Introduction

The term *biometrics* is derived from the Greek word βίος, *bios*, "life" and the suffix μέτρον, *metron*, "measure". It refers to measuring human physiological or behavioral characteristics such as face, fingerprint, iris, retina, DNA, hand geometry, voice, gait and signature [2–7]. Biometrics has now become a science of measuring a person's characteristics to identify or verify her/his identity automatically by means of computers. A system implementing this is called a biometric system which processes, encodes and matches some given traits against similar data in a database.

Biometric systems can be used to carry out identification or verification tasks depending on the specific request given to the system. An identification system selects the best matching subject out of a long list of candidates using trait measurements of the input subject, while a verification system matches the input subject against the claimed identity and then confirms or rejects that claim. Applications of biometrics include access control, surveillance, general identity management and human-computer interaction (HCI) (e.g. multi-media environments). They can be found in all kinds of scenarios such as consumer and residential, financial, health care, justice and law enforcement applications. As a result, biometrics is used on the front doors of thousands of businesses around the world, at the doors to the tarmacs of major airports, and at the entrances of other facilities where the combination of security and convenience is desired.

This dissertation is devoted to the biometrics using face as the trait known as face recognition. Compared with other biometric modalities, one important advantage of face

recognition is that it does not require the cooperation of the test subject. The main case of interest discussed in this work is matching face images acquired in the infrared (IR) band of the electromagnetic wave against face images acquired in visible light, in other words cross-spectral face recognition. The IR band includes near-infrared (NIR), short-wave infrared (SWIR), mid-wave infrared (MWIR) and long-wave infrared (LWIR). The main focus of this dissertation is to develop new high performance operators for cross-spectral face and partial face recognition.

1.1 Face Recognition: A Glance through History

Humans use faces to recognize individuals on a daily basis. Sir Francis Galton was the first to study this subject using face profiles to describe and identify individuals [8]. Advancements in computing capability over the past few decades now enable this process to be accomplished automatically. Pioneers of automated facial recognition include Woody Bledsoe, Helen Chan Wolf and Charles Bisson who during the 1960s developed the first semi-automated system for facial recognition. Their system required an administrator to locate features – such as eyes, ears, nose, and mouth – on photographs before it calculated distances and ratios to a common reference point, which were then compared to reference data. In the early 1970s, Goldstein, Harmon and Lesk used 21 specific subjective markers such as hair color and lip thickness to automate the recognition [9]. The problem of measurements and locations being manually computed still resides with their solutions. Later on, Kanade et al. [10] employed simple image processing methods to extract fiducial markers on the face and defined a vector of 16 facial parameters – ratios of distances, areas. They used a simple Euclidean distance measure for matching and achieved a peak performance of 75% on a database of 20 different people using 2 images per person.

Since the 1990s face recognition attracted more and more attention of researchers and has been an active research topic. The work of Kirby and Sirovich on Principal Component Analysis (PCA) [11] represents a milestone in the area of face recognition technique. They showed that less than one hundred values were required to accurately code a suitable aligned and normalized face. As a follow up, Turk and Pentland [12] discovered that while using the

eigenfaces techniques, the residual error could be used to detect faces in images, a discovery that enabled reliable real-time automated face recognition systems. As researchers' interest in face recognition continued, many other algorithms were developed. Of these methods the predominant ones can be categorized into two: geometric and photometric. Popular examples of the methods in face recognition literature are PCA, Linear Discriminant Analysis (LDA) [13], Elastic Bunch Graph Matching (EBGM) [14], the hidden Markov model (HMM) [15,16], Gabor filters [17], Local Binary Patterns (LBP) [18], etc.

The US Government has performed multiple evaluations to determine the capabilities and limitations of face recognition, and to encourage and direct future developments. The Face Recognition Technology Evaluation (FERET) [19] was sponsored by the Defense Advanced Research Products Agency (DARPA) from 1993 through 1997. It was an effort to encourage the development of face recognition algorithms and technology by assessing the prototypes of face recognition systems.

Face Recognition Vendor Tests (FRVT) [20] was performed in 2000, 2002, and 2006. These evaluations built upon the work of FERET. The two goals were to assess the capabilities of commercially available facial recognition systems and to educate the public on how to properly present and analyze results.

Face Recognition Grand Challenge (FRGC) [21] was the next step in the government development and evaluation process – to promote and advance face recognition technology designed to support existing face recognition efforts of the US Government. It evaluated the latest face recognition algorithms available. High-resolution face images, 3D face scans, and iris images were used in the tests.

1.2   Face Recognition System

A full face recognition system usually comprises the following blocks: image acquisition, face detection, preprocessing and normalization, feature extraction and matching.

- Image acquisition: Acquisition of a face image using cameras such as charge-coupled device (CCD) cameras, 3D cameras, IR cameras, etc. (See Figure 1.1 for an example);

- Face detection: Detection of the face region in the image (An example is shown in Figure 1.2);

- Preprocessing and normalization: The adjustment of pixel intensity, cropping, geometrical transformation, etc.;

- Feature extraction: Extraction of features characterizing the face pertaining to a specific subject;

- Matching: Comparison of the input feature set (the probe) with the reference sets (the gallery) using a certain metric (the matching score) and finally decision of whether the two images come from the same subject.

Figure 1.1: Image acquisition using a thermal infrared camera
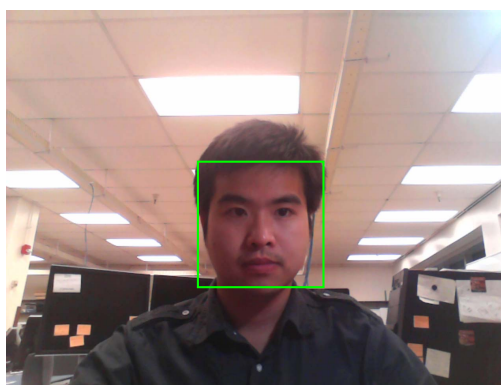
Figure 1.2: An example of face detection

A typical face recognition system has two operation modes: (a) enrollment and (b) recognition. The first time an individual uses such a system is called enrollment. During the enrollment, face information from an individual is captured and stored. The individual usually stands at a close distance from the capturing camera. In subsequent uses, face information is detected and compared with the information stored at the time of enrollment. Note that it is crucial that storage and retrieval of such systems themselves be secure if the face recognition system is to be robust. A block diagram of a face recognition system working in the enrollment mode is shown in Fig. 1.3.
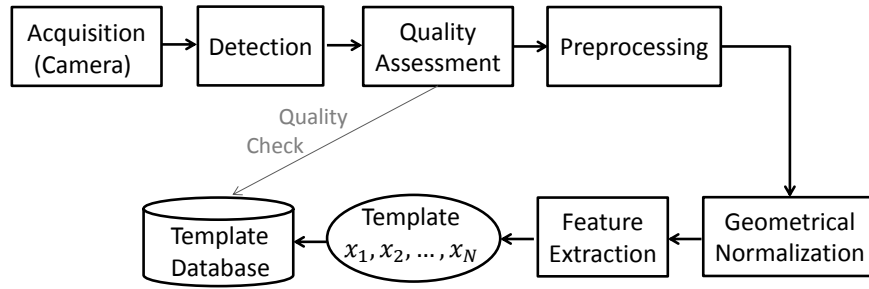
Figure 1.3: A face recognition system working in the enrollment mode.

The enrollment process in Fig. 1.3 can be summarized as follows: (1) a video frame (image) is acquired; (2) face and additional landmarks are detected; (3) a vector of quality measures is generated for the biometric sample; (4) the acquired face sample is discarded or retained depending on its quality; (5) the face image is geometrically normalized to a canonical form; (6) the normalized face image is pre-processed (transformed to gray scale, convolved with a filter, etc.); (7) a feature vector is extracted from the processed face image (encoding stage); (8) the obtained feature vector is stored as a template in the database to form a gallery image in a gallery set and (optionally) the quality measures of the sample are stored together with the template.

When the face recognition system is working in the recognition mode, the probe image can be acquired at a close or far distance and the individual may or may not be aware of being captured by a camera. The face recognition process is depicted in Fig. 1.4. In Fig. 1.4, note that the first seven steps (1)-(7) of the recognition process overlap with the steps
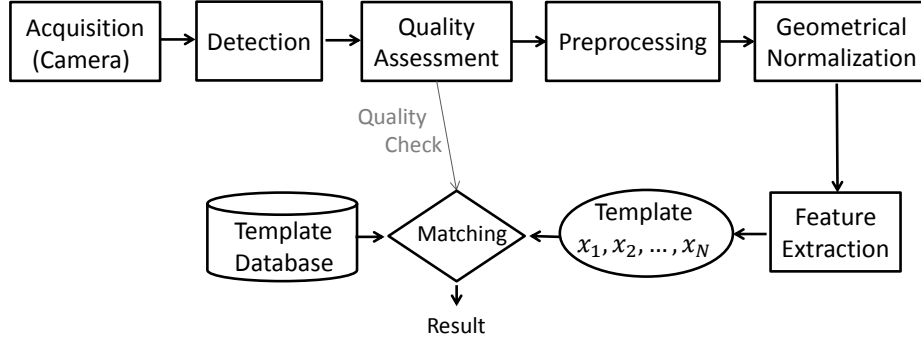
Figure 1.4: A face recognition system working in the recognition mode.

of the enrollment procedure. The last blocks perform matching: (8) the probe feature vector is compared against a gallery set of templates previously stored in a database and matching scores are generated. The matching block may employ a distance (dissimilarity) measure as a matching score or alternatively adopts a similarity score; (9) a verification or identification result is produced depending on the specific case.

The verification case performs a 1-to-1 matching. A probe is claimed to belong to a specific entry in the gallery. Therefore, it is compared to only this entry. In this mode of operation, biometric systems compare the obtained matching score to a given threshold to produce a genuine or impostor decision. In the identification case (1-to-N matching) the face recognition system searches through the entire gallery set by comparing each gallery entry to a probe. In this case, the obtained matching scores can be sorted by value to define the best match. This sorting results in ranking.

The implementation in this work follows the schematics depicted in Fig. 1.3 and Fig. 1.4. However, the focus is placed on designing the preprocessing, feature extraction and matching modules.

1.3   Evaluation and Performance

To evaluate the accuracy of a biometric system, score distributions for genuine and impostor comparisons are generated for the whole dataset. The precision of a biometric system is characterized by its ability to separate between the two distribution functions. It defines

a threshold setting the boundary between the two distributions: if the score of a comparison is bellow the threshold, it is considered as genuine. Otherwise, it is considered as impostor.

In an ideal case, the two distribution functions do not overlap with each other. The system can therefore perfectly distinguish between the genuine class and the impostor class (See Figure 1.5 (a)). However, in reality, the two distributions do overlap (See Figure 1.5 (b)). As a result, some of the genuine matching will be mistakenly taken as impostors which yields the false rejection rate (FRR). Likewise, some of the impostor matching will be mistakenly considered as a genuine class which yields the false acceptance rate (FAR).
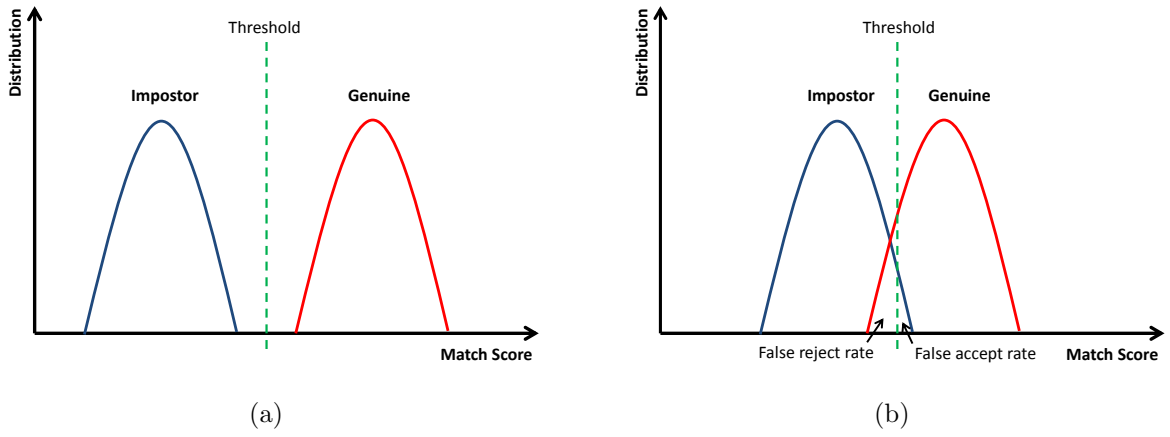
Figure 1.5: Distribution of genuine and impostor scores: (a) ideal case; (b) real case.

As explained earlier, the evaluation of a biometric system is done using the genuine and impostor distributions for a reference database. Estimation of FAR and FRR for a given threshold t is:

$$FAR(t) = \int_t^\infty p(s|H_0)ds, \tag{1.1}$$

$$FRR(t) = \int_{-\infty}^t p(s|H_1)ds, \tag{1.2}$$

An illustration of FAR and FRR vs different thresholds is shown in Figure 1.6 (a). When putting FAR and FRR or genuine acceptance rate (GAR) (i.e., $1 - FRR$) together in one plot, one gets the receiver operating characteristic (ROC) curve, as shown in Figure 1.6 (b).

(a)   (b)
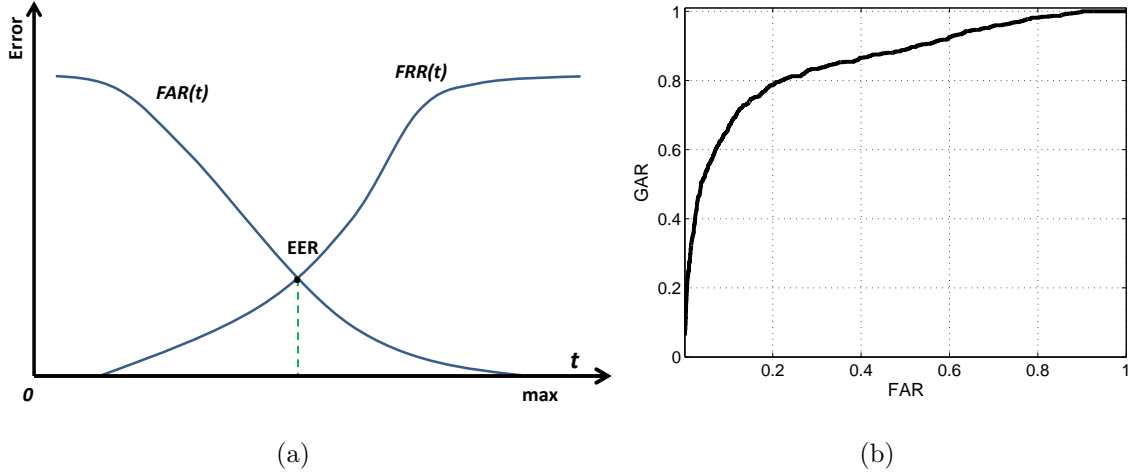
Figure 1.6: (a) Plot of FRR and FAR with different thresholds and (b) ROC curve.

1.4 New Topics

Although lots of advances in the research and applications of face recognition have been witnessed, many problems and difficulties still exist. One of them is that most methods developed for face recognition require a constrained condition. Unfortunately, the human face is not a unique or rigid object and there are numerous factors that cause the appearance of the face to vary. This problem becomes worse when the variations between the images of the same face due to these factors are larger than image variations due to change in face identity, which is not a rare case. The sources of variation in facial appearance [22] can be categorized into two groups: (a) intrinsic factors and (b) extrinsic ones. Intrinsic factors are due purely to the physical nature of the face and are independent of the observer. Examples include age, facial expression, and facial paraphernalia (e.g. facial hair, glasses, cosmetics). Extrinsic factors cause the appearance of the face to alter via the interaction of light with the face and the observer. These factors include illumination, pose, scale and imaging parameters (e.g., resolution, focus, imaging, noise, etc.).

Since factors such as illumination, age, facial expression and pose plague face recognition systems and prevent them from achieving high performance, recent research efforts have been made to explore alternate face modalities such as IR [23] and 3-D face model for face recognition [24]. New studies have also been devoted to fusion of multiple face modali-

ties and results have demonstrated improved performance compared with single modal face recognition [25]. There is also attention paid to the topics of aging in face recognition [26], expression recognition [27] and gender classification [28].

In the following subsections we will give a brief introduction to the first two topics: face recognition in the infrared spectrum and 3D face recognition.

1.4.1 Face Recognition in the Infrared Spectrum

IR is invisible electromagnetic radiation with longer wavelength than that of visible light, extending from the nominal red edge of the visible spectrum at 700 nm (frequency 430 THz) to 1 mm (300 GHz). Figure 1.7 shows the IR band in the electromagnetic spectrum. Most of the thermal radiation emitted by objects near room temperature is IR. Objects emit different amounts of IR energy according to their temperature and characteristics. Therefore, IR images or thermograms are good for capturing of the patterns associated with heat emission of an object. The range of human face and body temperature is quite uniform, varying from 35.5°C to 37.5°C providing a consistent thermal signature. It is known that even identical twins have different thermal patterns [29]. The range and sensitivity are well within the specification of current IR imaging technology. An example of LWIR images is shown in Figure 1.8.

Since an external source of illumination is not required for thermal IR or less affected by the illumination source for active IR (i.e., NIR and SWIR), it is obviously an advantage for face recognition using IR that it is less susceptible to lighting condition. This is in contrast to the difficulties encountered in the visible light spectrum due to physical diversity coupled with lighting, color, and shadow effects. Another advantage of face recognition in IR is seen when conducting surveillance at night or in harsh environments, which is one of the most recent applications of face recognition. Latest advancements in manufacturing of small and cheap imaging devices sensitive in active infrared range (near- and short- infrared) [30, 31] and the ability of these cameras to see through fog, rain, at night and operate at long ranges provided researchers with new type of imagery and posed new research problems [32–40]. As observed, active IR energy is less affected by scattering and absorption by smoke or dust
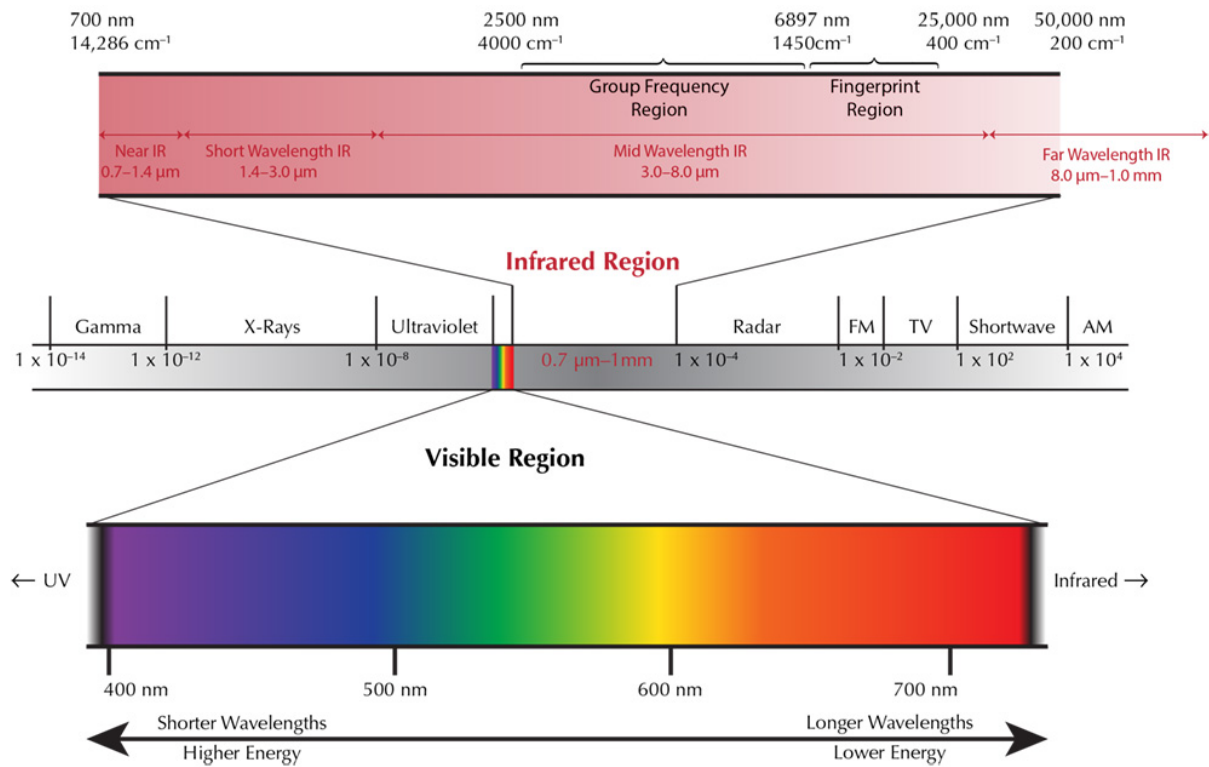
Figure 1.7: The infrared and visible bands in the electromagnetic spectrum.

Figure 1.8: An example of LWIR images

than visible light.

1.4.2   3-Dimensional Face Recognition

Face recognition from 3D range image data is a newly emerging trend which has been claimed to achieve improved recognition accuracy [41]. As a face is inherently a 3D object, it is beneficial to exploit information about the 3D structure of a face, such as the 3D contour and curvature of the forehead, eye sockets, cheeks, jaw, and chin [42]. An example of 3D face images is shown in Figure 1.9.

Figure 1.9: Face data from [1]: (a) two-dimensional intensity image, (b) a 2.5-D range image, and (c) a 3-D mesh (courtesy of [1]).

One advantage of 3D facial recognition is that it is not affected by changes in lighting like other techniques. It can also identify a face from a range of viewing angles, including a profile view [42, 43]. 3D data points from a face vastly improve the precision of facial recognition. To acquire 3D face data, one needs a stereo camera system, a structured light sensor, or a laser range scanner. The output is either range images or 3D polygonal meshes (or clouds). 3D research is enhanced by the development of sophisticated sensors that do a better job of capturing 3D face imagery. The sensors work by projecting structured light onto the face. Up to a dozen or more of these image sensors can be placed on the same CMOS chip – each sensor captures a different part of the spectrum [44].

With all the advantages mentioned above, 3D face recognition nonetheless has the following drawbacks or issues: (a) The complexity and computational cost is relatively high [45]; (b) Capture devices are more expensive. Compared with 2D cameras, existing 3D sensor technologies are still immature for practical applications. Existing problems include noise and artifacts, small depth of field, long acquisition time, and issues related to eye safety, points sampling, and accuracy; (c) Expression variation is still an issue. Even a perfect 3D

matching technique could be sensitive to expressions [27].

1.5   Dissertation Outline

This dissertation is dedicated to the topic of cross-spectral face recognition. It focuses on matching face images acquired in the electromagnetic wave spectrum of infrared against face images acquired in visible light.

Traditional face recognition systems are designed to process short range images acquired by a color camera. In this scenario (e.g. e-passport, and access control.) state-of-the-art face recognition algorithms are able to achieve high recognition rates (e.g. $> 99\%$). However, in a surveillance scenario where the acquisition is faced with difficult environments or has to be done at night time, the recognition task becomes very challenging. The conventional recognition algorithms are unable to perform. To improve the recognition rates of face algorithms operating in difficult settings, other imaging modalities operating within a different range of the electromagnetic spectrum (such as NIR, SWIR, MWIR and LWIR) have been introduced. In such applications, the cross-spectral recognition task can be very challenging due to the very distinct properties of the different imaging modalities.

This dissertation comprises seven chapters. Chapter 1 provides an introduction to the problem of face recognition. It reviews the history of face recognition, describes the typical structure of a face recognition system, explains the evaluation method for recognition performance and introduces new research topics in the area of face recognition (this dissertation falls into the scope of one of the new topics). Then it gives the outlines of the dissertation and summarizes the contributions of this work.

Chapter 2 discusses the problem of cross-spectral face recognition. It gives a detailed literature review of related research work on this topic, describes the framework for the cross-spectral face recognition system used throughout the dissertation, and finally compares and studies the performance of several baseline algorithms and several advanced algorithms for feature extraction.

In Chapter 3, we discuss three newly proposed methods for feature extraction in cross-spectral face recognition, namely Composite Multi-Lobe Descriptor (CMLD), Gabor-WLD-

LBP-HOG (GWLH) and Gabor Multi-Levels of Measurement (GMLM). The methods are evaluated on the Tactical Imager for Night/Day Extended-Range Surveillance (TINDERS) and Pre-Tactical Imager for Night/Day Extended-Range Surveillance (Pre-TINDERS) datasets.

Chapter 4 is devoted to a discussion of the problem of cross-spectral periocular recognition. The definition and reviews of this problem are given. All the methods talked about in previous chapters are compared.

Chapter 5 raises the problem of partial face recognition. Efforts are made to first define this problem and later identify the regions in a face image, which are the most informative for the purpose of cross-spectral face recognition. Methods from previous chapters are employed to find out the answer.

In Chapter 6 a technique called image quality parity is proposed for cross-spectral face recognition when there is a quality disparity between probes and a gallery. A quality measure named Adaptive Sharpness Measure is utilized. Quality parity by blurring using a Gaussian kernel function and enhancement with denoising and sharpening is proposed. An overview of related research work on image quality for biometric systems is also provided.

Chapter 7 summarizes the results developed in the previous chapters and concludes with a short description of items proposed as the future work.

1.6   Summary of Contributions

The main contributions of this dissertation are as follows:

1. A thorough comparative study on cross-spectral face recognition using some of the currently available or well-known operators has been conducted. We provided an overview of recent advances in the field of heterogeneous face recognition, emphasizing local operators developed for matching IR face probes to a gallery composed of high quality visible face images. A brief description of each individual and composite operators (10 in total) was provided. The list of individual operators included Gabor filters, LBP, GLBP, WLD and HOG. Composite operators include Gabor+LBP, Gabor+GLBP, Gabor+WLD, GOM, and Gabor+LBP+GLBP+WLD.

2. Three novel operators/encoding algorithms are designed for cross-spectral recognition between IR (NIR, SWIR, MWIR or LWIR) and visible spectral bands. The methods are demonstrated on two datasets each composed of data from 48 subjects: Pre-TINDERS and TINDERS datasets. Pre-TINDERS involves images acquired at a short distance (1.5 meters). TINDERS dataset includes long range images captured at two different operational distances (50 and 106 meters). The algorithms are shown to outperform or be comparable to the other basic and advanced algorithms currently available in the literature, in terms of performance.

3. The dissertation raises a new research topic – periocular recognition in the cross-spectral context. It reviews recent research work on periocular recognition as well as discusses the advantage of using periocular recognition as a new modality over face recognition. It then addresses the new problem by utilizing our proposed operators as the tool for feature extraction and compare them with other baseline algorithms. It further presents the results of matching SWIR, NIR and MWIR periocular probes to a gallery of visible periocular images. Both short (1.5 m) and long (50 m and 106 m) standoff distances were considered.

4. The new problem of cross-spectral partial face recognition is studied. We presented the results of partial face matching with probes being SWIR, NIR, and MWIR data and the gallery composed of visible face images. The heterogeneous face was partitioned into three non-overlapping regions in two different ways: the horizontal way and the characteristic way. In the first way we conducted two experiments – covering two out of three regions and sequential covering of the face, to find out which region is the most informative in terms of matching performance. In the second way we conducted an experiment to find out the most informative region by comparing the performance of the facial regions.

5. The dissertation addresses the problem of image quality disparity by proposing two approaches: blurring of the high-quality visible light images and enhancement of the low-quality infrared images. Both approaches are shown to be beneficial for cross-

spectral face recognition in presence of image quality disparity. An overview of related research work on image quality for biometric systems is also provided.

Chapter 2

Cross-Spectral Face Recognition

This chapter provides a comparative study of local operators recently proposed for heterogeneous face recognition. It also analyzes performance of each individual operator and demonstrates performance of composite operators. Basic local operators include Local Binary Patterns (LBP), Generalized Local Binary Patterns (GLBP), Weber Local Descriptors (WLD), Gabor filters, and Histograms of Gradients (HOG). They are directly applied to normalized face images. The composite operators dicussed are Gabor filters followed by LBP (Gabor+LBP), Gabor filters followed by WLD (Gabor+WLD), Gabor filters followed by GLBP (Gabor+GLBP), Gabor filters followed by LBP, GLBP and WLD (Gabor+WLD+LBP+GLBP), and Gabor Ordinal Measures (GOM).

When applying a composite operator to face images, images are first normalized and processed with a bank of Gabor filters and then local operators or a combination of local operators are applied to the outputs of Gabor filters. After a face image is encoded using the local operators, the outputs of local operators are converted to a histogram representation and then concatenated, resulting in a very long feature vector. Each component in the feature vector appears to contribute a small amount of information needed to generate a high fidelity matching score. A matching score is generated by means of the Kullback-Leibler distance between two feature vectors. The cross matching performance of heterogeneous face images is demonstrated on two data sets composed of active infrared and visible light face images. Both short and long standoff distances are considered.

2.1 Introduction

2.1.1 Topic Overview

Face recognition has been an active area of research over the past few decades. Many major advances have been reported in the literature. New applications have triggered new challenges, and new challenges have called for new research solutions. Surveillance at night or in harsh environments [46–48] is one of the most recent applications of face recognition. Latest advancements in manufacturing of small and cheap imaging devices sensitive in active infrared range (near- and short- infrared) [30, 31, 49] and the ability of these cameras to see through fog, rain, at night and operate at long ranges provided researchers with new type of imagery and posed new research problems [32–40]. As observed, active IR (ie. NIR and SWIR. See Table 2.1 for a detailed definition) energy is less affected by scattering and absorption by smoke or dust than visible light. SWIR cameras produce high signal-to-noise ratio (SNR) images under low light conditions or at night time. Also, SWIR light beams are not visible for the human eye which makes ongoing survailance unnoticeable during night time [31]. Furthermore, unlike visible spectrum imaging, active IR imaging can be used to extract not only exterior but also useful subcutaneous anatomical information [50]. Another advantage of the use of IR for face recognition over visible light is that thermal IR imaging is apparently invariant to changing illumination since the human face emits thermal energy, not reflected incident light. Therefore, changes in illumination appear to play less of a role in thermal infrared images, as opposed to the problem of illumination with visible light images, which is one of the most challenging problems to solve (See Chapter 1 Section 1.4).

All these advantages result in a very different appearance of face images in active and thermal IR range compared to face images in visible spectrum. Acknowledging these differences, many related questions can be posed. What type of information should be extracted from active and thermal IR images to successfully solve the problem of face recognition? How to match a face image in visible range to a face image in active IR and thermal range? The latter falls in the scope of heterogeneous face recognition. In the next subsection, we will first provide an overview of two existing general approaches to solve the problem of face recognition and later conduct a thorough review of recent research work related to ours and

discuss how the described approaches are utilized in these research work.

Table 2.1: Sub-division of the infrared band

Sub-Division Name	Abbreviation	Frequency	Wavelength	Photon Energy
Near-infrared	NIR	214–400 THz	0.75–1.4 µm	886–1653 meV
Short-wave infrared	SWIR	100–214 THz	1.4–3 µm	413–886 meV
Mid-wave infrared	MWIR	37–100 THz	3–8 µm	155–413 meV
Long-wave infrared	LWIR	20–37 THz	8–15 µm	83–155 meV
Far-infrared	FIR	0.3–20 THz	15–1,000 µm	1.2–83 meV

2.1.2   Review of Related Research

The literature identified two general categories of approaches to address the problem of face recognition: the holistic approach (also known as subspace analysis) and the local feature approach. The former represents the global photometric information of a human face using subspace projections. Examples include Principal Component Analysis (PCA), Independent Component Analysis (ICA), Linear Discriminant Analysis (LDA), Canonical-Correlation Analysis (CCA), Multilinear Subspace Learning (MSL) and their derivatives. Sirovich and Kirby [11] showed that PCA could be applied to a collection of face images to form a set of basis features which are known as eigenfaces. Later, Turk and Pentland [12, 51] expanded these results and presented the method of eigenfaces as well as a system for automated face recognition using eigenfaces. They showed a way of calculating the eigenvectors of a covariance matrix in such a way that makes it possible for computers at that time to perform eigen-decomposition on a large number of face images. Jutten and Herault [52] introduced the general framework for ICA and then Comon [53] refined it. ICA can be seen as a generalization of PCA, in which ICA generates a set of basis vectors that possess maximal statistical independence while PCA uses eigenvectors to determine basis vectors that capture maximal image variance. Motivated by the fact that much of the important information may be contained in the high-order relationship rather than that of the second-order, Bartlett at el. [54, 55] applied ICA to the problem of face recognition.

Fisher was the first to introduce the idea of LDA [13]. LDA determines a set of optimal discriminant basis vectors so that the ratio of the inter- and intra-class scatter matrices is maximized. It is primarily used to reduce the number of features to a more manageable number before classification. Each of the new dimensions is a linear combination of pixel values, which form a template. CCA was first introduced by Hotelling in 1936 [56]. Given two random vectors $X = (X_1, ..., X_n)$ and $Y = (Y_1, ..., Y_m)$, and assuming a correlation among the variables, CCA finds the linear combinations of X_i and Y_j that results in the maximum correlation with each other. Melzera et al. [57] applied CCA to face recognition and proposed appearance models based on kernel canonical correlation analysis.

The second category of approaches use local operators instead and have advantages such as more robustness to illumination and occlusion, less strictly controlled conditions, and involvement of very small training sets. Examples of operators used in this category include Gabor filters, Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), Weber Local Descriptor (WLD) and their generalizations and variants. Gabor filter is known to be a robust directional filter used for edge detection [17]. It has been found that simple cells in the visual cortex of mammalian brains can be modeled by Gabor functions [58,59]. A set of Gabor filters parameterized by different frequencies and orientations are shown to perform well as an image feature extraction tool. Therefore it has been widely used in image processing and pattern analysis applications [60–63]. LBP is a particular case of the texture spectrum model proposed by Wang et.al [64]. It was first introduced by Ojala and Pietikäinen [18,65] for texture classification and found to be a powerful tool. LBP was thereafter applied to face recognition as well as object detection [66, 67]. Due to its discriminative power and computational simplicity as well as robustness to monotonic changes of image intensity caused by illumination variations, LBP has been expanded into several variant forms (see for example, [68, 69]). HOG analysis was introduced by Dalal et al. [70] and was initially used for the purpose of object detection. This operator is similar to other operators such as edge orientation histograms and scale-invariant feature transform, but differs in that it is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy. Chen et. al [71] introduced the WLD operator inspired by Weber's law - an important psychological law quantifying the perception of change in a

given stimulus [72].

Most of the described methods above have been developed for intra-spectral matching, to be more specific to match visible light images. Some operators were tuned to work with heterogeneous face images. In the work of Chen et al. [73] a study of face recognition in thermal IR and visible spectral bands is performed, using PCA [10] and Faceit G5. It is shown that the performance of PCA in the visible light band is higher compared to the performance of PCA in the thermal IR bands, and that these data fused at the matching score level result in performance similar to the performance of the algorithm in visible band.

Pan et al. [74] demonstrate effectiveness of a hyperspectral approach within the NIR spectral band. They collect a dataset of face images acquired at 31 narrow spectral bands in the range 0.7-1.0 . Then a spectral reflectance vector evaluated in few face (square) regions at the different wavelengths is employed as the feature vector. The experiments performed on a hyperspectral dataset show that the adopted features are robust for recognition of individuals having different poses and facial expression.

In their work, Kong et al. [75] perform fusion of NIR and thermal IR face images in the Discrete Wavelet Transform domain employing images from the NIST/Equinox and the UTKIRIS [76] databases. They show that, when the fused images are fed to the Faceit recognition software, the resulting matching performance improves with respect to the case when the same face classes are compared within the same spectral band, NIR or thermal IR in this case.

Li et al. [77] propose a method to compare face images within the NIR spectral band with different illumination scenarios. Their face matcher is based on LBP operator to achieve illumination invariance and is applied to NIR images acquired at a short distance (less than one meter).

In their recent works Akhloufi and Bendada experimented with images from Equinox Database [49] (it includes visible, SWIR, MWIR, and LWIR images) and Laval University Multispectral Database [78] (includes visible, NIR,MWIR, LWIR data). The first work [79] evaluates recognition performance within each spectral band by using a set of known face matching techniques. In the second work (performed on the same data) [80] a classic Local Ternary Pattern (LTP) and a new Local Adaptive Ternary Pattern (LATP) operators are

adopted to extract features from images. The authors involve multi-resolution analysis in the "texture space" to fuse images from different spectral bands. They report that the fusion of face images from different spectral bands leads to improved recognition rates with respect to the case when images are matched within the same spectral band.

Lin et al. [81] introduce a Common Discriminant Feature Extraction (CDFE) method that brings images from different modalities (visible light – optical images and photographs, NIR and sketches) in a common feature space. It is shown that the proposed algorithm outperforms PCA, LDA (Linear Discriminant Analysis), kernel PCA and kernel LDA [82] in the visible (optical) versus NIR comparison and also when visible (photo) images are matched against sketches.

Liao et al. [83] apply a Multiscale Block Local Binary Patterns (MB-LBP) descriptor to NIR and visible face images. Both types of images are preprocessed with Difference of Gaussian (DOG) filters and then encoded with the MB-LBP operator. A Gentle AdaBoost is applied to select features, and a regularized LDA method is used to match processed data. The method is tested on a multispectral dataset of 52 face classes. The implemented approach is shown to outperform CDFE, PCA-CCA and LDA-CCA [84] methods when visible images are matched against NIR images.

In their paper Yi et al. [85] encode images captured at NIR and visible spectrum by adopting a Laplacian of Gaussian (LoG) filter. The filtered images are further converted into binary images that are locally partitioned in small patches. The method compares common patches (partial faces) in visible and NIR images. The experiments are performed on MBGC data [86, 87]. The proposed method is compared to CDFE, PCA-CCA and LDA-CCA.

The work of Klare and Jain [88] employs a method based on LBP and Histogram of Gradient (HOG) features, followed by a random sampling LDA algorithm to reduce the dimensionality of feature vectors. This encoding strategy is applied to NIR and color images for their cross-spectral matching. The results are shown to outperform Cognitec's FaceVACS [89].

Maeng et al. [90] are the first to report the results of long range cross spectral face matching, where long range NIR images are matched against visible face images. The paper introduces a new long range NIR database called Near-Infrared Face Recognition at a

Distance Database (NFRAD-DB). Face recognition performance was evaluated using Face-VACS, DoG-SIFT, and DoG-MLBP methods. The experiments involve 50 long range NIR classes and more than 10,000 short range visible face images. The achieved rank 1 recognition performance is 28 percent, which is a promising result for long range cross spectral face recognition.

This chapter focuses on a discussion of local operators (algorithms from the second category) for heterogeneous face recognition. The methodology for feature extraction and heterogeneous matching adopted in this chapter does not require training data, which justifies its importance in practice. Once local operators are developed, they can be applied to any heterogeneous data (we particularly focus on matching visible images to active IR images) and do not require any estimation or learning of parameters or retraining of the overall face recognition system.

We present and compare several feature extraction approaches applied to heterogeneous face images. Face images (in visible spectrum and active IR) may be first processed with a bank of Gabor filters parameterized by orientation and scale parameters followed by an application of a bank of local operators. The operators encode both the magnitude and phase of filtered (or non-filtered) face images. The application of an operator to a single image results in multiple magnitude and phase outputs. The outputs are mapped into a histogram representation, which constitutes a long feature vector. Feature vectors are cross-matched by applying a symmetric Kullback-Leibler distance. The combination of Gabor filters and local operators offers an advantage of both the selective nature of Gabor filters and the robustness of these operators.

In addition to known local operators such as LBP, generalized LBP (GLBP), WLD and HOG, we also present a recently developed operator named Gabor Ordinal Measures by Chai et al. [91]. Performance of Gabor filters, LBP, GLBP, WLD, and HOG used both individually and in combinations are demonstrated on both the Pre-TINDERS and TINDERS datasets [92]. These datasets contain color face images, NIR and SWIR face images acquired at a distance of 1.5, 50,and 106 meters.

2.2   Recognition System Framework

As discussed in Section 1.2 from Chapter 1, a typical face recognition (the same as heterogeneous face recognition) system can be described by five modules: image acquisition, face detection, preprocessing, feature extraction and matching. Since our work focuses on the latter three modules, we omit the former two in the following chapters hereafter. A simplified version of such a heterogeneous face recognition system is shown in the block-diagram in Figure 2.1. In this work, the preprocessing module implements an alignment, cropping and normalization of heterogeneous face images. The feature extraction module performs filtering, applies local operators and represents the outputs of local operators in a form of histograms. The matching module applies a symmetric Kullback-Leibler distance to histogram representations of heterogeneous face images to generate a matching score. A functional description of each of the three modules is provided in the following subsections.

```
PREPROCESSING  →  FEATURE EXTRACTION  →  MATCHING
```

Figure 2.1: A block diagram of a typical face recognition system.

2.2.1   Preprocessing

In this work, the preprocessing module implements image alignment, cropping, and normalization. For alignment, positions of the eyes are used to transform the face to a canonical representation. Geometric transformations such as rotation, scaling and translation are applied to each face image with the objective to project eyes to a fixed position. Figure 2.2 (a), (b) and (d) illustrate the processing steps. In our work, the anchor points - the fixed positions of the eyes - are manually selected. However, this process can be automated by means of a Haar-based detector trained on heterogeneous face images [93], as an example.

The aligned face images are further cropped to an area of size 120×112 (see Figure 2.2 (b) and (d)). After being cropped, images undergo an intensity normalization. Color images are converted to grayscale images using a simple linear combination of the original R, G and

B channels (see Figure 2.2 (c)). Active IR images - SWIR and NIR images - are preprocessed using a simple nonlinear transformation given by $\log(1 + X)$, where X is the input image, as shown in Figure 2.2 (e). The log-transformation redistributes the original darker pixels over a much broader range and compresses the range of the original brighter pixels. The transformed image is brighter and has a better contrast than the original image while the gray variation (trend) of the pixels is still preserved since the transformation is monotonic.

(a)   (b)   (c)   (d)   (e)
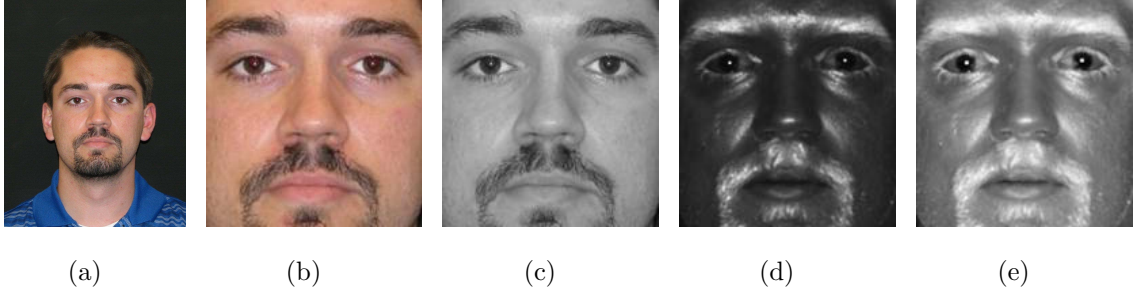
Figure 2.2: Preprocessing of the face: (a) original color image, (b) aligned and cropped color face, (c) gray scale conversion of (b), (d) aligned and cropped SWIR face, and (e) log-tranformation of (d).

2.2.2   Feature Extraction

Gabor Filter

As recently demonstrated by Tan et al. [94], Xie et al. [95], and Nicolo et al. [37,38], a two-step encoding of face images, where encoding with local operators is preceded by Gabor filtering, leads to considerably improved recognition rates. Therefore, many combinations of operators analyzed in this chapter, involve filtering with a bank of Gabor filters as a first step. The filter bank includes 2 different scales and 8 orientations resulting in a total of 16 filter responses. The mathematical description of the filter is given as:

$$G(\mathbf{z}, \theta, s) = \frac{\|\mathcal{K}(\theta, s)^2\|}{\sigma^2} \exp\left[-\frac{\|\mathcal{K}(\theta, s)\|^2 \|\mathbf{z}\|^2}{2\sigma^2}\right] \left[e^{i\mathcal{K}^T(\theta, s)\mathbf{z}} - e^{-\frac{\sigma^2}{2}}\right], \qquad (2.1)$$

where $\mathcal{K}(\theta, s)$ is the wave vector and σ^2 is the variance of the Gaussian kernel. The magnitude and phase of the wave vector determine the scale and orientation of the oscillatory term and $z = (x, y)$ is a pixel in an input image. The wave vector can be expressed as

$$\mathcal{K}(\theta, s) = \mathcal{K}_s(cos\phi_\theta, sin\phi_\theta), \qquad (2.2)$$

where $\mathcal{K}_s$ is known as a scale parameter and ϕ_θ is an orientation parameter. The adopted parameters for the complex vector in the experiments of this chapter are set to $\mathcal{K}_s = (\pi/2)^{s/2}$ with $s \in \mathbb{N}$ and $\phi_\theta = \theta\pi/8$ with $\theta = 1, 2, ..., 8$. The Gaussian kernel has the standard deviation $\sigma = \pi$.

A normalized and preprocessed face image $I(z)$ is convolved with a Gabor filter $G(z, \theta, s)$ at orientation ϕ_θ and scale $\mathcal{K}_s$ resulting in a filtered image $Y(z, \theta, s) = I(z) * G(z, \theta, s)$, where $*$ stands for convolution.

Weber Local Descriptor

The WLD operator consists of two joint parts: a differential excitation operator and a gradient orientation descriptor. In this chapter we adopt only the differential excitation operator to encode the magnitude filter response, resulting in a robust representation of face features.

The differences between the neighboring pixels of a central pixel are calculated and normalized by the pixel value itself. The summation of these normalized differences is further normalized by a monotonic function such as a tangent function. Finally, quantization is performed to output the WLD value.

The mathematical definition of WLD used in this chapter is given as:

$$WLD_{l,r,N}(x) = \mathcal{Q}_l \left\{ \tan^{-1} \left[\sum_{i=1}^{N} \left(\frac{x_i - x}{x} \right) \right] \right\}, \tag{2.3}$$

where x_i are the neighbors of x at radius r and N is the total number of neighbors (see Figure 2.3). $\mathcal{Q}_l$ is a uniform quantizer with l quantization levels.
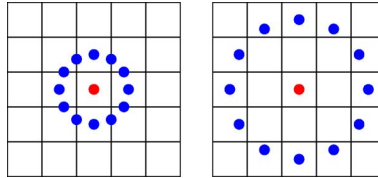
Figure 2.3: Illustration of the neigboring pixels (N=12) of a central pixel at different radii: the left corresponds r=1; the right r=2.

Local Binary Pattern

A uniform LBP operator is described as

$$LBP^{\mathcal{U}}_{r,N}(x) = \mathcal{U}\left\{\sum_{i=1}^{N} \mathcal{I}\{x_i - x\}2^i\right\}, \tag{2.4}$$

where x_i are the neighbors of the pixel x at radius r, N is the total number of neighbors. $\mathcal{U}$ is the uniform pattern mapping and $\mathcal{I}(\cdot)$ is the unit step function:

$$\mathcal{I}(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases} \tag{2.5}$$

A binary pattern is *uniform* if it contains at most two bit-wise transitions from 0 to 1 or from 1 to 0 when the bit sequence is recorded circularly. For example, the sequence 011111111000 is a 12-bit uniform pattern while the sequence 010001011111 is not uniform. The uniform mapping $\mathcal{U}(d)$ is defined as

$$\mathcal{U}(d) = \begin{cases} d, & \text{if } d_B \text{ is uniform} \\ M, & \text{otherwise} \end{cases} \tag{2.6}$$

where d_B is the binary form of a number d and M is the total number of uniform patterns formed using N bits. Throughout this chapter, we work with $N = 12$-bit sequences, which results in $M = 134$ uniform patterns.

Generalized Local Binary Pattern

A uniform GLBP operator is a generalization of the encoding method introduced in [96] by introducing a varying threshold t rather than a fixed one. Encoding of Gabor phase response by GLBP adds up to encoding of Gabor magnitude response using LBP and leads to improved performance compared to the performance of each of them [38]. The uniform generalized binary operator is defined as

$$GLBP^{\mathcal{U}}_{r,N,t}(x) = \mathcal{U}\left\{\sum_{i=1}^{N} \mathcal{T}_t\{x_i - x\}2^i\right\}, \tag{2.7}$$

where x_i is the i-th neighbor of x at radius r (we set $r = 1, 2$ in our experiments) and N is the total number of neighbors. $\mathcal{U}(\cdot)$ is the uniform pattern mapping described in the

previous subsection (see Sec. 2.2.2). $\mathcal{T}_t(\cdot)$ is a thresholding operator based on threshold t. It is defined as

$$\mathcal{T}_t(x) = \begin{cases} 1, & |x| \le t \\ 0, & |x| > t \end{cases} \tag{2.8}$$

The values for the thresholds in this chapter were evaluated experimentally and set to $t = \pi/2$.

Gabor+WLD+LBP+GLBP

A fusion of extracted features often leads to improved recognition performance. As shown in [38,97], LBP and WLD applied to the magnitude of Gabor filtered images combined with GLBP applied to the phase of Gabor filtered images yielded a significant performance boost. Details of this fusion scheme can be found in [38,97]. A block diagram of the fusion approach is displayed in Figure 2.4.
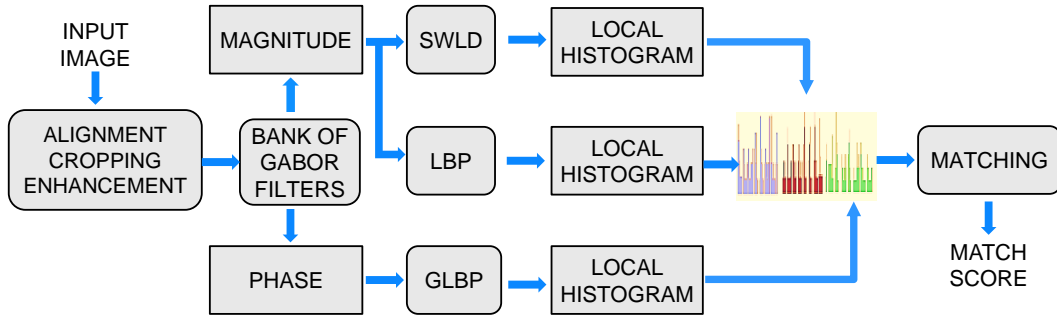
Figure 2.4: A block diagram of the fusion scheme of Gabor+WLD+LBP+GLBP.

Gabor Ordinal Measures

Gabor Ordinal Measures (GOM) is a recently developed local feature operator [91]. This operator combines Gabor filters (see Sec. 2.2.2) with ordinal measures, a measurement level which records the information about ordering of multiple quantities [98]. Following GOM, Chai et al. extracted a histogram representation and applied a dimensionality reduction by means of LDA to filtered and encoded face data.

The ordinal measure in [91] is modified using a smoothed neighborhood described by a Gaussian smoothing function. Therefore, the ordinal measure filter $f_{om}(\mathbf{z})$ can be expressed as:

$$
\begin{aligned}
f_{om}(\mathbf{z}) = &\, C_p \sum_{i=1}^{N_p} \frac{1}{\sqrt{2\pi}\sigma_{p,i}} \exp \left[\frac{-(\mathbf{z} - \mu_{p,i})^T (\mathbf{z} - \mu_{p,i})}{2\sigma_{p,i}^2} \right] \\
&- C_n \sum_{i=1}^{N_n} \frac{1}{\sqrt{2\pi}\sigma_{n,i}} \exp \left[\frac{-(\mathbf{z} - \mu_{n,i})^T (\mathbf{z} - \mu_{n,i})}{2\sigma_{n,i}^2} \right]
\end{aligned}
\tag{2.9}
$$

where $\mathbf{z} = (x, y)$ is the location of a pixel. $\mu_{p,i}$ and $\sigma_{p,i}$ denote the central position and the scale of the i-th positive lobe of a 2D Gaussian function, while $\mu_{n,i}$ and $\sigma_{n,i}$ denote that of the i-th negative lobe of the same Gaussian function. N_p and N_n are the numbers of positive and negative lobes, respectively, while constant coefficients C_p and C_n keep the balance between positive and negative lobes, i.e., $C_p N_p = C_n N_n$.

2.2.3   Matching

Each encoded response (the output of each local operator) is divided into 210 non-overlapping square blocks of size 8×8. Blocks are displayed in the form of histograms and the number of bins is chosen the same as the level of the encoders mentioned in the previous section (eg. 135 in our experiments). Then 135-bin histograms of all blocks are normalized and concatenated to be treated as a probability mass function, resulting in a vector of length $135 \times 210 = 28,350$ for each encoded response. The length of the feature vector was selected empirically to maximize the cross-matching performance. Vectors of all encoded responses will be further concatenated and thus the total size of a feature vector corresponding to an input face image is $28,350 \times P$, where P is the number of encoded responses, (E.g., $P = 96$ for both the case of combination of operators. See subsections 2.2.2).

When the distance between two feature vectors (histograms in our case) is evaluated, it is expressed as a sum of distances for all feature vector pairs. A sum of two Kullback-Leibler distances [99] is used as the distance metric to compare the feature vectors of heterogeneous images (refer to the work of Nicolo for comparison of using different distance metrics [97]). For two images A and B with the feature vectors H_A and H_B, respectively, the symmetric

Kullback-Leibler distance is defined as:

$$D_{KL}(A, B) = \sum_{k=1}^{K} (H_A(k) - H_B(k)) \log \frac{H_A(k)}{H_B(k)}, \tag{2.10}$$

where K is the length of the feature vectors H_A or H_B.

2.3   Datasets

In our experiments we use two datasets Pre-TINDERS (Tactical Imager for Night/-Day Extended-Range Surveillance) and TINDERS collected by the Advanced Technologies Group, West Virginia High Tech Consortium (WVHTC) Foundation [100]. A summary of the datasets can be found in Table 2.2.

Pre-TINDERS is composed of 48 frontal face classes of total 576 images acquired at three wavelengths – visible light, 980 nm NIR and 1550 nm SWIR. Images are acquired at a short standoff distance of 1.5 m in a single session. Four images per class are available in each spectral band. A 980 nm light source is used to illuminate the face in the NIR spectral band while a 1550 nm light source is used in the SWIR spectral band. The original resolutions of the acquired images (see Figure 2.5) are 640×512 (png format) for both NIR and SWIR images and 1600×1200 (jpg format) for color images.

TINDERS is composed of 48 frontal face classes each represented by visible, NIR (980 nm) at two standoff distances (50 and 106 m), and SWIR at two standoff distances (50 and 106 m) images. At each distance and spectrum, four or five images per class are available. A total of 478 images with the resolution 640×512 (png format) are available in SWIR band. A total of 489 images with the resolution 640×512 (png format) are available in the NIR band. The visible (color) images with the resolution 480×640 (jpg format) are collected at a short distance and in two sessions (3 images per session), and all of them have neutral expression, resulting in a total of 288 images. Sample images from the Pre-TINDERS and TINDERS datasets are shown in Figure 2.5.

It is important to note that although the original resolution of images in Pre-TINDERS and TINDERS is varying, we crop and normalize them to be the same size for each experiment described below. This is done to ensure a fair comparison.
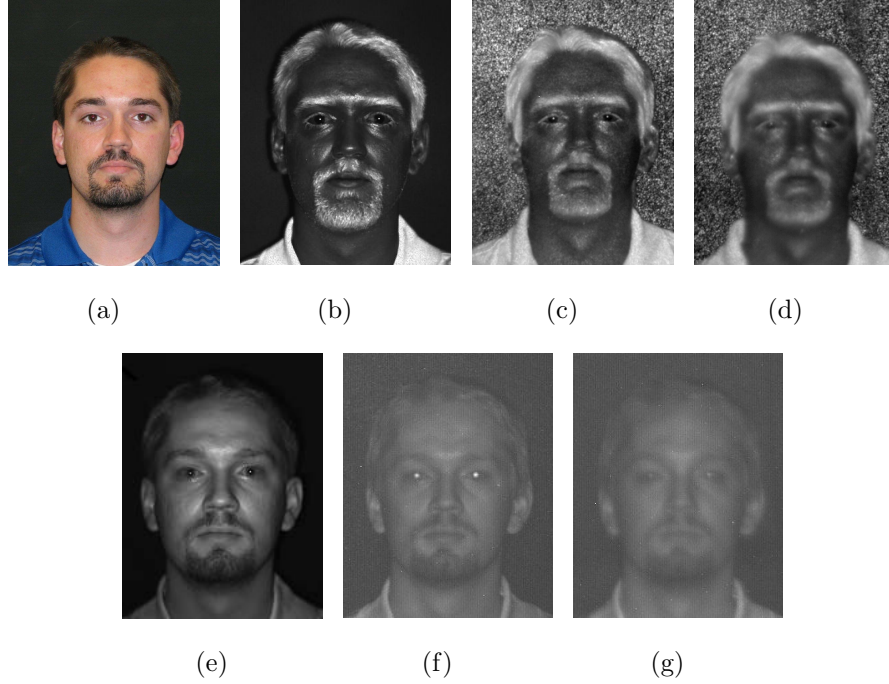
Figure 2.5: Sample images: (a) visible, (b) SWIR at 1.5 m, (c) SWIR at 50 m, (d) SWIR at 106 m, (e) NIR at 1.5 m, (f) NIR at 50 m, and (g) NIR at 106 m.

2.4   Performance Evaluation

In this section, we analyze the performance of various local operators used for encoding heterogeneous face images. In our experiments, galleries are composed of visible light face images, while NIR and SWIR face images are presented as probes. We match NIR and SWIR face images collected at 1.5, 50, and 106 m to visible light face images acquired at a distance 1.5 m.

For both SWIR and NIR spectra (at both short and long standoff distances), a total of 10 operators (including individual operators and their combinations) are implemented. We order and number them as: (1) LBP, (2) WLD, (3) GLBP, (4) HOG, (5) Gabor filter, (6) Gabor filter followed by LBP applied to the magnitude image (Gabor+LBP), (7) Gabor filter followed by WLD applied to the magnitude image (Gabor+WLD), (8) Gabor filter followed by GLBP applied to the phase image (Gabor+GLBP), (9) Gabor filter followed by LBP, GLBP and WLD (Gabor+LBP+GLBP+WLD) and (10) GOM. The parameters in the experiments are chosen as follows. The number of orientations and radii for Gabor filters

Table 2.2: Summary of the datasets

Dataset	Class	Total # Images	Spectrum	Distance	Original Resolution
Pre-TINDERS	48	576	visible NIR SWIR	1.5 m	visible: 1600 × 1200 NIR: 640 × 512 SWIR: 640 × 512
TINDERS	48	1255	visible NIR SWIR	visible: 1.5 m NIR & SWIR: 50 m and 106 m	visible: 640 × 480 NIR: 640 × 512 SWIR: 640 × 512

are set to 8 and 2, respectively. The number of radii for LBP, GLBP, and WLD is chosen as 2, and the number of neighbors around the central pixel is set to 12. The same parameters are used in operators to encode short and long range images.

The results of matching are displayed in the form of Receiver Operating Characteristic (ROC) curves. We plot Genuine Accept Rate (GAR) versus False Accept Rate (FAR). Summaries of Equal Error Rates (EER), d-prime values, and GARs at the FAR set to 0.1 and 0.001 are provided in tables.

2.4.1 Matching SWIR Probes against Visible Gallery

Our first experiment involves matching SWIR face images to visible face images. The heterogeneous images are encoded using the ten individual or composite operators as described earlier in this section. The performance of the individual encoders can be treated as benchmarks. The results of matching parameterized by different standoff distances are shown in Figures 2.6, 2.7 and 2.8. In these experiments, visible light images form the gallery set. All SWIR images are used as probes.

Short Standoff Distance

For the case of the short standoff distance (the Pre-TINDERS dataset), the performance of single operators such as HOG, LBP, WLD, GLBP and Gabor filters is inferior to the performance of the composite operators where Gabor filters are followed by LBP, WLD,
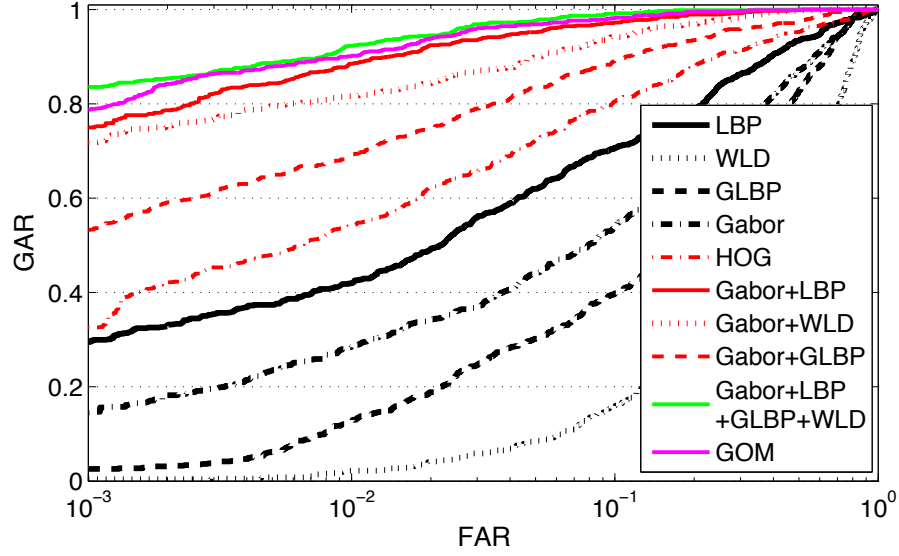
Figure 2.6: ROC curves: matching SWIR probes at 1.5 m to visible gallery.
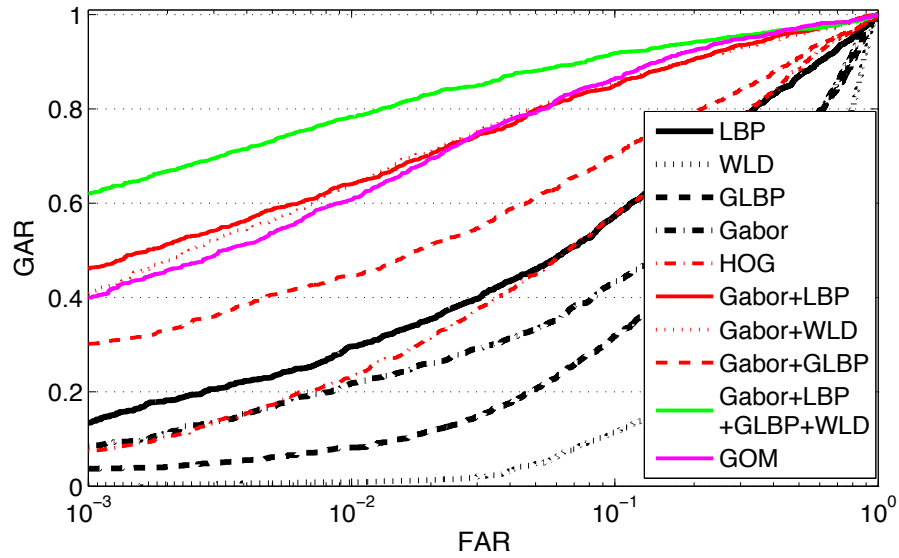
Figure 2.7: ROC curves: matching SWIR probes at 50 m to visible gallery.

and GLBP. It is also inferior to the performance of Gabor+LBP+GLBP+WLD and GOM, the other two composite multi-lobe operators. Within the group of single operators, HOG outperforms the other four operators closely followed by LBP and then Gabor filters. WLD appears to be less suitable for encoding heterogeneous face images in the framework of the cross-spectral matching.
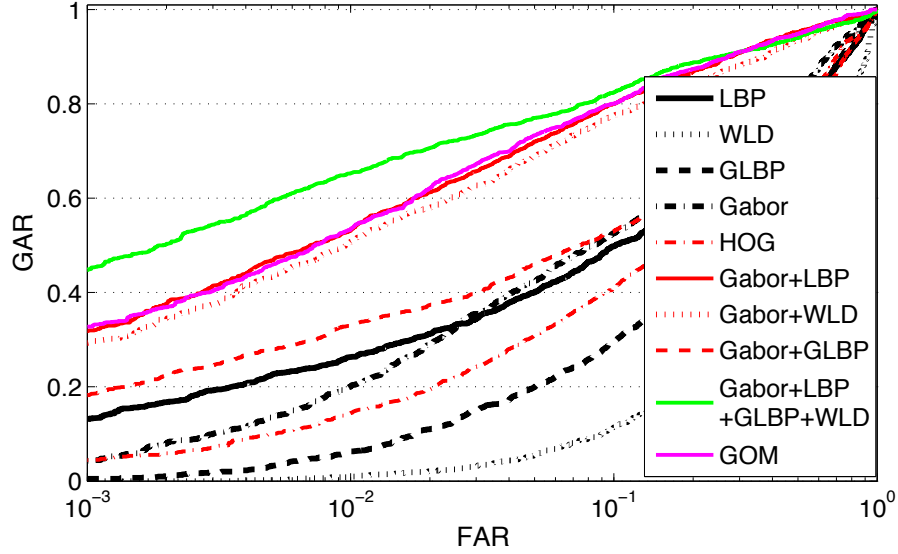
Figure 2.8: ROC curves: matching SWIR probes at 106 m to visible gallery.

Within the group of composite operators, the top four, following closely together, are Gabor+LBP+GLBP+WLD, GOM, Gabor+LBP, and Gabor+WLD. Gabor+GLBP performs slightly inferior to the top four. Table 2.3 presents a summary of EERs, d-prime values and GAR values at FAR set to 0.1 and 0.001 values.

Long Standoff Distance

SWIR images at longer standoff distances (50 and 106 m in the case of TINDERS dataset) experience some loss of quality due to air turbulence, insufficient illumination, and optical effects during data acquisition. This immediately reflects on the values of matching scores. Figure 2.7 and Figure 2.8 display the results of cross-spectral comparison parameterized by 50 m and 106 m standoff distances, respectively. Gallery images are retained from the previous session. Note that in both figures, Gabor+LBP, Gabor+WLD and GOM display a very similar performance. They are closely followed by Gabor+GLBP. The top performance in both cases is demonstrated by Gabor+LBP+GLBP+WLD. Once again, composite operators outperform single operators, which was anticipated. However, at longer standoff distances matching performance of all the operators and their combinations but Gabor+LBP+GLBP+WLD drops nearly two times for the case of 50 m and 2.5 times for the

Table 2.3: EERs and GAR values: matching SWIR probes at 1.5 m to visible gallery.

Method	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER(%)	d-prime
WLD	15.89	0.39	41.40	0.40
LBP	70.70	29.56	20.61	1.66
GLBP	39.71	2.60	33.46	0.98
Gabor	54.04	14.71	27.35	1.24
HOG	80.47	32.55	15.36	1.86
Gabor+WLD	94.14	71.88	7.68	2.74
Gabor+LBP	97.27	75	4.82	3.09
Gabor+GLBP	89.19	53.39	10.68	2.35
GOM	98.18	78.78	3.64	3.18
Gabor+WLD+LBP+GLBP	**99.09**	**83.59**	**3.13**	**3.2**4

Table 2.4: EERs and GAR values: matching SWIR probes at 50 m to visible gallery.

Method	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER(%)	d-prime
WLD	11.55	0.21	49.59	0.045
LBP	57.29	13.45	25.28	1.24
GLBP	31.86	3.71	37.07	0.65
Gabor	43.35	8.33	34.97	0.82
HOG	57.42	7.56	25.42	1.25
Gabor+WLD	85.57	40.90	12.74	2.19
Gabor+LBP	85.01	46.15	12.89	2.25
Gabor+GLBP	70.10	30.18	20.51	1.56
GOM	86.41	39.98	11.97	2.27
Gabor+WLD+LBP+GLBP	**91.88**	**62.11**	**8.90**	**2.57**

case of 106 m. EERs, d-prime values and GARs at FAR set to 0.1 and 0.001 are summarized in Tables 2.4 and 2.5.

Table 2.5: EERs and GAR values: matching SWIR probes at 106 m to visible gallery.

Method	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER(%)	d-prime
WLD	11.39	0.10	48.69	0.0038
LBP	49.79	13.19	31.11	0.94
GLBP	29.31	0.49	36.49	0.57
Gabor	52.57	4.31	28.67	1.09
HOG	41.04	4.44	33.68	0.78
Gabor+WLD	77.57	29.31	16.96	1.83
Gabor+LBP	80.00	31.81	15.83	1.99
Gabor+GLBP	53.06	18.19	32.65	0.88
GOM	80.07	32.78	14.78	2.02
Gabor+WLD+LBP+GLBP	**82.50**	**44.79**	**14.17**	**2.00**

2.4.2   Matching NIR Probes against Visible Gallery

In the second experiment, NIR face images (probes) are matched to short range visible face images (gallery). The results of matching parameterized by the standoff distances of 1.5 m, 50 m, and 106 m are shown in Figures 2.9, 2.10 and 2.11, respectively.

Short Standoff Distance

Among the group of single operators, LBP and HOG outperform the other operators, followed by GLBP and Gabor. Similar to the the case of SWIR probe images, WLD operator performs poorly. All composite operators demonstrate a relatively high performance with ROC curves closely following one another. Gabor+WLD+LBP+GLBP appears to outperform the other four composite operators. It is closely followed by GOM and then by Gabor+WLD, Gabor+LBP and Gabor+GLBP. Table 2.6 summarizes the values of EERs, d-primes and GARs at FAR equal to 0.1 and 0.001.
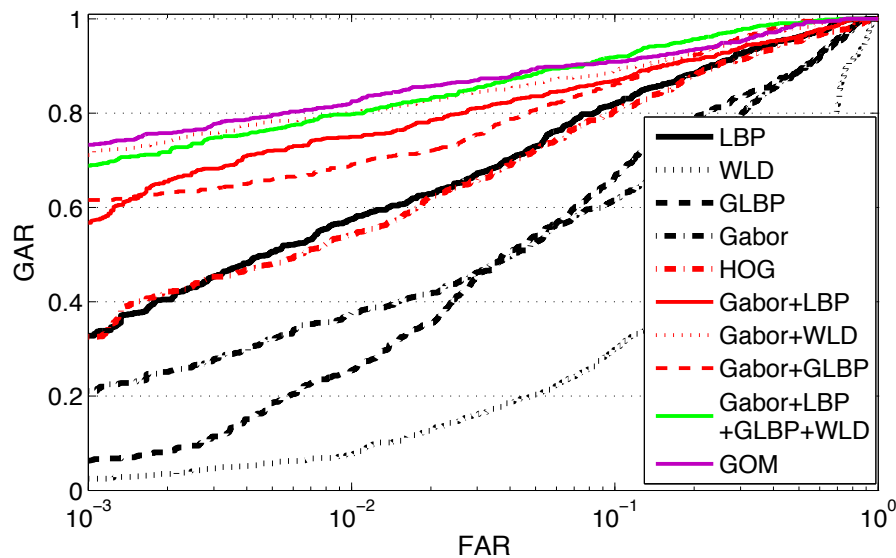
Figure 2.9: The results of cross matching short range (1.5 m) NIR probes and visible gallery images.
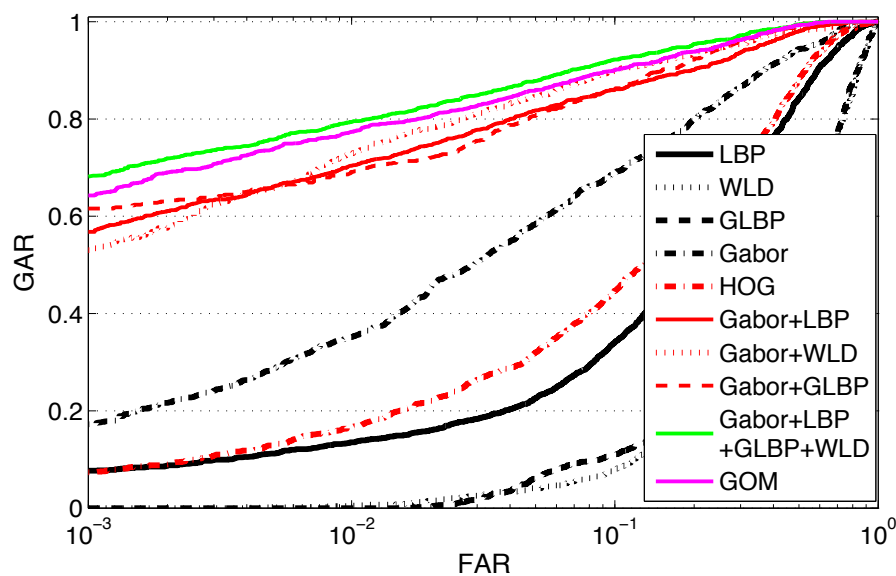
Figure 2.10: The results of cross matching long range (50 m) NIR probes and visible gallery images.
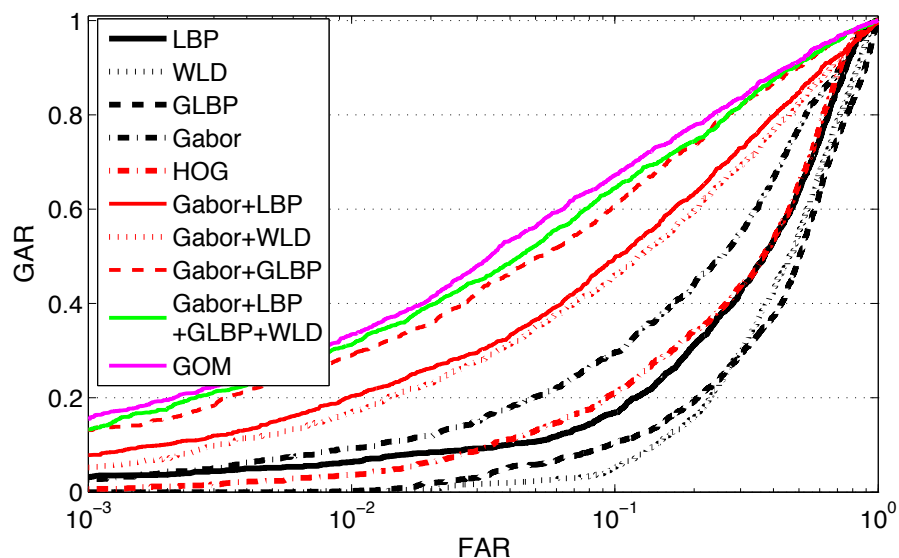
Figure 2.11: The results of cross matching short range (106 m) NIR probes and visible gallery images.

Table 2.6: EERs and GAR values: matching NIR probes at 1.5 m to visible gallery.

Method	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER(%)	d-prime
WLD	29.82	2.47	44.27	0.44
LBP	82.03	32.81	14.36	2.12
GLBP	66.54	6.38	20.57	1.467
Gabor	61.46	21.09	23.57	1.43
HOG	65.23	23.96	22.03	1.68
Gabor+WLD	89.19	71.098	10.54	2.38
Gabor+LBP	86.98	56.77	11.82	2.29
Gabor+GLBP	86.20	61.595	12.23	2.29
GOM	90.89	73.31	9.27	**2.59**
Gabor+WLD+LBP+GLBP	**91.93**	**68.88**	**8.73**	2.48

Long Standoff Distance

Long range NIR probes display a cardinally different performance. As can be seen from Figure 2.5, NIR images at 106 m have much lower contrast and overall quality compared to NIR images at 50 m. This difference in image quality immediately reflects on the matching

Table 2.7: EERs and GAR values: matching NIR probes at 50 m to visible gallery.

Method	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER(%)	d-prime
WLD	5.39	0.1	48.12	0.072
LBP	34.17	7.70	31.74	0.99
GLBP	11.06	0.1	49.01	0.12
Gabor	68.98	17.23	19.98	1.66
HOG	44.68	7.35	29.98	1.16
Gabor+WLD	89.85	53.011	10.07	2.40
Gabor+LBP	86.13	56.79	12.54	2.33
Gabor+GLBP	92.02	69.89	8.66	2.73
GOM	90.06	64.29	10.00	2.65
Gabor+WLD+LBP+GLBP	**92.23**	**68.21**	**8.71**	**2.66**

Table 2.8: EERs and GAR values: matching NIR probes at 106 m to visible gallery.

Method	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER(%)	d-prime
WLD	7.91	0.1	49.40	0.067
LBP	16.95	3.18	43.30	0.45
GLBP	10.52	0.1	50.21	0.038
Gabor	29.66	2.61	36.87	0.72
HOG	21.12	0.64	42.87	0.52
Gabor+WLD	45.97	5.23	30.48	1.05
Gabor+LBP	49.72	7.84	28.43	1.10
Gabor+GLBP	60.88	13.14	23.16	1.44
GOM	**67.30**	**15.53**	**21.65**	**1.58**
Gabor+WLD+LBP+GLBP	64.48	13.28	23.24	1.49

performance of the two sets of probes (50 m probes and 106 m probes). This also reflects on the interplay among 10 operators. Figure 2.10 and Figure 2.11 display the cross matching results for the two stand off distances (50 m and 106 m, respectively). Comparing the composite operators in terms of their performance, NIR at 50 m shows that Gabor+LBP+GLBP+WLD

and GOM perform equally well. Their performance is very close to the performance they demonstrate at 1.5 m. Note it is only slightly degraded. These three ROCs are closely followed by the ROCs of Gabor+GLBP and Gabor+WLD. At 106 m NIR probes do not perform as well. In fact, the performance of NIR images encoded with composite operators drops at least three times compared to the performance of the same operators applied to NIR at 50 m. Figure 2.11 indicates that GOM followed by Gabor+LBP+GLBP+WLD and Gabor+GLBP, where GLBP is applied to phase images, seem to be more robust to degraded image quality in NIR spectrum compared to other composite operators. Among single operators, Gabor and HOG still outperform other single operators for both standoff distances. Tables 2.7 and 2.8 present a summary of EERs, d-primes and GARs at FAR set to 0.1 and 0.001 for the case of 50 m and 106 m standoff distances, respectively.

In addition to matching performance, computation time is also evaluated and compared among all the operators that appear in this chapter. We evaluate both the encoding time and matching time for each operator. Both mean and standard deviation (std.) are calculated. The computation time is evaluated on a PC with an Intel Core i5 CPU at 3.2 GHz and a 8 GB RAM memory. An example of the computation time for the case of encoding and matching SWIR 1.5 m probes is provided in Table 2.9.

2.5 Summary

This chapter presented an overview of recent advances in the field of heterogeneous face recognition, emphasizing the topic of local operators developed for matching IR face probes to a gallery composed of high quality visible face images. A brief description of each individual and composite operator (10 in total) was provided. The list of individual operators included LBP, GLBP, WLD and Gabor filters. Composite operators included Gabor+LBP, Gabor+GLBP, Gabor+WLD, GOM, and Gabor+LBP+GLBP+WLD.

We considered a very specific framework for cross-matching heterogeneous face images, assuming that each image is aligned, cropped and enhanced at first. It was then filtered and encoded using local operators. The output images were converted into a histogram

Table 2.9: Comparison of computation time: Encoding and matching of SWIR probes at 1.5 m.

Method	Computation Time for Encoding (s)		Computation Time for Matching (s)	
	Mean	Std.	Mean	Std.
WLD	0.1591	0.0051	0.0129	0.0009
LBP	0.1649	0.0077	0.0135	0.0011
GLBP	0.1568	0.0047	0.0124	0.0011
Gabor	2.6248	0.0259	0.3202	0.0037
HOG	0.0340	0.0262	0.0043	0.0009
Gabor+WLD	2.9444	0.0780	0.1710	0.0046
Gabor+LBP	2.7791	0.0378	0.1610	0.0045
Gabor+GLBP	2.7733	0.0468	0.1452	0.0044
GOM	6.5721	0.1321	0.3048	0.0059
Gabor+WLD+LBP+GLBP	3.4120	0.0552	0.4735	0.0108

representation and compared against histogram representations of images in the gallery by means of a symmetric Kullback-Leibler distance. This cross-matching approach does not require any training or learning and it is shown to be robust when applied to a variety of heterogeneous datasets.

We presented the results of matching SWIR and NIR facial images to visible facial images. Both short (1.5 m) and long (50 m and 106 m) standoff distances were considered. The results were documented in figures and tables. We presented ROC curves as well as GARs at two specific levels of FAR, EERs and d-prime values. Conclusions from the experimental results are made as follows:

- The combination of Gabor filters followed by other local operators substantially out-performed the original LBP and the other individual operators;

- As the standoff distance increased, the matching performance of all the methods dropped. This drop was attributed to a relatively low quality of imagery at long standoff distances (SWIR vs. visible and NIR vs. visible).

Chapter 3

New Operators for Feature Extraction

The purpose of this chapter is to develop a family of powerful operators capable of improving the performance of cross-spectral face recognition systems (which can also be applied to partial face). We will describe three novel operators, namely Gabor+WLD+LBP+HOG (GWLH), Composite Multi-Lobe Descriptor (CMLD) and Gabor Multi-Levels of Measurement (GMLM). GWLH is a composite operator which combines individual operators LBP, Histograms of Gradients (HOG) and Weber Local Descriptor (WLD) preceded by Gabor filters. CMLD is realized by combing the multi-lobe forms of individual operators (such as LBP) which are modified to consider local information within a smoothed neighbourhood. GMLM is based on the concept of multiple *levels of measurement* [98]. It employs operations at different levels of measurement on a face image to extract different types of information and fuses them together to take advantage of the complementary information.

After feature extraction using one of the new operators, matching scores are generated by means of a Kullback-Leibler distance between two feature vectors. The cross matching performance of the three operators is demonstrated on three datasets composed of heterogeneous face images acquired in the NIR, SWIR, MWIR, LWIR and visible light spectra. To demonstrate the advantages of the three operators, we compare them with both basic and advanced methods employed for face recognition. Both short and long standoff distances are considered.

The outline of this chapter is as follows. Section 3.1 provides an introduction to two different categories of methods for face recognition and points out our goal of designing

operators belonging to the second category. Section 3.2 explains the structure of GWLH. Section 3.3 describes the details how CMLD is constructed and illustrates its advantage. Section 3.4 introduces the operator GMLM and demonstrates its advantage. Section 3.5 summarizes the work described in this chapter and presents our observations and conclusions.

3.1 Introduction

In the literature there exists two general categories of approaches for addressing the problem of face recognition: the holistic approach (or image-based) and the configurative (or local feature-based) approach. The former analyzes the global photometric information of a human face using subspace projections. Examples include Principal Component Analysis (PCA) [12], Independent Component Analysis (ICA) [55], Linear Discriminant Analysis (LDA) [13], Canonical-Correlation Analysis (CCA) [56], and other subspace methods. The second category of approaches has been introduced more recently. They rely on the application of local operators, filters and descriptors to extract local features. The approaches have many advantages such as a requirement of very small training sets, more robustness to illumination and occlusion, and less strict controlled conditions. Examples of local operators and descriptors include Gabor filters [17, 60–63], Histogram of Oriented Gradients (HOG) [70], Local Binary Patterns (LBP) [65–67], Weber Local Descriptor (WLD) [71, 72] and their generalizations and variants.

Most of the methods mentioned above were developed for intra-spectral matching, to be more specific for matching visible light images versus visible light images. Some operators were tuned to work with heterogeneous face images. For example, Chen et al. [73] conducted a face recognition study in thermal IR and visible spectral bands using PCA and Faceit G5. They showed that the performance of PCA in visible spectral band is higher compared to the performance of PCA in thermal IR spectral band, and that these data fused at the matching score level resulted in a performance similar to the performance of the algorithm in visible band. Li et al. [77] proposed a method to compare face images within the NIR spectral band under different illumination scenarios. Their face encoder involved the LBP operator to achieve illumination invariance and was applied to NIR images acquired at a short distance.

Akhloufi et al. [79] experimented with images from database including visible, SWIR, MWIR, and thermal infrared images. They adopted a classic Local Ternary Pattern (LTP) and a new Local Adaptive Ternary Pattern (LATP) operator for feature extraction. The work of Klare and Jain [101] employed a method based on LBP and HOG features, followed by a random sampling LDA algorithm to reduce the dimensionality of feature vectors. This encoding strategy is applied to NIR and color images for their cross-spectral matching. The results are shown to outperform Cognitec's FaceVACS [89].

The focus of this chapter is on the development of new operators falling in the second category. We propose three different operators: Gabor+WLD+LBP+HOG (GWLH), Composite Multi-Lobe Descriptor (CMLD) and Gabor Multi-Levels of Measurement (GMLM). They are introduced with the purpose to ensure a robust recognition performance when matching IR (NIR, SWIR or MWIR) face probes to visible face image galleries. The first method GWLH fuses Gabor filters with WLD, LBP and HOG to extract more useful information from the face which are complementary to each other. The second method combines a Gaussian bell function with LBP, generalized LBP, and WLD and modifies them into multi-lobe operators with smoothed neighborhood. When applied to a face image previously processed with Gabor filters, the operator encodes both the intensity and the information about the location and orientation of edges in the image. It demonstrates robustness to noise and poor image quality. The third method is based on the concept of levels of measurement proposed by Stevens [98]. It suggests a grouping of all possible operators that can only be applied to pixels in an image into levels. Each level is then applied to an image to extract features relevant to this specific level. Since the operators used at different levels are very distinct, they often extract useful complimentary information from an image. This information can be later fused resulting in a successful matching. Like the first and second methods, the multiple levels of measurements are applied to the Gabor filter responses of the face images to ensure robustness.

After feature extraction utilizing those three methods, face images are converted into the form of histograms, and a symmetric Kullback-Leibler distance is applied to the histograms to generate a matching score. Performance of the proposed operator is demonstrated on two active IR datasets, Pre-TINDERS and TINDERS [100], and is compared to the performance

of two basic operators, LBP and HOG, and two composite operators, Gabor followed by LBP and Gabor Ordinal Measures (GOM).

3.2   Gabor+WLD+LBP+HOG

In this work, we describe a new composite operator that involves a bank of Gabor filters followed by the application of three local operators. They are WLD, LBP and HOG applied to magnitude and/or phase of the output of Gabor filters. The composite operator is used to encode heterogeneous face images. We compare performance of the composite operator with the performance of (1) single local operators such as LBP, HOG, and Gabor filters, (2) a combination of LBP and HOG [102], and (3) two state-of-the-art composite operators such as Gabor filters followed by WLD, LBP and GLBP [38] and Gabor filters followed by LBP and HOG [103]. We further demonstrate that the new composite operator outperforms the other operators when applied to match NIR, SWIR or MWIR probe images to a gallery of visible light face images. In addition to varying the spectral band of the probes, we also consider a short (1.5 m) and long (50 and 105 m) standoff distances in the case of NIR and SWIR. The poor performance of the operators at the long standoff distances is linked to the quality of heterogeneous face images at that distances. We also show that the performance gap in matching heterogeneous face images increases as the quality of active IR face images increases. The quality of the heterogeneous images is measured in terms of the sharpness metric introduced by Yao et al. [104].

3.2.1   Histogram of Oriented Gradients

HOG was introduced by Dalal and Triggs in their work [70]. The essential thought behind the HOG operator is that local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions.

An input image is computed using Gaussian smoothing followed by a derivative mask such as the following basic 1-D mask $[-1, 0, 1]$. The directional derivatives can be expressed

as

$$\begin{cases} G_x(x,y) & = I(x+1,y) - I(x-1,y) \\ G_y(x,y) & = I(x,y+1) - I(x,y-1), \end{cases} \tag{3.1}$$

where $I(x,y)$ is the input image, and $G_x(x,y)$, $G_y(x,y)$ denote the derivatives along x and y directions respectively. Then the magnitude and phase components of the gradient can be calculated as

$$\begin{cases} M(x,y) & = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \\ \alpha(x,y) & = \tan^{-1}[G_y(x,y)/G_x(x,y)], \end{cases} \tag{3.2}$$

where $M(x,y)$ and $\alpha(x,y)$ are the magnitude and phase, respectively.

The next step is spatial and orientation binning. A weighted vote is calculated at each pixel for an edge orientation histogram channel based on the orientation of the gradient at that pixel, and the votes are accumulated into orientation bins over small local regions called cells (cells can be either rectangular or circular). The orientation bins are evenly spaced over $0° - 180°$ ("unsigned" gradient) or $0° - 360°$ ("signed" gradient). The vote is a function of the gradient magnitude at the pixel, very often the magnitude itself. The descriptor vector is thereafter normalized over non-overlapping blocks using the L_1 or L_2 norms, or their variants. An example of using L_2 normalization is given as:

$$\mathbf{v}^* = \mathbf{v}/\sqrt{\|\mathbf{v}\|_2^2 + \epsilon^2}, \tag{3.3}$$

where $\mathbf{v}$ is the non-normalized descriptor vector and ϵ is a small constant.

3.2.2   Fusion of Operators

In this work we fuse Gabor filters with WLD, LBP and HOG to form a new composite operator named Gabor+WLD+LBP+HOG (GWLH). The details of Gabor filters, LBP and WLD can be found in Section 2.2.2 in Chapter 2. A detailed description of HOG is provided in Section 3.2.1 in this chapter. The combination of WLD, LBP and HOG encodes both the orientation and intensity information residing in edges and their distribution in an image. A bank of Gabor filters at different orientations and scales are applied to an input image to generate a set of filter responses. The magnitude and phase of the output of a filter is then entered simultaneously into three single operators, WLD, LBP, and HOG. WLD and LBP

are applied to the magnitude of the output, and HOG and LBP are applied to the phase of the output. After the application of each operator, the encoded outputs are converted into a histogram representation and concatenated to obtain the final feature vector (see Sec. 2.2.3 for more details). A block diagram demonstrating the details of the proposed fusion scheme is displayed in Figure 3.1.
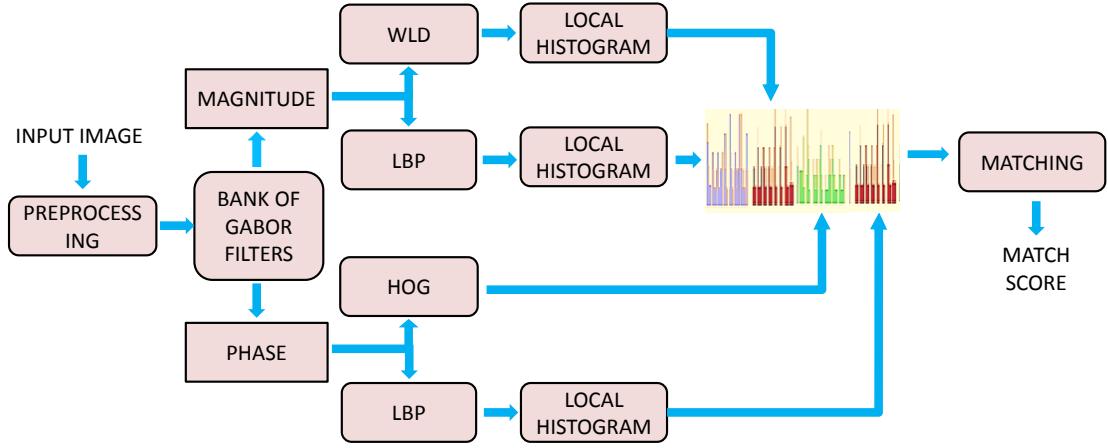
Figure 3.1: Block diagram of the fusion scheme.

3.2.3 Experimental Results and Analysis

This subsection presents the numerical results and analysis of matching heterogeneous face images. The heterogeneous faces are encoded using six algorithms: (1) LBP, (2) HOG, (3) Gabor filters, (4) Gabor filters combined with LBP (Gabor+LBP), (5) Gabor filters combined with LBP, GLBP and WLD (Gabor+LBP+GLBP+WLD), and (6) our proposed operator, Gabor+WLD+LBP+HOG (or GWLH). The first three operators are applied as single operators, while the latter three are compound operators fused using different combinations of the three simple operators and WLD. The performance of the six considered schemes is displayed as Receiver Operating Characteristic (ROC) curves in Figure 3.2 - Figure 3.5.

Matching SWIR to Visible Images

Our first experiment involves matching SWIR face images to visible face images. Visible light images form the gallery while all SWIR images are used as test images. Both the SWIR and visible images are encoded using the six algorithms mentioned above. The results of matching are shown in Figure 3.2 (a) - (c) for the standoff distances 1.5 m, 50 m and 106 m, respectively. A summary of GARs, EERs and d-prime values is given in Table 3.1, 3.2 and 3.3.

The comparison in each of the three cases of standoff distances clearly demonstrates the advantage of using the three compound operators, Gabor+LBP, Gabor+LBP+GLBP+WLD, GWLH, over the other three individual operators, LBP, HOG and Gabor filters. However, as the standoff distance increases to 50 m and then to 106 m, the difference in matching performance of three compound operators and the others three individual operators diminishes. Performance of all algorithms drops significantly. Note that the useful information that helps matching heterogeneous images is contained in the gradients of intensity and in their relative distribution, that is, their relative location and their density. Due to long standoff distances the overall quality of SWIR images in TINDERS dataset is reduced. This, in turn, affects the quality of informative gradients and their distribution in SWIR images compared to visible images or SWIR images in Pre-TINDERS dataset.

Among the three compound operators, Gabor+LBP+GLBP+WLD appears to be the best for the cases of 1.5 m and 50 m standoff distances while GWLH appears to be the best for the cases of 106 m standoff distance.

Matching NIR to Visible Images

The results of matching NIR face images to visible face images are shown in Figure 3.3 (a) - (c). The analyzed standoff distances are 1.5 m, 50 m and 106 m. A summary of the GARs, EERs and d-prime values are shown in Table 3.4, Table 3.5 and Table 3.6 for each of the considered cases of standoff distances. Again, this experiment demonstrates the advantage of the compound operators over other operators, especially our proposed GWLH operator. In all cases of standoff distances, GWLH, Gabor+LBP+GLBP+WLD and Gabor+LBP are
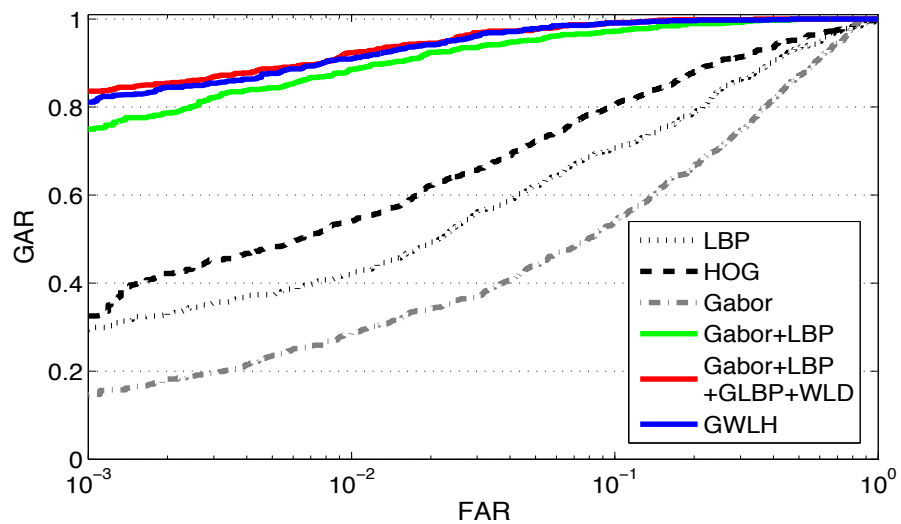
(a)
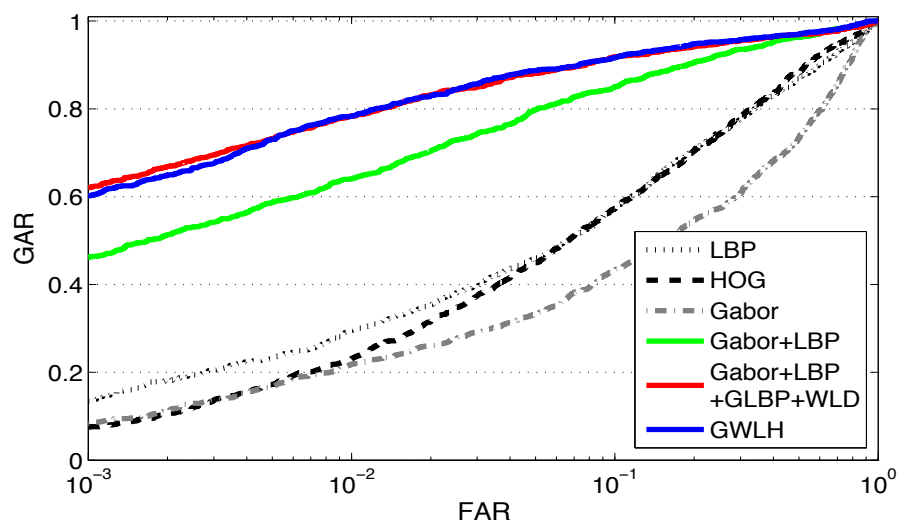
(b)
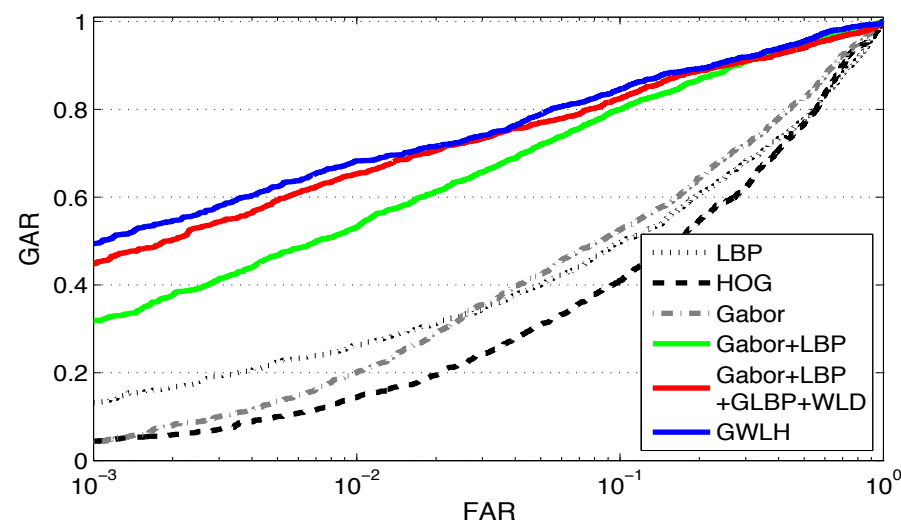
(c)

Figure 3.2: Cross-spectral matching of face using GWLH: (a) SWIR 1.5 m, (b) SWIR 50 m and (c) SWIR 106 m.

Table 3.1: GARs, EERs and d-prime values: matching SWIR probes at 1.5 m to visible gallery using GWLH.

METHOD	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER (%)	d-prime
LBP	70.70	29.56	20.61	1.66
HOG	80.47	32.55	15.36	1.86
Gabor	54.04	14.71	27.35	1.24
Gabor+LBP	97.27	75	4.82	3.09
Gabor+WLD +LBP+GLBP	**99.09**	**83.59**	**3.13**	**3.24**
GWLH	99.09	81.12	3.25	3.23

Table 3.2: GARs, EERs and d-prime values: matching SWIR probes at 50 m to visible gallery using GWLH.

METHOD	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER (%)	d-prime
LBP	57.29	13.45	25.28	1.24
HOG	57.42	7.56	25.42	1.25
Gabor	43.35	8.33	34.97	0.82
Gabor+LBP	85.01	46.15	12.89	2.25
Gabor+WLD +LBP+GLBP	**91.88**	**62.11**	**8.90**	**2.57**
GWLH	91.60	60.22	8.90	2.62

shown to have better performance than LBP, HOG and Gabor filters, though the performance of all algorithms becomes poorer as the standoff distance increases. This requires taking the quality of images into account, since NIR images at large standoff distances are extremely noisy due to insufficient illumination (see Figure 2.5).

Matching MWIR to Visible Images

In the third experiment, we match MWIR face images to a gallery of visible face images. The results shown in Figure 3.4 are for 200 classes. Color images (one per class) form the gallery while MWIR images (two per class) serve as the probes. It is clear that both our
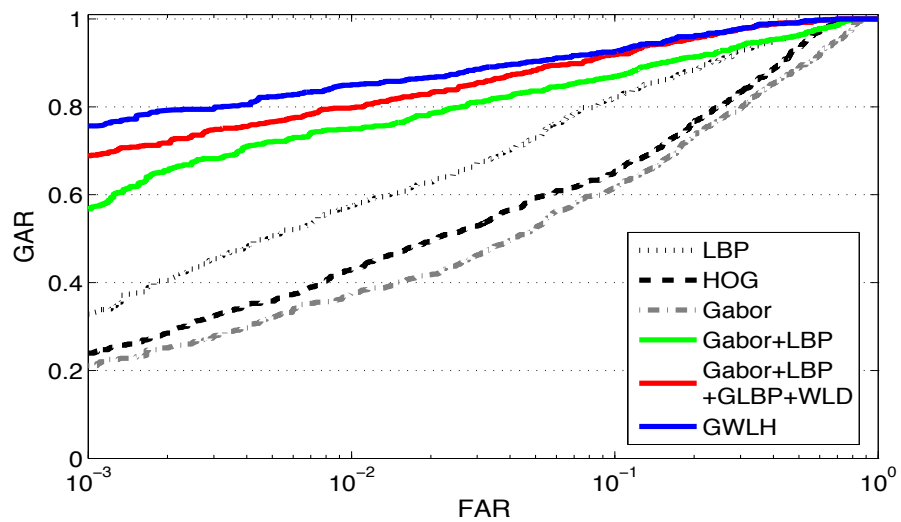
(a)
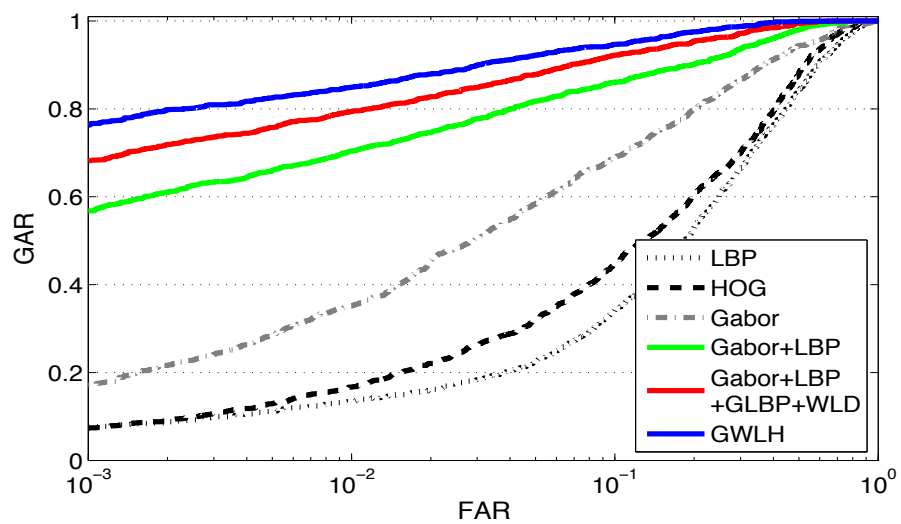
(b)
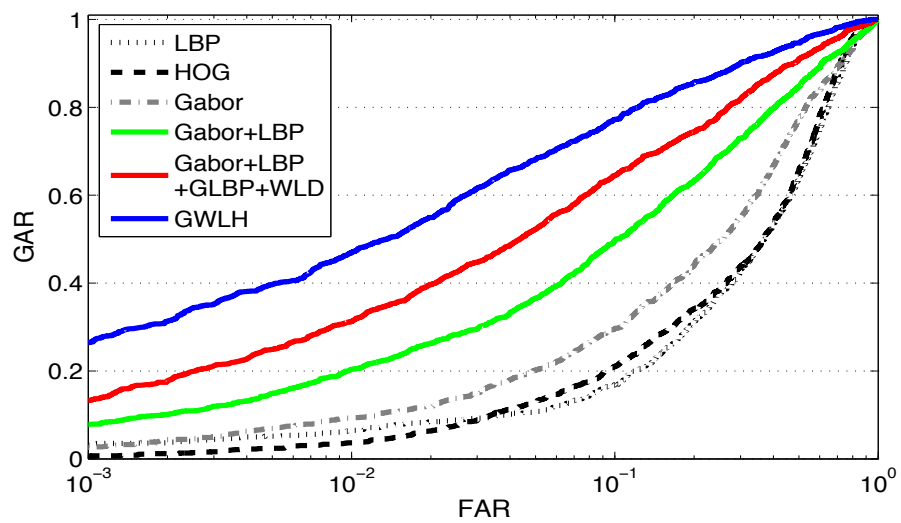
(c)

Figure 3.3: Cross-spectral matching of face regions using GWLH: (a) NIR 1.5 m, (b) NIR 50 m and (c) NIR 106 m.

Table 3.3: GARs, EERs and d-prime values: matching SWIR probes at 106 m to visible gallery using GWLH.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
LBP	49.79	13.19	31.11	0.94
HOG	41.04	4.44	33.68	0.78
Gabor	52.57	4.31	28.67	1.09
Gabor+LBP	80.00	31.81	15.83	1.99
Gabor+WLD +LBP+GLBP	82.50	44.79	14.17	2.00
GWLH	**84.65**	**49.44**	**13.13**	**2.20**

Table 3.4: GARs, EERs and d-prime values: matching NIR probes at 1.5 m to visible gallery using GWLH.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
LBP	82.03	32.81	14.36	2.12
HOG	65.23	23.96	22.03	1.68
Gabor	61.46	21.09	23.57	1.43
Gabor+LBP	86.98	56.77	11.82	2.29
Gabor+WLD +LBP+GLBP	91.93	68.88	8.73	2.48
GWLH	**92.58**	**75.65**	**8.07**	**2.63**

algorithm and other algorithms are not designed to deal with such a large spectral gap between visible light and MWIR. Nevertheless, the three compound operators are shown to be more powerful than the other three individual operators, with GWLH to be the best.

Due to different imaging nature, MWIR images are determined by the distribution of heat radiation at a subject's face while visible images are formed as patterns characterizing the reflectivity properties of subject's skin. The only common information for heterogeneous matching is the gradient lines and their relative distribution and density. MWIR and visible face regions do not contain much of this type of information in common. A summary of the GARs, EERs and d-prime values is shown in Table 3.7.

Table 3.5: GARs, EERs and d-prime values: matching NIR probes at 50 m to visible gallery using GWLH.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
LBP	34.17	7.70	31.74	0.99
HOG	44.68	7.35	29.98	1.16
Gabor	68.98	17.23	19.98	1.66
Gabor+LBP	86.13	56.79	12.54	2.33
Gabor+WLD +LBP+GLBP	92.23	68.21	8.71	2.66
GWLH	**94.68**	**76.33**	**6.58**	**3.01**

Table 3.6: GARs, EERs and d-prime values: matching NIR probes at 106 m to visible gallery using GWLH.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
LBP	16.95	3.18	43.30	0.45
HOG	21.12	0.64	42.87	0.52
Gabor	29.66	2.61	36.87	0.72
Gabor+LBP	49.72	7.84	28.43	1.10
Gabor+WLD +LBP+GLBP	64.48	13.28	23.24	1.49
GWLH	**77.26**	**26.48**	**16.52**	**1.93**

Matching LWIR to Visible Images

In the last experiment, we match LWIR face images to a gallery of visible face images: color images form the gallery while LWIR images serve as the probes. The results are shown in Figure 3.5. Again, all the six operators achieve low performance due to the great gap between the heterogeneous bands. Same as in the case of matching MWIR face images to visible light images, LWIR images display heat distribution while visible light images characterize the reflectivity properties. Nonetheless, the three compound operators are shown to be superior than the other three individual ones. A summary of the GARs, EERs and d-prime values is shown in Table 3.8.
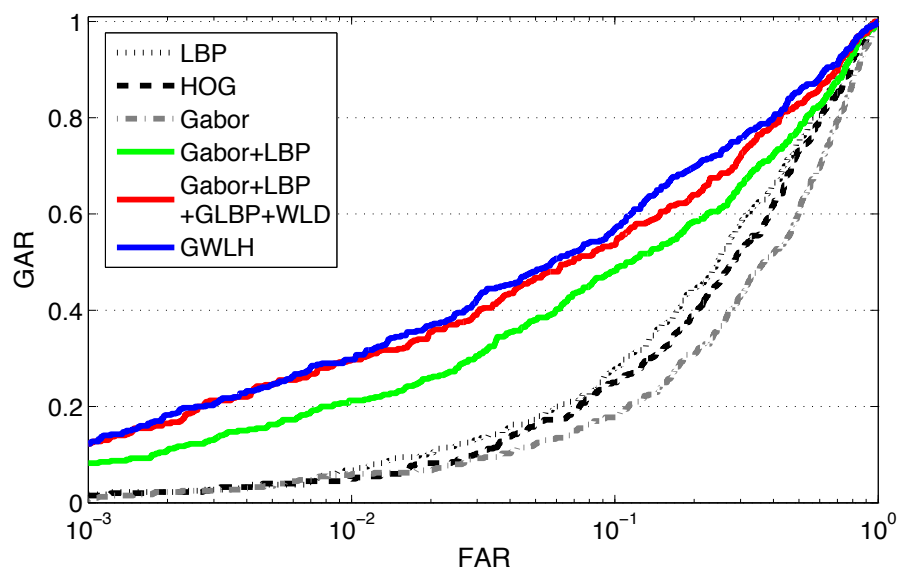
Figure 3.4: Cross-spectral matching of the face using GWLH: MWIR.
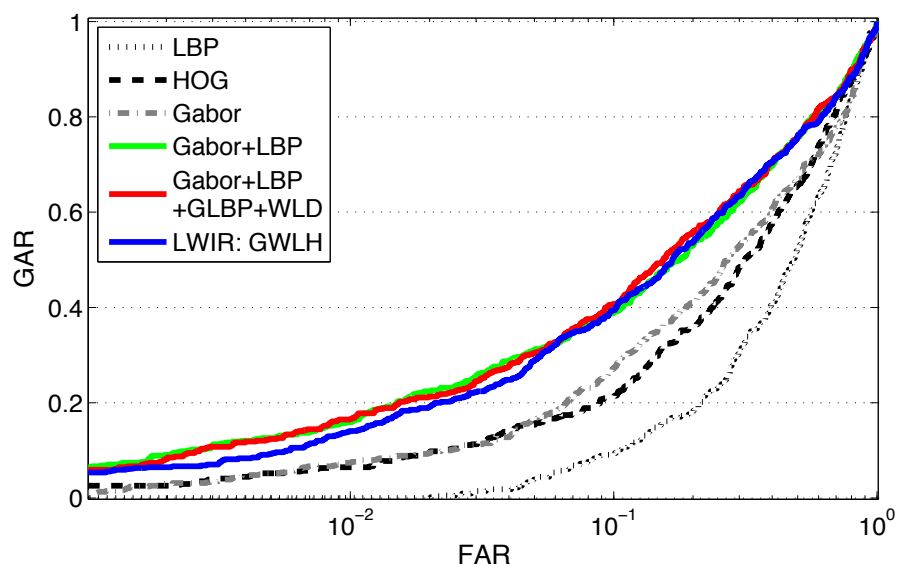
Figure 3.5: Cross-spectral matching of the face using GWLH: LWIR.

Table 3.7: GARs, EERs and d-prime values: matching MWIR probes to visible gallery using GWLH.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
LBP	27.75	1.75	37.00	0.70
HOG	25.00	1.50	39.27	0.57
Gabor	17.75	0.75	45.05	0.34
Gabor+LBP	48.25	8.25	32.25	0.97
Gabor+WLD +LBP+GLBP	53.50	12.25	29.01	1.18
GWLH	**57.00**	**12.25**	**26.50**	**1.25**

Table 3.8: GARs, EERs and d-prime values: matching LWIR probes to visible gallery using GWLH.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
LBP	9.26	0.001	49.08	0.04
HOG	21.30	2.59	41.18	0.41
Gabor	27.41	1.11	39.82	0.38
Gabor+LBP	38.89	**6.48**	34.64	**0.74**
Gabor+WLD +LBP+GLBP	**40.74**	5.93	34.08	0.71
GWLH	39.44	5.37	**33.52**	0.68

Quality and Performance

As anticipated, the quality of active and passive IR (i.e., SWIR, NIR, MWIR and LWIR) probes affects the matching performance. In this work, the quality of the probes is a function of the standoff distance. We use an adaptive sharpness measure [104] to calculate the image quality of the probes in SWIR, NIR , MWIR and LWIR spectra at all the standoff distances, as shown in Table 3.9. From the results, the sharpness measure value decreases as standoff distance increases in both cases of SWIR and NIR spectra. This is in consistence with the visual perception of the quality of images in the datasets. The overall sharpness measure values of SWIR images are higher compared to the sharpness measure values of NIR images.

It is further observed that although the matching performance of SWIR images at the short standoff distance is lower than that of NIR images in general, the performance of SWIR data degrades with increasing standoff distance slower than that of NIR data does.

Table 3.9: Sharpness measure of the SWIR, NIR, MWIR and LWIR images at different standoff distances

STATISTICS OF SHARPNESS MEASURE	SWIR 1.5 m	SWIR 50 m	SWIR 106 m	NIR 1.5 m	NIR 50 m	NIR 106 m	MWIR	LWIR
Mean	0.5835	0.5112	0.4391	0.4390	0.3910	0.3741	0.3496	0.2946
Standard Deviation	0.0707	0.0732	0.0730	0.0595	0.0461	0.0642	0.0633	0.1429

3.3   Composite Multi-Lobe Descriptor

3.3.1   General Structure

Chai et al. [91] have recently shown that the application of Gabor filters to visible light images followed by the application of multi-lobe ordinal measures to filtered images yields a robust feature extraction method. Inspired by this result, we introduce two smooth multi-lobe kernel functions, with features specified by the original LBP and WLD, with multiple smooth lobes specified by a Gaussian bell function and their parameters, and with a discrete function specifying lobe placement parameters.

We thereafter construct a new compound operator – Composite Multi-Lobe Descriptor (CMLD) [105, 106], which combines multi-lobe kernel functions and LBP, GLBP and WLD and modifies their original forms into multi-lobe functions with smoothed neighborhoods. The new descriptor encodes both the magnitude and phase information of an input image. They are applied to the outputs of Gabor filters. The introduction of the multi-lobe functions with smoothed neighborhoods makes CMLD robust against noise and poor image quality. An example illustrating their robustness is presented in Figure 3.6. The three number matrices positioned in the middle of the figure demonstrate the process of generating a noisy 3×3 image (at the bottom of the figure). The left side of the figure presents the result of encoding
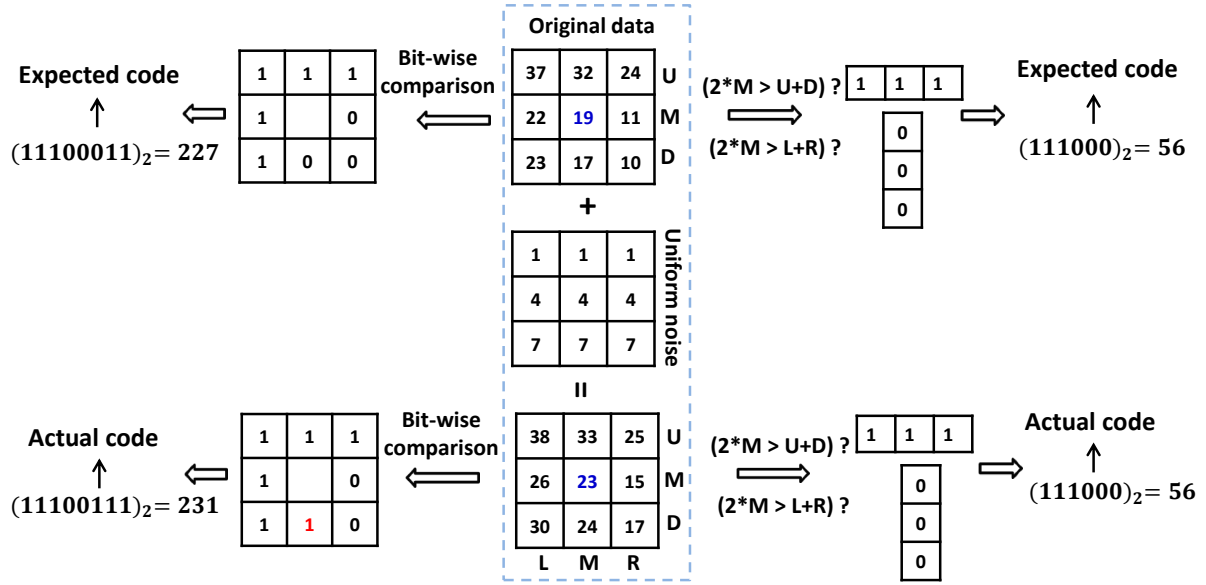
Figure 3.6: An example illustrating the robustness of MLLBP.

the original clean image with LBP and the result of encoding the noisy image with LBP. Note that the expected code (encoding of the original image) and the actual code (encoding of the noisy image) disagree. The right side of the figure presents encoding of the original and the noisy images with MLLBP. In this case, the actual and expected codes are the same.

3.3.2   Multi-Lobe Descriptors

The kernel function of multi-lobe LBP (MLLBP) denoted as $f_{MLLBP}(\mathbf{z}; \theta, L)$ is mathematically described as

$$
\begin{aligned}
f_{MLLBP}(\mathbf{z}; \theta, L) = & \; C_p \sum_{l=1}^{N_p} \frac{1}{\sqrt{2\pi}\sigma_{l,\theta,L}} \exp\left[\frac{-(\mathbf{z} - \mu_{l,\theta,L})^T(\mathbf{z} - \mu_{l,\theta,L})}{2\sigma_{l,\theta,L}^2}\right] \\
& - C_n \sum_{k=1}^{N_n} \frac{1}{\sqrt{2\pi}\hat{\sigma}_{k,\theta,L}} \exp\left[\frac{-(\mathbf{z} - \hat{\mu}_{k,\theta,L})^T(\mathbf{z} - \hat{\mu}_{k,\theta,L})}{2\hat{\sigma}_{k,\theta,L}^2}\right]
\end{aligned}
\tag{3.4}
$$

where $\mathbf{z} = (x, y)$ is the location of a pixel. $\mu_{l,\theta,L}$ and $\sigma_{l,\theta,L}$ denote the central position and the scale of the l-th positive lobe of the kernel function with orientation θ and total number of lobes L, while $\hat{\mu}_{k,\theta,L}$ and $\hat{\sigma}_{k,\theta,L}$ denote the central position and the scale of the k-th negative lobe of the same kernel. N_p and N_n are the numbers of positive and negative lobes, respectively. Coefficients C_p and C_n balance the contribution of positive and negative

lobes, i.e., $C_p N_p = C_n N_n$. An illustration of the constructed functions is provided in Figure 3.7.

(a)   (b)
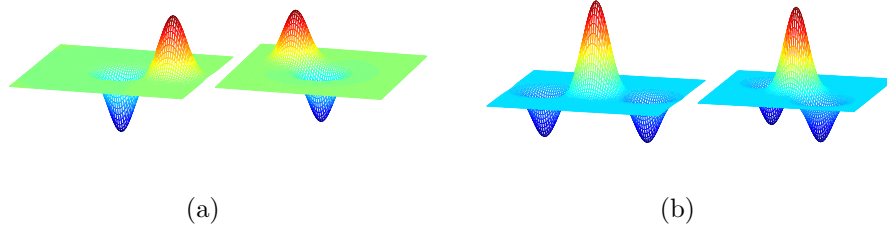
Figure 3.7: Examples of multi-lobe kernel functions at different orientations: (a) bi-lobe and (b) tri-lobe.

To encode an image with MLLBP, the kernel function $f_{MLLBP}(\mathbf{z})$ is applied to an image $I(\mathbf{z})$ (the magnitude response of a Gabor filter as given in (2.1)) followed by the application of the binary quantizer and the uniform pattern mapping as specified in (2.6) and (3.10),

$$MLLBP_N(\mathbf{z}) = \mathcal{U}\left\{\sum_{i=1}^{N} \mathcal{I}\left[I(\mathbf{z}) * f_{MLLBP}^{(i)}(\mathbf{z})\right] \cdot 2^i\right\}, \tag{3.5}$$

where $*$ stands for convolution, and $f_{MLLBP}^{(i)}(\mathbf{z})$ denotes the i-th member in the ensemble of all the kernel functions for MLLBP at all orientations and numbers of lobe, $\{f_{MLLBP}(\mathbf{z}; \theta, L) : \theta = 1, 2, ..., \Theta; L = 2, 3, ..., M\}$, with Θ and M being the total number of orientations and the maximum number of lobes, respectively.

The multi-lobe version of GLBP (MLGLBP) involves the same kernel function as MLLBP. After application of $f_{MLLBP}(\mathbf{z})$ to an image $I(\mathbf{z})$, the output is processed with the thresholding function $\mathcal{T}_t(\cdot)$ and the uniform pattern mapping $\mathcal{U}(\cdot)$ described in (2.7) and (2.8), respectively:

$$MLGLBP_N(\mathbf{z}) = \mathcal{U}\left\{\sum_{i=1}^{N} \mathcal{T}_t\left[I(\mathbf{z}) * f_{MLLBP}^{(i)}(\mathbf{z})\right] \cdot 2^i\right\}. \tag{3.6}$$

The multi-lobe WLD (MLWLD) kernel function is designed following the outline for developing the MLLBP kernel function with the difference that MLWLD involves the original WLD operator instead of the original LBP. The mathematical expression for the kernel function of MLWLD is as follows

$$f_{MLWLD}(\mathbf{z}; \theta, L) = \sum_{l=1}^{L} \frac{C_l}{\sqrt{2\pi}\sigma_{l,\theta,L}} \exp\left[-\frac{(\mathbf{z} - \mu_{l,\theta,L})^T(\mathbf{z} - \mu_{l,\theta,L})}{2\sigma_{l,\theta,L}^2}\right], \tag{3.7}$$

where $\mu_{l,\theta,L}$ and $\sigma_{l,\theta,L}$ are the center and the scale of the kernel function at orientation θ, and number of lobes L, respectively. $\{C_l\}$ are the coefficients to keep a balance between the positive and negative lobes. The ensemble of all the kernel functions for MLWLD at each orientation and number of lobes is denoted by $\{f_{MLWLD}(\mathbf{z}; \theta, L) : \theta = 1, 2, ..., \Theta; L = 2, 3, ..., M\}$, where Θ is the total orientations and M is the maximum number of lobes.

To complete the feature extraction with MLWLD, the kernel function $f_{MLWLD}(\mathbf{z})$ is applied to an input image $I(\mathbf{z})$ (again, $I(\mathbf{z})$ is the magnitude of a Gabor filter response described in (2.1)) and then transformed by means of the *arctangent* function and the quantizer $\mathcal{Q}_l$ as specified in (2.3):

$$MLWLD_N(\mathbf{z}) = \mathcal{Q}_l \left\{ \tan^{-1} \left[\sum_{i=1}^{N} \frac{I(\mathbf{z}) * f_{MLWLD}^{(i)}(\mathbf{z})}{I(\mathbf{z})} \right] \right\}, \tag{3.8}$$

where $f_{MLWLD}^{(i)}(\mathbf{z})$ is the i-th element of $\{f_{MLWLD}(\mathbf{z}; \theta, L)\}$.

The multi-lobe version of the individual operators (ie. LBP, GLBP and WLD) can be viewed as a generalization of their original forms. Alternatively, the original forms can be seen as the limits of their multi-lobe kernel functions. An illustration of the relationship between MLLBP and LBP is provided in Figure 3.8. It demonstrates the transformation of MLLBP to LBP as the number of filter lobes, L reduces to 2 and the skewness of the lobes, σ approaches zero.
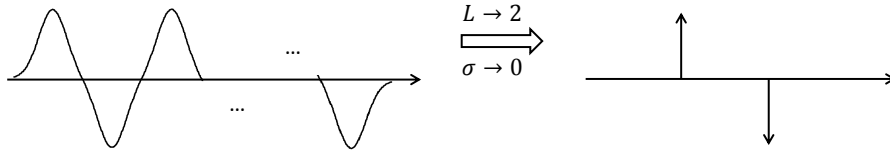
Figure 3.8: Illustration of the relationship between multi-lobe LBP and LBP (in 1D).

3.3.3 Experimental Results

In this section we describe several cross spectral matching experiments and summarize the results of matching SWIR and NIR probe images to a gallery of high quality visible images. Results are presented for both short and long standoff distances. The heterogeneous

images are encoded using five operators: (1) LBP, (2) HOG, (3) Gabor followed by LBP (Gabor+LBP), (4) GOM, and (5) CMLD. The results of matching are displayed in the form of Receiver Operating Characteristic (ROC) curves. We plot Genuine Accept Rate (GAR) versus False Accept Rate (FAR). Summaries of Equal Error Rates (EER), d-prime values, and GARs at the FAR set to 0.1 and 0.001 are provided in tables.

Matching SWIR Probes against Visible Gallery

The first experiment involves matching SWIR face images (the probes) to visible face images (the gallery). The performance of the individual original operators such as HOG and LBP can be treated as benchmarks. The results of matching parameterized by different stand-off distances are shown in Figure 3.9, 3.10 and 3.11.

Short Standoff Distance For the case of the short standoff distance (Pre-TINDERS dataset), the performance of the single operators, LBP and HOG, is inferior to the performance of the combined operators Gabor+LBP, GOM and CMLD. Among the combined operators, CMLD performs best. Table 3.10 presents a summary of EERs, d-prime values and GAR values at FAR set to 0.1 and 0.001 values.
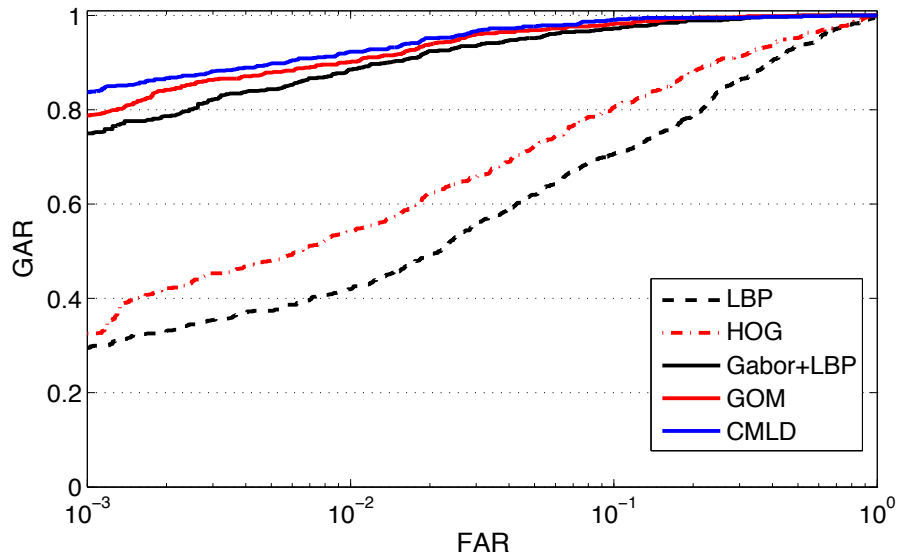
Figure 3.9: ROC curves: matching SWIR probes at 1.5 m to visible gallery using CMLD.

Table 3.10: EERs, d-prime and GAR values: matching SWIR probes at 1.5 m to visible gallery using CMLD.

Method	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER(%)	d-prime
LBP	70.70	29.56	20.61	1.66
HOG	80.47	32.55	15.36	1.86
Gabor+LBP	97.27	75.00	4.82	3.09
GOM	98.18	78.78	3.64	3.18
CMLD	**99.09**	**83.72**	**3.12**	**3.29**

Long Standoff Distance   SWIR images at longer standoff distances (50 m and 106 m in the case of TINDERS dataset) experience some loss of quality due to air turbulence, insufficient illumination, and optical effects during data acquisition. This immediately reflects on the matching performance. Figure 3.10 and 3.11 display the results of cross-spectral matching for 50 m and 106 m standoff distances, respectively. Gallery images are retained from the previous session. Once again, the three composite operators outperform the single operators, which was anticipated. However, at longer standoff distances their matching performance (especially the values of GAR at FAR=10^{-3}) drops nearly 2 times for the case of 50 m and 2.5 times for the case of 106 m. EERs, d-prime values and GARs at FAR set to 0.1 and 0.001 are summarized in Table 3.11 and 3.12.

Table 3.11: EERs, d-prime and GAR values: matching SWIR probes at 50 m to visible gallery using CMLD.

Method	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER(%)	d-prime
LBP	57.29	13.45	25.28	1.24
HOG	57.42	7.56	25.42	1.25
Gabor+LBP	85.01	46.15	12.89	2.25
GOM	86.41	39.98	**11.97**	2.27
CMLD	**86.76**	**45.73**	12.03	**2.31**
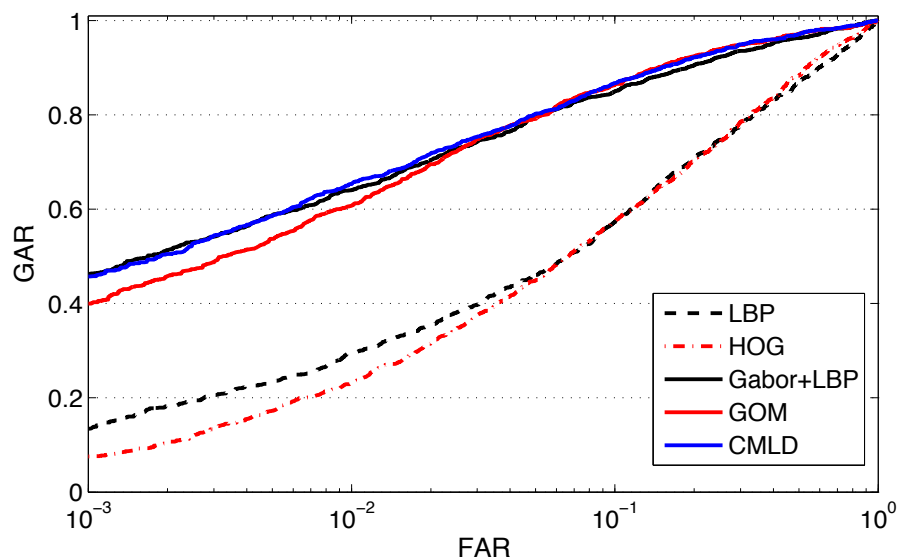
Figure 3.10: ROC curves: matching SWIR probes at 50 m to visible gallery using CMLD.
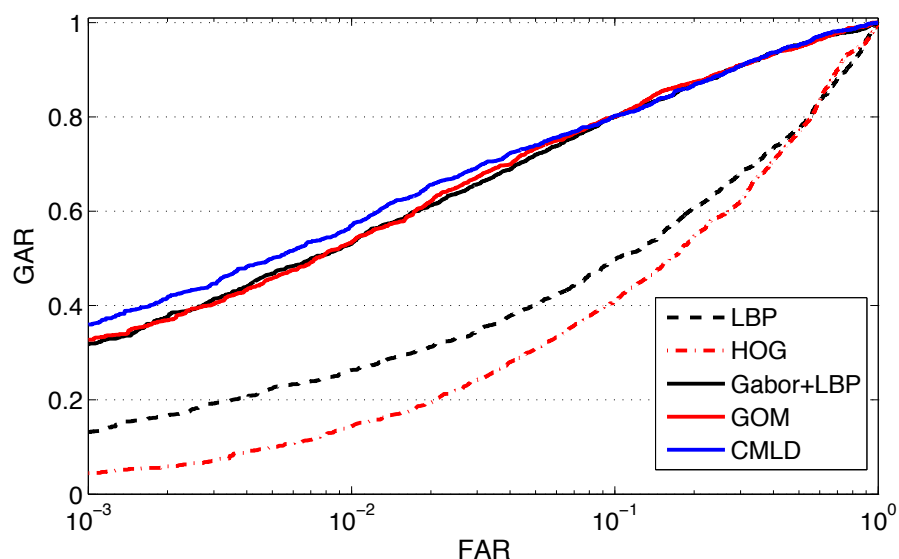
Figure 3.11: ROC curves: matching SWIR probes at 106 m to visible gallery using CMLD.

Matching NIR Probes against Visible Gallery

In the second experiment, NIR face images (probes) are matched to short range visible face images (gallery). The results of matching parameterized by three stand-off distances are shown in Figure 3.12, 3.13 and 3.14, respectively.

Table 3.12: EERs, d-prime and GAR values: matching SWIR probes at 106 m to visible gallery using CMLD.

Method	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER(%)	d-prime
LBP	49.79	13.19	31.11	0.94
HOG	41.04	4.44	33.68	0.78
Gabor+LBP	80.00	31.81	15.83	1.99
GOM	80.07	32.78	**14.78**	2.02
CMLD	**80.28**	**35.97**	15.76	**2.04**

Short Standoff Distance   Similar to the the case of SWIR probe images, all composite operators demonstrate a relatively high performance compared to the individual operators. CMLD appears to outperform the other two composite operators, Gabor+LBP and GOM. Table 3.13 summarizes the EERs, d-primes and GARs at FAR equal to 0.1 and 0.001 for this case.
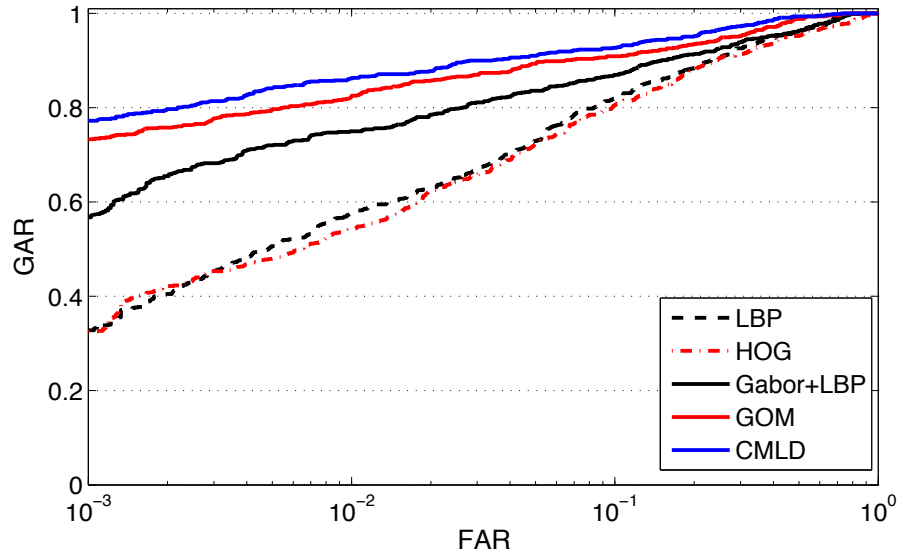
Figure 3.12: ROC curves: matching NIR probes at 1.5 m to visible gallery using CMLD.

Standoff Distance   Long range NIR probes display a cardinally different performance. As can be seen from Figure 2.5, NIR images at 106 m have much lower contrast and overall quality compared to NIR images at 50 m. This difference in image quality immediately

Table 3.13: EERs, d-prime and GAR values: matching NIR probes at 1.5 m to visible gallery using CMLD.

Method	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER(%)	d-prime
LBP	82.03	32.81	14.36	2.12
HOG	65.23	23.96	22.03	1.68
Gabor+LBP	86.98	56.77	11.82	2.29
GOM	90.89	73.31	9.27	2.59
CMLD	**92.71**	**77.21**	**7.68**	**2.72**

reflects on the matching performance of the two sets of probes (50 m probes and 106 m probes). This also reflects on the interplay among the 5 operators. Figure 3.13 and Figure 3.14 display the cross matching results for 50 m and 106 m standoff distances, respectively. Comparing the composite operators in terms of their performance, NIR at 50 m shows that CMLD performs slightly better than GOM, followed by Gabor+LBP. At 106 m NIR probes do not perform as well. In fact, the performance of composite operators drops at least three times compared to the performance of the same operators applied to NIR at 50 m. Figure 3.14 presents the only case in our experiments when GOM performs slightly better than CMLD. Similar to the results of all other experiments, the composite operators substantially outperform the individual operators.

Table 3.14 and Table 3.15 present a summary of EERs, d-primes and GARs at FAR set to 0.1 and 0.001 for the case of 50 m and 106 m standoff distances, respectively.

Table 3.14: EERs, d-prime and GAR values: matching NIR probes at 50 m to visible gallery using CMLD.

Method	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER(%)	d-prime
LBP	34.17	7.70	31.74	0.99
HOG	44.68	7.35	29.98	1.16
Gabor+LBP	86.13	56.79	12.54	2.33
GOM	90.06	64.29	10.00	2.65
CMLD	**90.76**	**67.51**	**9.52**	**2.65**
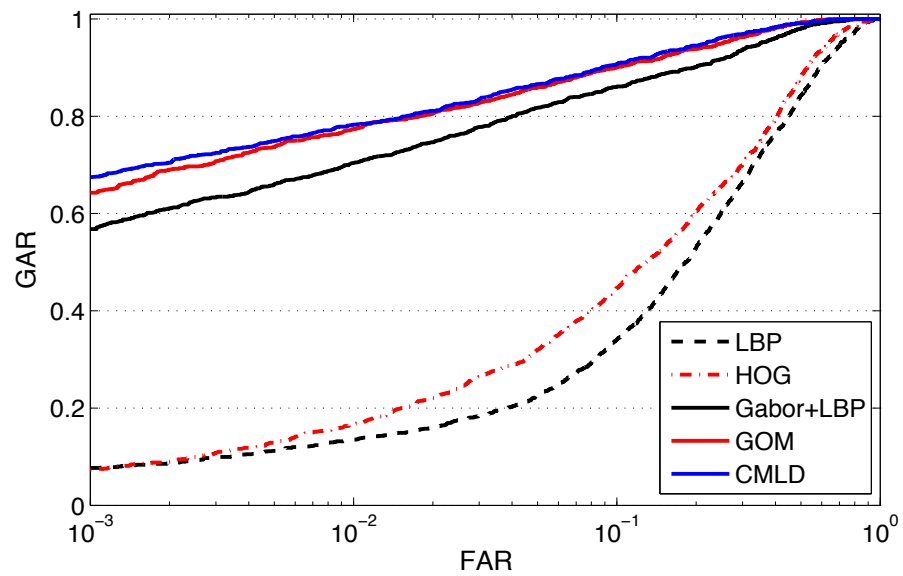
Figure 3.13: ROC curves: matching NIR probes at 50 m to visible gallery using CMLD.
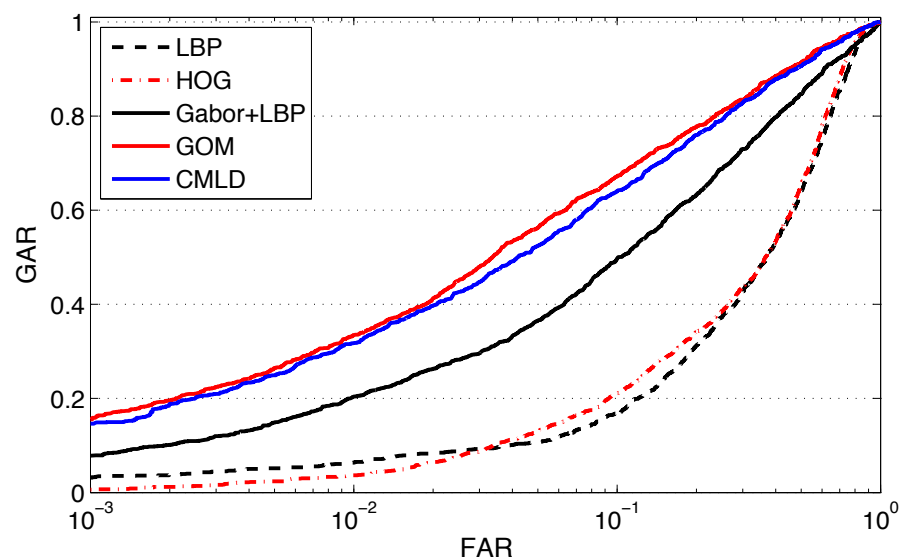
Figure 3.14: ROC curves: matching NIR probes at 106 m to visible gallery using CMLD.

Table 3.15: EERs, d-prime and GAR values: matching NIR probes at 106 m to visible gallery using CMLD.

Method	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER(%)	d-prime
LBP	16.95	3.18	43.30	0.45
HOG	21.12	0.64	42.87	0.52
Gabor+LBP	49.72	7.84	28.43	1.10
GOM	**67.30**	**15.53**	**21.65**	**1.58**
CMLD	64.12	14.62	22.55	1.51

3.4 Gabor Multi-Levels of Measurement

3.4.1 Levels of Measurement

A level of measurement, also known as a scale of measure, refers to the nature of information within the values assigned to a variable (some quantity) to be measured and the relationship among the values. Psychologist Stanley S. Stevens proposed a typology (the best known one) with four levels: the nominal, the ordinal, the interval, and the ratio levels [98].

The nominal level is often referred to as the qualitative level, and measurements made at the other three levels are called quantitative data. Usage of the concept has been witnessed in many disciplines such as natural sciences, linguistics and political science. Examples of these classifications include taxonomic ranks in biology, parts of speech in grammar and political affiliation in politics. A very common case can be seen in describing the gender of human beings, as given in Figure 3.15.

The nominal level of measurement simply "names" the attributes of a variable to be measured uniquely by assigning certain numerical values (as in the gender example in Figure 3.15). Neither ordering of the attributes is implied, nor arithmetic or logical operations on the assigned values are meaningful. In ordinal measurement, however, the attributes can be rank-ordered. Larger values suggest a greater amount of a certain quality. Nonetheless, the distances (the difference) between values can not be defined. Therefore, the interval between any two values is not interpretable in such an ordinal measure. When it comes to interval measurement the distances between attributes do have meanings. For instance,
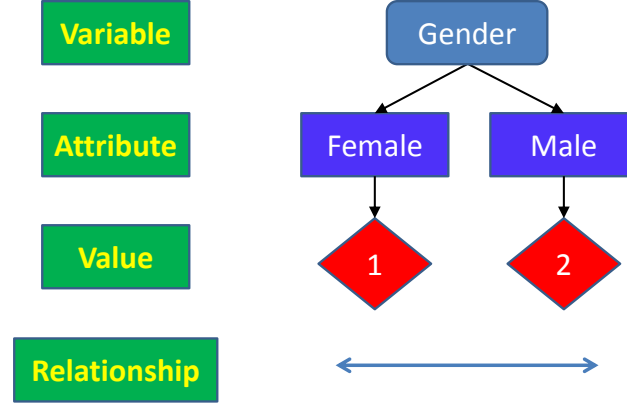
Figure 3.15: Nominal measurement for the gender.

when measuring the temperature (in Celsius), the distance between 15° and 25° is the same as the distance between 30° and 40°. But note that ratios between values still do not make sense – 40° is not twice as hot as 20°. Finally, in ratio measurement ordering, distance and ratio are all meaningful and there is always an absolute zero defined. This means that one can construct a meaningful fraction (or ratio) with a ratio variable. Practice of this level of measurement is commonly seen in sciences such as measuring length, mass and force, to name a few.

The four levels of measurement have increasing complexity (from nominal to ratio) and measure different types of information. Different mathematical or logical operations are defined at each level. Different operators or encoders used for face recognition can be classified into corresponding levels by examining the core operations involved in the operator. A summary of the four levels of measurement and their properties as well as example face recognition operators are given in Table 3.16.

In this section we design another new operator, named Interval Measure Descriptor (IMD), in addition to the other two new operators we designed in previous sections of this chapter. So far no other operators (to the best of our knowledge) described in the literature are working at this level. We further propose to fuse it with operators at other levels to see if a performance improvement can be gained by utilizing the complementary information. The fused operator is named Gabor Multi-Levels of Measurement (GMLM).

Table 3.16: Summary of the four levels of measurement with example operators.

Measurement	Arithmetic and Logical Operations	Complexity	Meaning	Example Operator
nominal	$=, \neq$	lowest	categories	none
ordinal	$=, \neq, <, >$	mediate	orders	LBP
interval	$=, \neq, <, >, +, -$	mediate	distance meaningful	IMD (ours)
ratio	$=, \neq, <, >, +, -, \times, \div$	highest	absolute zero meaningful	HOG, WLD

3.4.2   Operator at the Interval Level

We propose an operator acting at the interval level of measurement and name it Interval Measurement Descriptor (IMD). Given an input image, the operator encodes the difference of every pixel and its neighbors (In our work we consider 8 neighbors). The range of the difference is divided into K intervals, uniformly or non-uniformly. An illustration of a K-interval division scheme is given in Figure 3.16. Each interval will then be assigned an integer code. The final encoding is generated by concatenating the codes of all the neighbors. The mathematical definition of IMD is given by

$$IMD(x) = \sum_{i=1}^{8} \mathcal{S}(x_i - x)K^i, \tag{3.9}$$

where x_i is a neighbor of the central pixel x within the input image and K is the total number of intervals used. $S(\cdot)$ is the assigning function defined as (assuming $K = 4$)

$$\mathcal{S}(x) = \begin{cases} 0, & -255 \leq x < -128, \\ 1, & -128 \leq x < 0, \\ 2, & 0 \leq x < 128, \\ 3, & 128 \leq x \leq 255. \end{cases} \tag{3.10}$$

3.4.3   Gabor Multi-Levels of Measurement

In order to boost the recognition performance by utilizing complementary information contained in the different levels of measurement, we further propose to fuse IMD with oper-
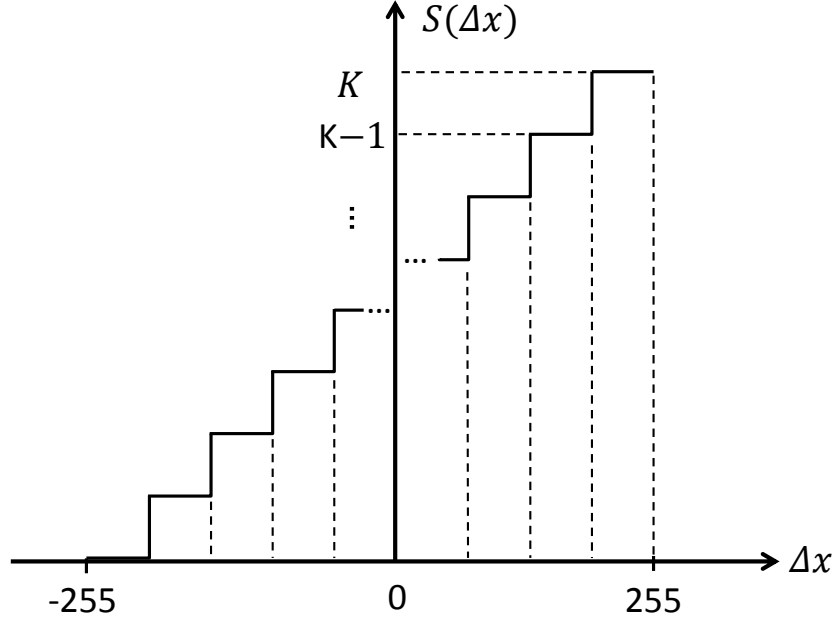
Figure 3.16: Encoding with Interval Measurement Descriptor.

ators at other levels. An input image is first passed through a bank of Gabor filters. Then the responses of the Gabor filters are passed through a set of operators working at three different levels of measurement: LBP at the ordinal level, the proposed operator IMD at the interval level, and HOG and WLD at the ratio level. The encoding results after the set of operators will be concatenated. The final compound operator is named Gabor Multi-Levels of Measurement (GMLM). A block diagram explaining the structure of this design is provided in Figure 3.17.

3.4.4   Experimental Results

In this subsection we present the cross-spectral matching results using the proposed new operator GMLM. The datasets involved are: Pre-TINDERS and TINDERS which consist of SWIR and NIR images collected at both short and long standoff distances, PCSO which consists of MWIR images collected at a short standoff distance, and Q-FIRE consisting of LWIR images collected at a short standoff distance. More details for the datasets can be found in Table 3.17. The matching performance using GMLM is compared to the performance using two other operators: Gabor+LBP+GLBP+WLD and Gabor+LBP. The results are
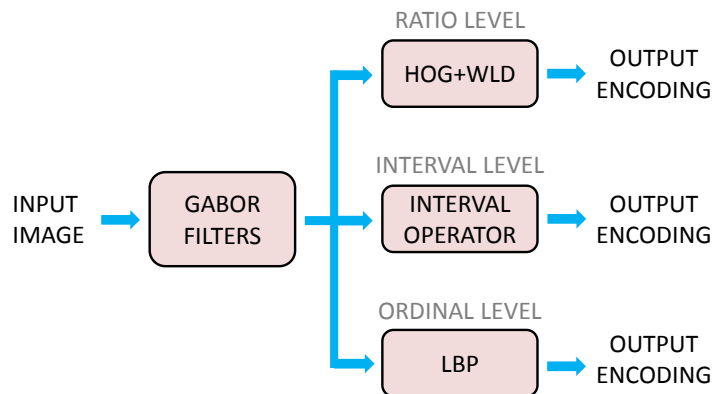
Figure 3.17: Block diagram for Gabor Multi-Levels of Measurement.

displayed as ROC curves. We plot Genuine Accept Rate (GAR) versus False Accept Rate (FAR). Summaries of Equal Error Rates (EER), d-prime values, and GARs at the FAR set to 0.1 and 0.001 are provided in tables.

Table 3.17: Summary of the datasets.

DATASET	CLASS	TOTAL # IMAGES	SPECTRUM	ACQUISITION DISTANCE	ORIGINAL RESOLUTION
Pre-TINDERS	48	576	visible NIR SWIR	1.5 m	visible: 1600×1200 NIR: 640×512 SWIR: 640×512
TINDERS	48	1255	visible NIR SWIR	visible: 1.5 m NIR & SWIR : 50 m and 106 m	visible: 480×640 NIR: 640×512 SWIR: 640×512
PCSO	1000	3000	visible MWIR	1.5 m	visible: 480×600 MWIR: 480×640
Q-FIRE	82	431	visible LWIR	2 m	visible: 1920×1080 LWIR: 726×480

Matching SWIR to Visible Images

The first experiment involving GMLM is conducted to match SWIR face images (the probes) to visible face images (the gallery). The performance of the other two operators

can be treated as benchmarks. The results of matching parameterized by different stand-off distances are shown in Figure 3.18 (a) - (c).

For the case of the short standoff distance (1.5 m), the performance of Gabor+LBP+GLBP+WLD and GMLM is better than the performance of Gabor+LBP, with the performance of the former two operators being very close to each other. As the standoff distance increases, the performance gap between the former two operators and the third operator becomes larger. When the distance reaches 106 m, GMLM performs better than Gabor+LBP+GLBP+WLD. A summary of GAR at FAR=0.1 and 0.001, EER and d-prime values is given in Tables 3.18, 3.19 and 3.20 for the cases of 1.5 m, 50 m and 106 m, respectively.

Table 3.18: GAR, EER and d-prime:Matching SWIR images at 1.5 m against visible light images using GMLM.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
Gabor+LBP	97.27	75.00	4.82	3.09
Gabor+LBP GLBP+WLD	99.09	**83.59**	**3.13**	3.24
GMLM	**99.09**	81.38	3.50	**3.24**

Table 3.19: GAR, EER and d-prime:Matching SWIR images at 50 m against visible light images using GMLM.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
Gabor+LBP	85.01	46.15	12.89	2.25
Gabor+LBP GLBP+WLD	**91.88**	**62.11**	8.90	2.57
GMLM	91.60	58.54	**8.89**	**2.60**

Matching NIR to Visible Images

The second experiment is for matching probes of NIR face images against a gallery of visible face images. Again, the performance of the two other operators can be treated as
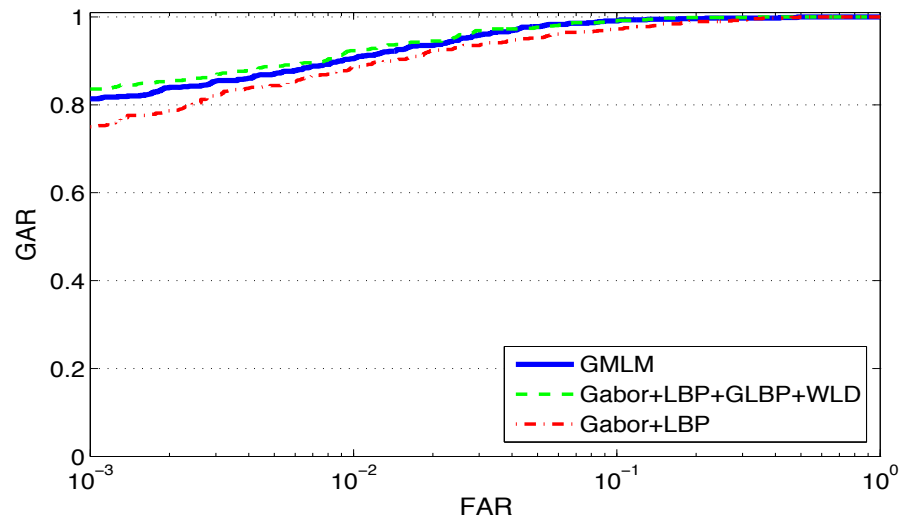
(a)
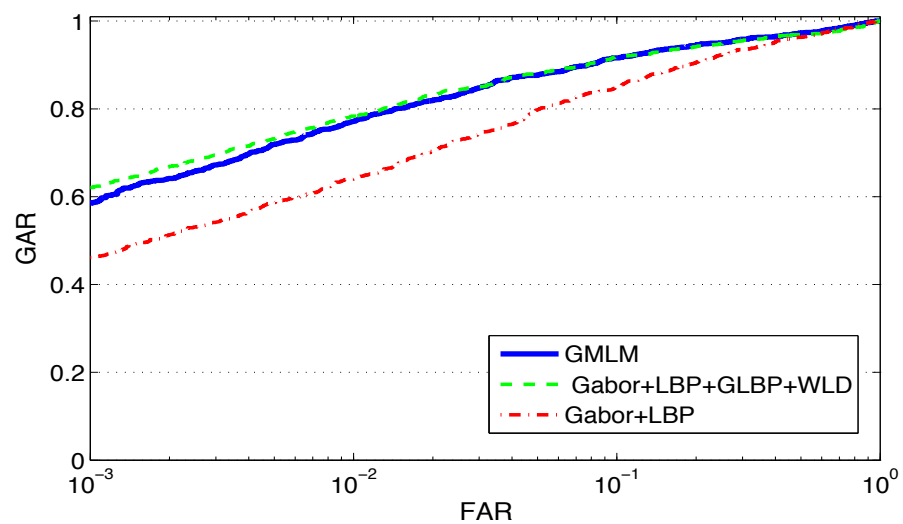
(b)
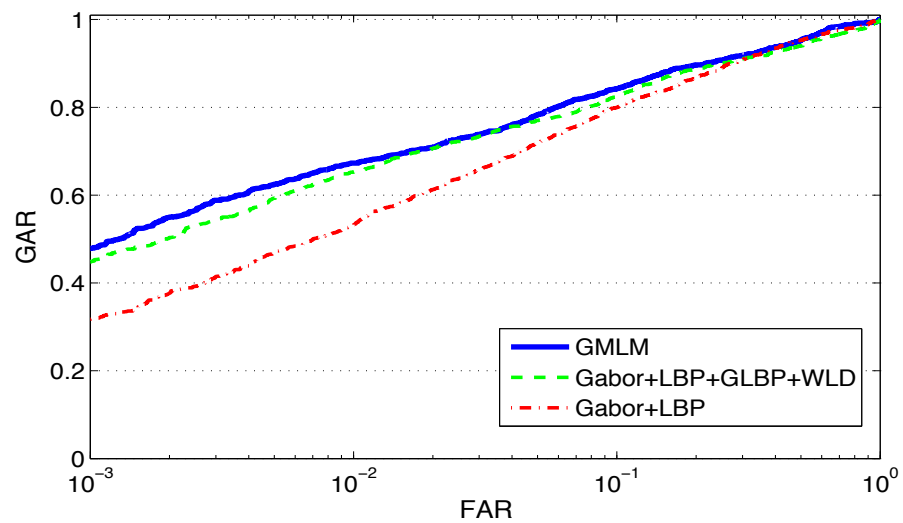
(c)

Figure 3.18: Matching SWIR against visible light images using GMLM: (a) SWIR 1.5 m, (b) SWIR 50 m, and (c) SWIR 106 m.

Table 3.20: GAR, EER and d-prime:Matching SWIR images at 106 m against visible light images using GMLM.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
Gabor+LBP	80.00	31.81	15.83	1.99
Gabor+LBP GLBP+WLD	82.50	44.79	14.17	2.00
GMLM	**84.24**	**47.85**	**13.26**	**2.19**

benchmarks. The results of matching parameterized by different stand-off distances are shown in Figure 3.19 (a) - (c).

For all cases of the standoff distances, i.e., 1.5 m, 50 m and 106 m, the performance of GMLM is the best among the three, with GMLM and Gabor+LBP+GLBP+WLD significantly better than Gabor+LBP. As the standoff distance increases, the gap between the former two operators and the third operator becomes larger. Furthermore, GMLM appears to be more advantageous than the other two operators especially when the distance becomes larger. More matching parameters such as EER, d-prime can be found in Table 3.21, 3.22 and 3.23.

Table 3.21: GAR, EER and d-prime:Matching NIR images at 1.5 m against visible light images using GMLM.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
Gabor+LBP	86.98	56.77	11.82	2.29
Gabor+LBP GLBP+WLD	**91.93**	68.88	**8.73**	2.48
GMLM	90.89	**74.74**	9.38	**2.55**

Matching MWIR to Visible Images

The third experiment is conducted to match MWIR face probes against the visible light face gallery. Again, the performance of the two other operators can be treated as benchmarks. The results of matching parameterized by different stand-off distances are shown in Figure
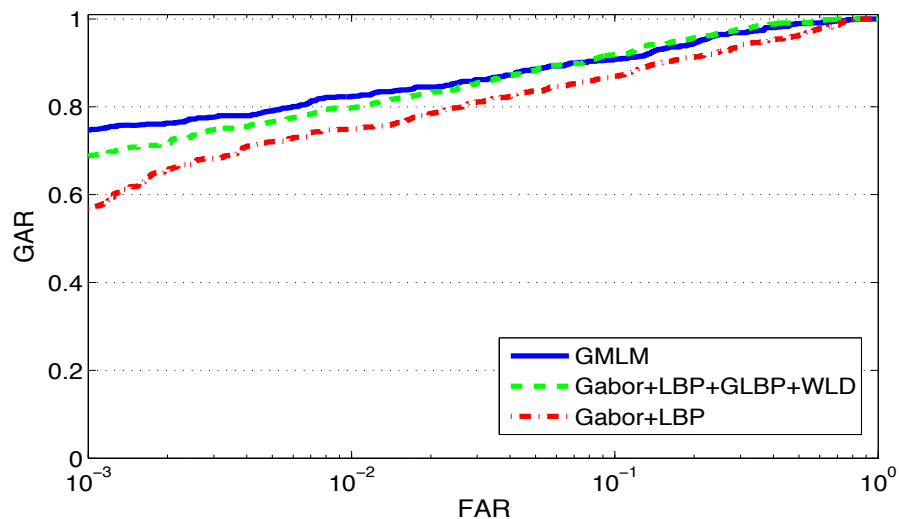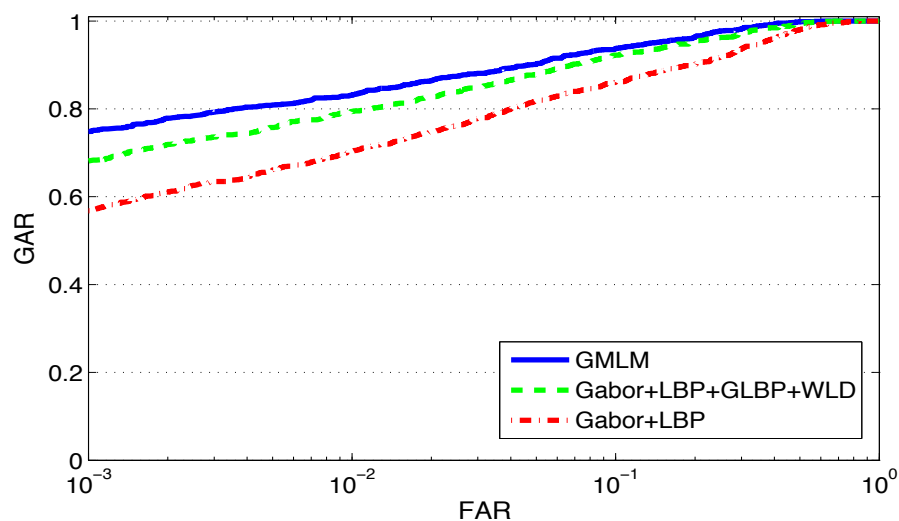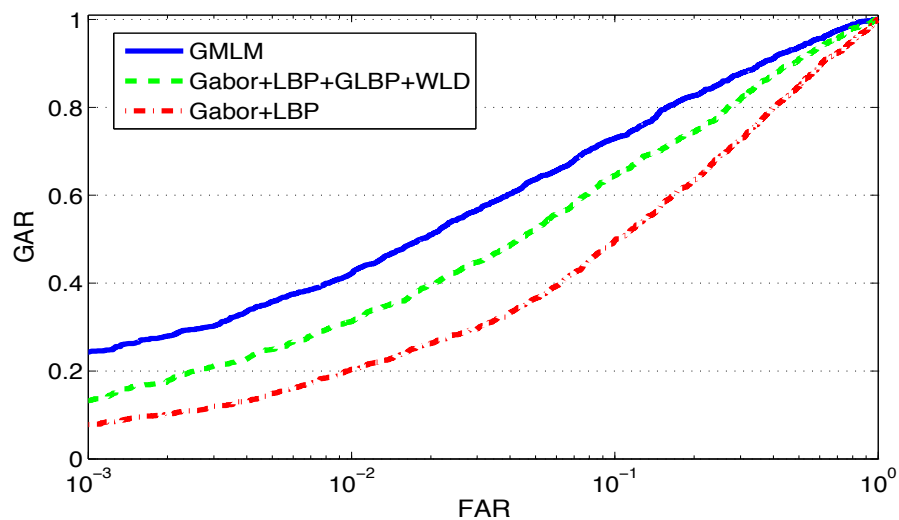
Figure 3.19: Matching NIR against visible light images using GMLM: (a) NIR 1.5 m, (b) NIR 50 m, and (c) NIR 106 m.

Table 3.22: GAR, EER and d-prime:Matching NIR images at 50 m against visible light images using GMLM.

METHOD	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER (%)	d-prime
Gabor+LBP	86.13	56.79	12.54	2.33
Gabor+LBP GLBP+WLD	92.23	68.21	8.71	2.66
GMLM	**93.70**	**74.86**	**7.38**	**2.92**

Table 3.23: GAR, EER and d-prime:Matching NIR images at 106 m against visible light images using GMLM.

METHOD	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER (%)	d-prime
Gabor+LBP	49.72	7.84	28.43	1.10
Gabor+LBP GLBP+WLD	64.48	13.28	23.24	1.49
GMLM	**72.95**	**24.29**	**18.13**	**1.79**

3.20.

For the only case of short standoff distance (the PCSO dataset), the performance of GMLM is slightly better than Gabor+LBP+GLBP+WLD and they both are better than Gabor+LBP. More matching parameters such EER, d-prime can be found in Table 3.24.

Table 3.24: GAR, EER and d-prime:Matching MWIR images against visible light images using GMLM.

METHOD	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER (%)	d-prime
Gabor+LBP	48.25	8.25	32.25	0.97
Gabor+LBP GLBP+WLD	53.50	**12.25**	29.01	1.18
GMLM	**57.00**	12.00	**26.75**	**1.22**
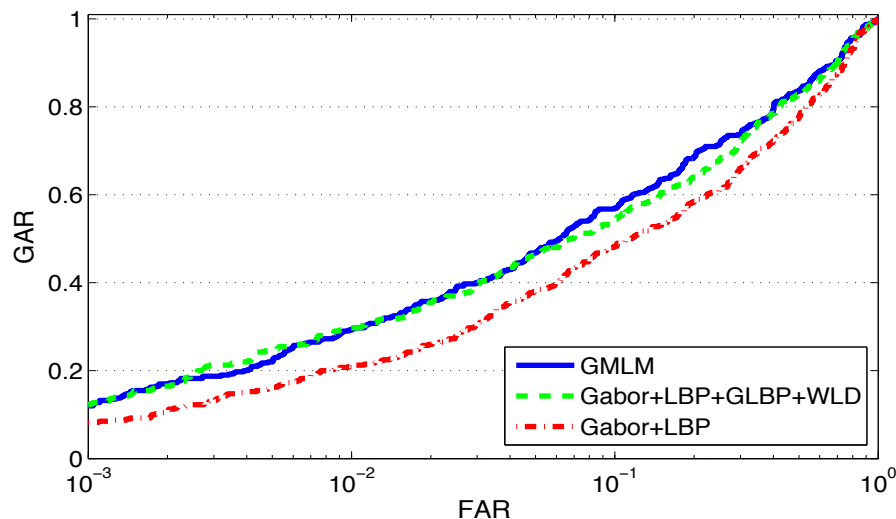
Figure 3.20: Matching MWIR against visible light images using GMLM.

Table 3.25: GAR, EER and d-prime:Matching LWIR images against visible light images using GMLM.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
Gabor+LBP	38.89	**6.48**	34.64	**0.74**
Gabor+LBP GLBP+WLD	**40.74**	5.93	**34.08**	0.71
GMLM	39.44	4.63	34.25	0.66

Matching LWIR to Visible Images

The last experiment is conducted to match LWIR face probes against the visible light face gallery. The other two operators, Gabor+LBP+GLBP+WLD and Gabor+LBP, serve as benchmarks. The results of matching parameterized by different stand-off distances are shown in Figure 3.21.

For the only case of standoff distance at 2 m (the Q-FIRE dataset), the performance of all the three operators is very close to each other: all lower than the performance of their own in the SWIR, NIR and MWIR cases. More matching parameters such EER, d-prime can be found in Table 3.25.
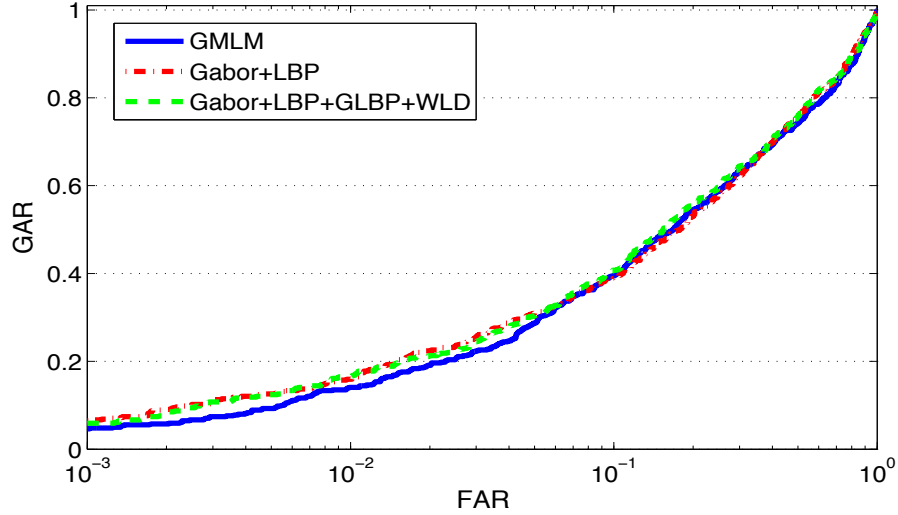
Figure 3.21: Matching LWIR against visible light images using GMLM.

Table 3.26: Comparison of computation time: Encoding and matching of SWIR probes at 1.5 m.

Method	Computation Time for Encoding (s)		Computation Time for Matching (s)	
	Mean	Std.	Mean	Std.
CMLD	10.2252	0.2328	0.8871	0.0250
GWLH	8.3186	0.0837	1.2698	0.0743
GMLM	13.2129	0.0883	1.4694	0.0441

3.5 Summary

This chapter focuses on designing local feature-based operators. Three new operators, CMLD, GWLH and GMLM, are proposed for extraction and encoding of facial features. After implementing the three operators, we present and analyze the experimental results of matching SWIR, NIR or MWIR facial images to visible light images. Different standoff distances varying from short (1.5 m) to long (50 m and 106 m) are considered.

The new operators outperform other popular operators such as LBP, HOG, Gabor followed by LBP (Gabor+LBP), and GOM. As the standoff distance increases, the matching performance of all operators drops. This drop is attributed to a relatively low quality of SWIR and NIR face images at long standoff distances.

Computation time – in addition to matching performance, is also evaluated and compared among the three newly proposed operators (See Section 2.4 of Chapter 2 for more details). An example for the case of encoding and matching SWIR 1.5 m probes is provided in Table 3.26.

Chapter 4

Cross-Spectral Periocular Recognition

This chapter is dedicated to a new problem termed *cross-spectral periocular recognition*. Due to its advantages over other biometric modalities (such as face and iris) in some special aspects, periocular recognition has attracted attention from researchers but mostly in the intra-spectral case. In this chapter we consider the cross-spectral case of periocular recognition which gains its own importance in some special application scenarios. We look into the feasibility of using the periocular region as a modality in a cross-spectral recognition context. We experiment with the usage of two operators that we have recently developed for the purpose of cross-spectral face recognition, namely CMLD and GWLH (See Chapter 3). To demonstrate the advantages of the two operators, we compare them with both basic and state-of-the-art methods. The cross matching performance of the two operators is demonstrated on periocular datasets generated from three heterogeneous face datasets with images acquired in NIR, SWIR, MWIR and visible light spectra. Both short and long standoff distances are considered.

The outline of this chapter is as follows. Section 4.1 provides an introduction to the new problem of cross-spectral periocular recognition. Section 4.2 defines the periocular region used in this chapter and describes the datasets involved. Section 4.3 describes the general structure of the recognition system. Section 4.4 presents experimental results when using some baseline algorithms for this new problem. In Section 4.5 we present the performance evaluation of using our newly proposed operator CMLD while in Section 4.6 we present the performance evaluation for another new operator GWLH. The final section 4.7 summarizes

the observations and conclusions of this chapter.

4.1   Introduction

Periocular recognition has been an active research area in the past several years [107–109]. Periocular regions can be categorized as a part of face, perhaps most visible in the presence of face occlusion (See Figure 4.1). On the other hand, it can add to reliable iris recognition (in visible and NIR spectra), especially when the quality of iris images is low. Although many research challenges such as unconstrained subject's presentation, uneven illumination, and partial occlusions have been previously addressed in the literature [107, 110–112], many challenges remain. Furthermore, as new practical applications evolve, new challenges offered by the applications arise and hence a need for development of new algorithms to mitigate them.
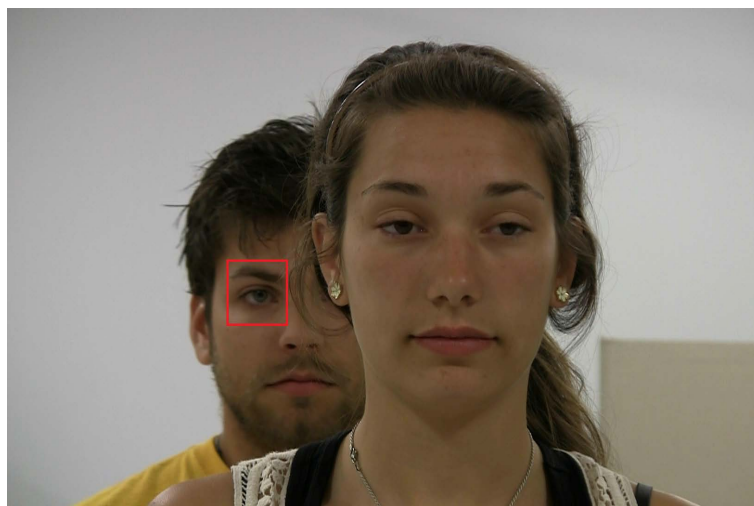
Figure 4.1: Advantage of periocular recognition under face occulsion.

Surveillance at night or in harsh environments is one of the most recent applications. The latest advances in manufacturing of small and cheap imaging devices sensitive in the active infrared range (NIR and SWIR) [30, 31] and the ability of these cameras to see through fog, rain, at night and operate at long ranges provided researchers with a new type of imagery and posed new research problems [35–38, 105, 113]. As observed, active IR energy is less affected by scattering and absorption by smoke or dust than visible light. Also, unlike visible

spectrum imaging, active IR imaging can be used to extract not only exterior but also useful subcutaneous anatomical information. This results in a very different appearance of images in active IR range compared to image in visible spectrum. Acknowledging these differences, many related questions can be posed. What type of information should be extracted from active IR images to successfully solve the problem of periocular recognition? How can we match a periocular image in the active IR or MWIR spectral band to a periocular image in the visible light band? The latter falls in the scope of heterogeneous periocular recognition.

However, results of heterogeneous matching of periocular regions have been barely reported so far [39, 40, 114], which motivates the research work conducted in this chapter. If addressed, heterogeneous matching of periocular regions will provide a baseline for the development of new improved algorithms for heterogeneous face and periocular recognition at night or in challenging environments. In addition to cross-spectral matching we explore the effect of varying standoff distances on the recognition performance.

A few publications on cross-spectral face recognition have appeared in the literature. Most of them were focused on algorithms for and analyses of matching NIR, SWIR, MWIR or LWIR face images to a gallery of visible face images [38, 83, 101, 115–117]. Some publications assumed short standoff distances, while others explored the case of varying standoff distances [35, 38, 115]. Popular algorithms such as Local Binary Pattern (LBP), Scale-Invariant Feature Transform (SIFT), Histogram of Oriented Gradients (HOG) [70], Gabor Ordinal Measures (GOM) [91] and their variants have been used for feature extraction and matching in the past [118]. In this work, we use three composite operators, Gabor+WLD+LBP+GLBP, GWLH and CMLD (see Chapter 2), and compare them with other individual and composite operators. We demonstrate that the three operators can be adapted to successfully match heterogeneous periocular regions at short (1.5 m) and long (50 and 106 m) standoff distances.

4.2   Periocular Region and Datasets

The periocular region is defined as the rectangular region centered around an eye in our work. For simplification, we only choose the right eye for usage in the experiments. Figure 4.2 gives an illustration of the defined geometry of the periocular region.
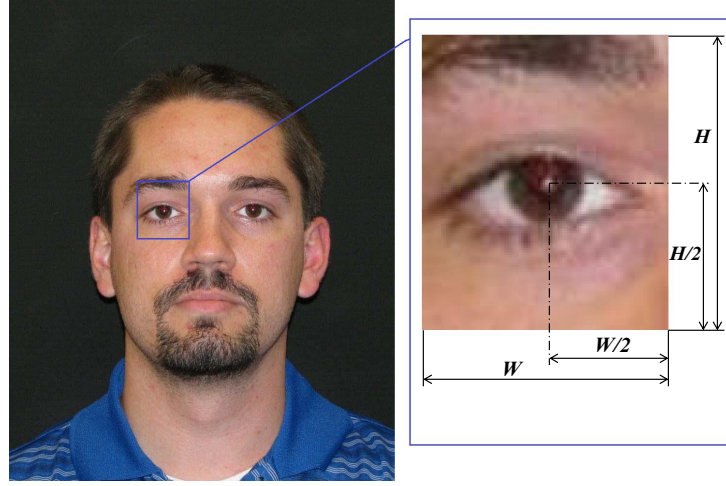
Figure 4.2: Definition of the periocular region

In our experiments we use four face datasets. The first two datasets, Pre-TINDERS (Tactical Imager for Night/Day Extended-Range Surveillance) and TINDERS are collected by the Advanced Technologies Group, West Virginia High Tech Consortium (WVHTC) Foundation. A detailed description of the two datasets is provided in Section 2.3 from Chapter 2. The third dataset is collected by Pinellas County Sheriff's Office (PCSO) [119]. The fourth dataset Quality-Face/Iris Research Ensemble (Q-FIRE) is collected by Clarkson University [120].

The PCSO dataset is composed of color and MWIR images of 1000 subjects. Images are acquired at a short standoff distance of 1.5 m. Each class is represented by two MWIR images collected in two different sessions and one color image. The resolution of all images in both visible and MWIR spectral bands is 620×480. Images collected in both bands are in JPEG format.

The Q-FIRE dataset comprises color and LWIR face images of 82 subjects. Images are acquired at a short standoff distance of 2 m. Each class is represented by two or four color and two or four LWIR images collected in two different visits. The resolution of all images in the visible band is 1920×1080 while the resolution for the MWIR spectral band is 726×480. All images are in BMP format.

A summary of all the four datasets can be found in Table 3.17 from Chapter 3. Sample

periocular images cropped from the original face images in the PCSO and Q-FIRE datasets are shown in Figure 4.3.
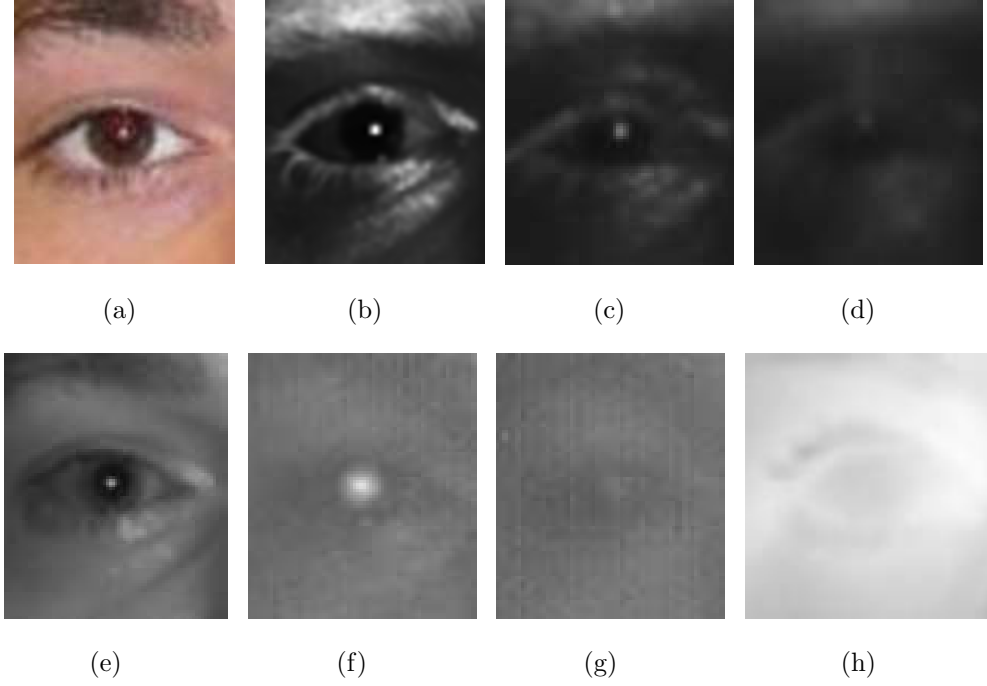
Figure 4.3: Sample periocular images cropped from the original face images in the four datasets: (a) visible light, (b) SWIR 1.5 m, (c) SWIR 50 m, (d) SWIR 106 m, (e) NIR 1.5 m, (f) NIR 50 m, (g) NIR 106 m and (h) LWIR 2 m.

4.3   Recognition System Framework

4.3.1   Preprocessing and Matching

The preprocessing and matching steps are the same as those used for cross-spectral face recognition (see Section 2.2 of Chapter 2). Preprocessing steps include image alignment, cropping and a simple intensity normalization. Geometric transformations are applied to all original face images and then alignment using the position of the eyes and nose is implemented. Fig. 4.4 (a)-(d) gives an illustration of the preprocessing step.

During matching, each encoded response of an operator is divided into non-overlapping square blocks. Then histograms of each block are normalized and concatenated, resulting in a template of features. The distance (the matching score) between two images is evaluated

(a)   (b)   (c)   (d)
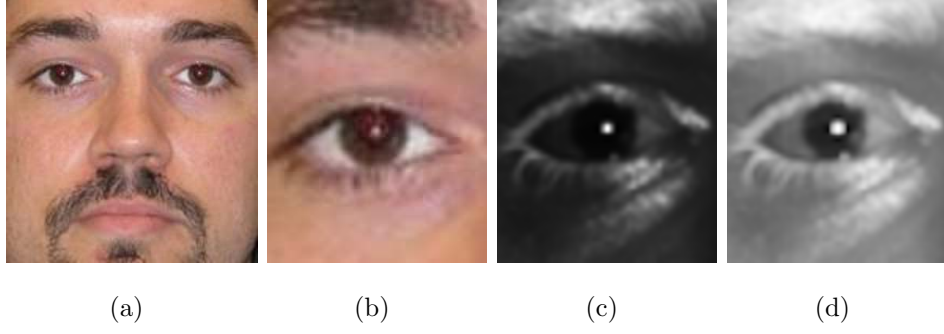
Figure 4.4: Preprocessing of the periocular region: (a) aligned and cropped face, (b) aligned and cropped eye, (c) original eye in SWIR and (d) log-transformed eye.

as a sum of symmetric I-divergence.

4.3.2   Feature Extraction

The next step after preprocessing is feature extraction. In this chapter, we employ operators Gabor+WLD+LBP+GLBP, CMLD, GWLH and compare them with other basic operators. The details of CMLD, GWLH can be found in Section 3.3 and Section 3.2 of Chapter 3, respectively. An example of the feature extraction step using GWLH is given in Fig. 4.5.

(a)   (b)   (c)   (d)   (e)
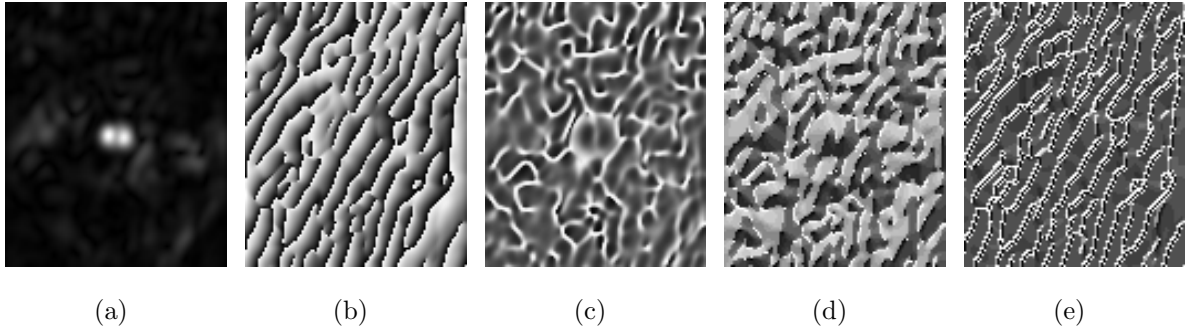
Figure 4.5: Filtering and encoding of a periocular image: (a) - (b) the magnitude and phase of a Gabor filter response at scale 3 and orientation 0°, respectively; (c) - (e) the results of encoding Gabor magnitude with WLD, Gabor magnitude with LBP, and Gabor phase with LBP, respectively (all at radius 1).

4.4   Cross-Spectral Periocular Recognition Using Gabor+WLD+LBP+GLBP

The combined operator Gabor+WLD+LBP+GLBP adopts a Gabor filter-based approach at the initial encoding stage, followed by an encoding scheme that involves three operators (as described in Section 4.3.2) – WLD, LBP and GLBP to extract robust features across different spectral bands. These three operators are designed to encode both magnitude and phase of filtered images. Details of this scheme can be found in [38].

4.4.1   Matching SWIR against visible

Our first experiment involves matching SWIR periocular regions to visible periocular regions. The heterogeneous images are encoded using three algorithms: (1) Gabor+Weber+LBP+GLBP (our algorithm), (2) the original LBP and (3) the GOM-based algorithm [91]. The performance of the LBP is used as a benchmark. The results of matching are displayed as Receiver Operating Characteristic (ROC) curves in Figure 4.6. Visible light images form the gallery. All SWIR images are involved as the test images. The results are shown for right eye only. The curves are parameterized by three standoff distances. Note that for the case of short standoff distance (Pre-TINDERS dataset) the value of Genuine Accept Rate (GAR) is 0.75 at False Accept Rate (FAR) set to 0.01 for our algorithm compared to the GAR of 0.28 and 0.56 at the same FAR for LBP and GOM, respectively. This comparison clearly demonstrates the advantage of using our algorithm over both LBP and GOM for short standoff distances. However, as the standoff distance increases to 50 m and then to 106 m, the difference in matching performance of the three algorithms diminishes too. Performance of all algorithms drops significantly.

Note that the useful information that helps matching heterogeneous images is contained in the gradients of intensity and in their relative distribution, that is, in their relative location and their density. Due to long standoff distances, the overall quality of SWIR images in the TINDERS dataset is reduced. This, in turn, affects the quality of informative gradients and their distribution in SWIR images compared to visible images or SWIR images in the Pre-

TINDERS dataset. As demonstrated, these low quality gradients can be detected equally well (or equally poorly) by LBP, GOM and our algorithm.
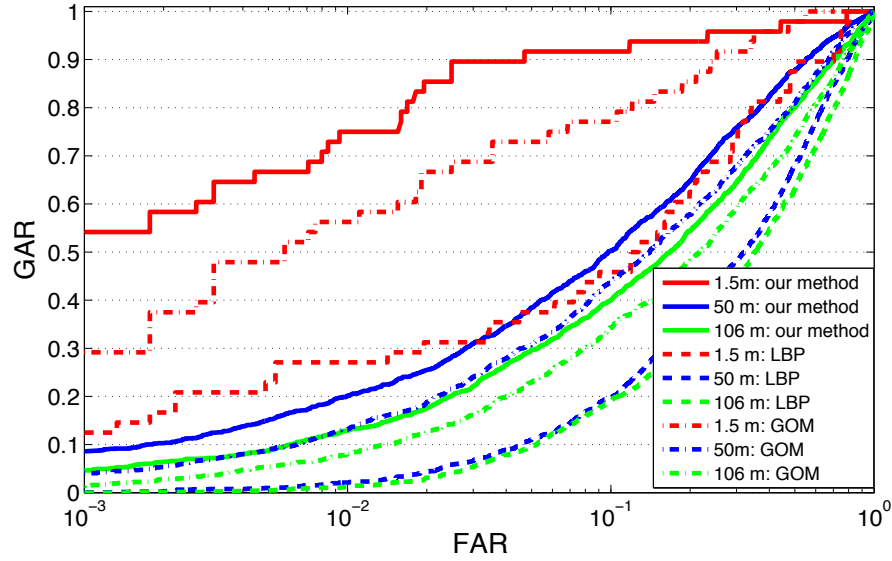
Figure 4.6: Matching SWIR vs. visible. Short and long standoff distances.

4.4.2   Matching NIR against visible

The results of matching NIR periocular regions to visible periocular regions are shown in Figure 4.7. Two standoff distances are tested: 50 m and 106 m. At 50 m, our algorithm reaches a GAR value of 0.44 at FAR set to 0.1 while LBP and GOM reach GAR values of 0.37 and 0.40 at the same value of FAR, respectively. When the standoff distance increases to 106 m, our algorithm, LBP and GOM produce GAR values of 0.22, 0.15 and 0.17, respectively, at FAR equal to 0.1. Again, this experiment demonstrates that a small region of the face becomes a weak identifier for the case of heterogeneous face matching. Apart from this, NIR images at large standoff distances are extremely noisy due to insufficient illumination (see Fig. 4.3). This requires taking the quality of the images into account.

4.4.3   Matching MWIR against visible

In our last experiment, we match MWIR periocular regions to visible periocular regions. The results shown in Figure 4.8 are for 200 periocular classes (right eye only). Color images
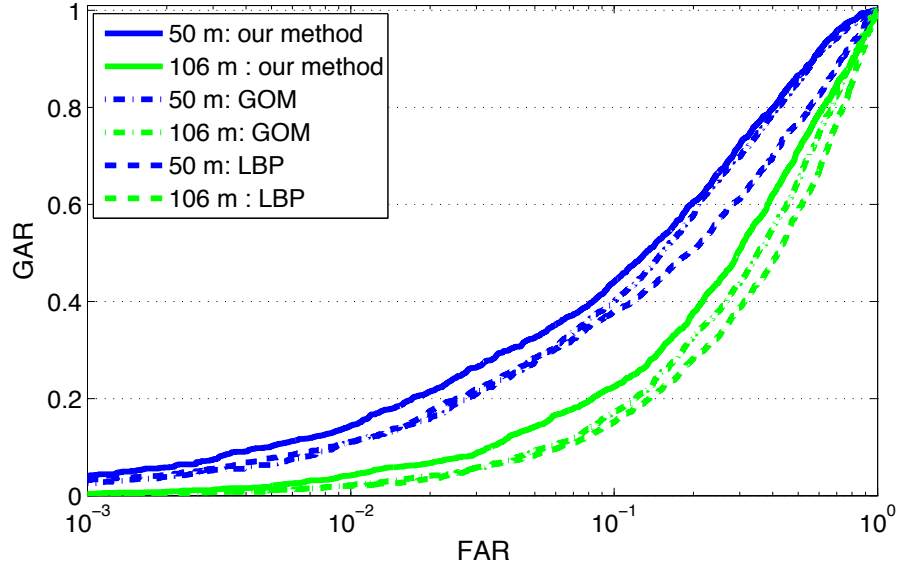
Figure 4.7: Matching NIR vs. visible. Long standoff distance.

constitute the gallery while MWIR images (two per class) are test images. Again, LBP, GOM and our algorithm are not designed to deal with such a large spectral gap between MWIR and visible light. Due to very distinct imaging nature, MWIR and visible periocular regions do not contain much intensity distribution and edge information in common. Nevertheless, we achieved GAR of 0.35, 0.24 and 0.36 at FAR set to 0.1 by applying our algorithm, LBP and GOM respectively.

4.5 Cross-Spectral Periocular Recognition Using CMLD

In this section we utilize the operator CMLD [105] proposed for cross-spectral face recognition (refer to Chapter 3 for details) to study the problem of cross-spectral periocular recognition and present the numerical results and analysis of matching heterogeneous periocular regions. Prior to matching, heterogeneous periocular regions are extracted from the heterogeneous face images as mentioned in Section 4.3.1. To demonstrate the advantage of our CMLD operator, we compare it with two other operators – one individual and one compound: (1) LBP and (2) Gabor+HOG+LBP (or GLH). The performance of the three considered schemes is displayed as ROC curves in Fig. 4.9 - Fig. 4.12.
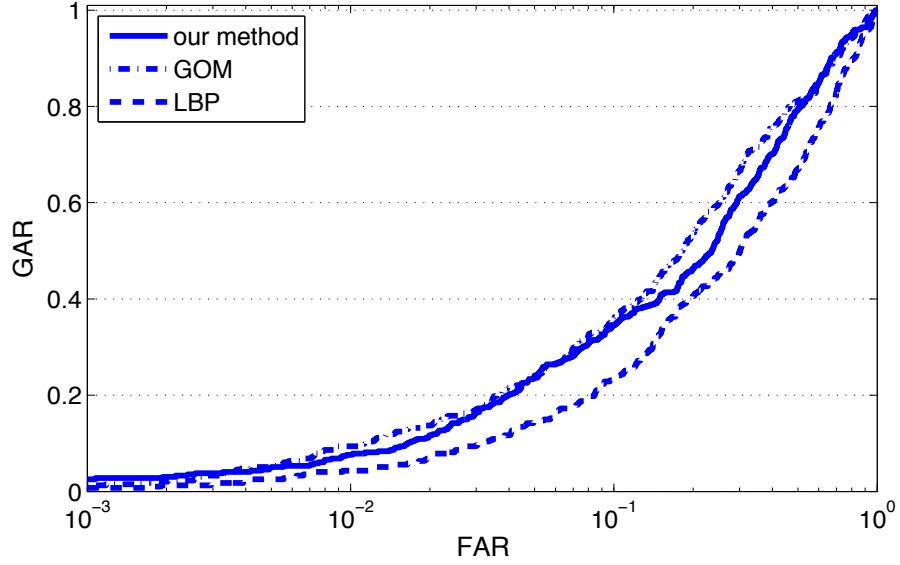
Figure 4.8: Matching MWIR vs. visible. Short standoff distance.

4.5.1   Matching SWIR to visible images

Our first experiment involves matching SWIR periocular images to visible periocular images. Visible light images form the gallery while all SWIR images are involved as test images. Both the SWIR and visible images are encoded using the seven algorithms mentioned above. The results of matching are shown in Fig. 4.13 (a) - (c) for the standoff distances 1.5 m, 50 m and 106 m, respectively. A summary of GARs, EERs and d-prime values are shown in Tables 4.1, 4.2 and 4.3.

The comparison in each of the three cases of standoff distances clearly demonstrates the advantage of using our operator, CMLD, over the other two aoperators. However, as the standoff distance increases to 50 m and then to 106 m, the difference in matching performance of our algorithm and the others diminishes too. Performance of all algorithms drops significantly.

Note that the useful information that helps matching heterogeneous images is contained in the gradients of intensity and in their relative distribution, that is, their relative location and their density. Due to long standoff distances the overall quality of SWIR images in TINDERS dataset is reduced. This, in turn, affects the quality of informative gradients and their distribution in SWIR images compared to visible images or SWIR images in the

Pre-TINDERS dataset.

Table 4.1: GARs, EERs and d-prime values: matching SWIR probes at 1.5 m to visible gallery using CMLD.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
LBP	50.26	13.28	27.87	1.22
GLH	82.55	40.23	14.07	2.23
CMLD	**93.62**	**46.09**	**7.83**	**2.61**

Table 4.2: GARs, EERs and d-prime values: matching SWIR probes at 50 m to visible gallery using CMLD.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
LBP	19.82	0.100	40.98	0.46
GLH	38.52	6.65	32.01	0.96
CMLD	**44.89**	**5.32**	**31.09**	**1.01**

Table 4.3: GARs, EERs and d-prime values: matching SWIR probes at 106 m to visible gallery using CMLD.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
LBP	19.24	0.100	42.80	0.32
GLH	30.76	3.13	35.49	0.75
CMLD	**34.44**	**1.74**	**35.40**	**0.70**

4.5.2   Matching NIR to visible images

The results of matching NIR periocular images to visible periocular images are shown in Fig. 4.10 (a) - (c). The analyzed standoff distances are 1.5 m, 50 m and 106 m. A summary of the GARs, EERs and d-prime values are shown in Table 4.4, Table 4.5 and Table 4.6 for each of the considered cases. The results demonstrate the advantage of CMLD over the other two operators in the case of short standoff distance but not at longer standoff distances. Again,

the performance of all algorithms drops in the cases of long standoff distances. This requires taking the quality of images into account, since NIR images at large standoff distances are extremely noisy due to insufficient illumination (see Fig. 4.3).

Table 4.4: GARs, EERs and d-prime values: matching NIR probes at 1.5 m to visible gallery using CMLD.

METHOD	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER (%)	d-prime
LBP	95.18	45.83	7.03	2.89
GLH	89.97	56.12	10.02	2.54
CMLD	**98.31**	**73.70**	**4.52**	**3.22**

Table 4.5: GARs, EERs and d-prime values: matching NIR probes at 50 m to visible gallery using CMLD.

METHOD	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER (%)	d-prime
LBP	44.12	4.13	28.86	1.14
GLH	26.89	2.73	40.23	0.52
CMLD	**33.54**	**2.24**	**33.52**	**0.87**

Table 4.6: GARs, EERs and d-prime values: matching NIR probes at 106 m to visible gallery using CMLD.

METHOD	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER (%)	d-prime
LBP	**22.46**	0.42	**39.01**	**0.49**
GLH	12.43	0.10	46.27	0.18
CMLD	**15.25**	**0.07**	44.12	0.31

4.5.3   Matching MWIR to visible images

In our third experiment, we match MWIR periocular images to a gallery of visible periocular images. The results shown in Fig. 4.11 are for 200 periocular classes. Color images (one per class) form the gallery while MWIR images (two per class) serve as the probes. It

is clear that both our algorithm and the other algorithms are not designed to deal with such a large spectral gap. MWIR images display the distribution of heat in a subject's periocular region while visible images characterize the reflectivity properties of subject's skin. Again, the only common information for heterogeneous matching lies at the edge gradients and their relative distribution and density. MWIR and visible periocular regions do not share much of this type of information. Nonetheless, CMLD is shown to be of higher performance than the other two operators. A summary of the GARs, EERs and d-prime values is shown in Table 4.7.

Table 4.7: GARs, EERs and d-prime values: matching MWIR probes to visible gallery using CMLD.

METHOD	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER (%)	d-prime
LBP	23.60	0.76	39.85	0.48
GLH	23.86	0.76	40.07	53.51
CMLD	**32.49**	**1.02**	**34.26**	**0.78**

4.5.4 Matching LWIR to visible images

In our last experiment, we match LWIR periocular images to a gallery of visible periocular images. The results of cross-spectral matching are shown in Fig. 4.12. Visible images form the gallery while LWIR images serve as the probes. As expected, the largest spectral gap considered results in quite degraded performance of matching heterogeneous periocular regions. LWIR and visible periocular regions do not contain significant edge information in common. Nonetheless, CMLD is shown to more powerful than the other two operators. A summary of the GARs, EERs and d-prime values is shown in Table 4.8.

4.6 Cross-Spectral Periocular Recognition Using GWLH

This section presents numerical results and an analysis of matching heterogeneous periocular regions using GWLH [121] (Refer to Chapter 3 for details). Prior to matching,

Table 4.8: GARs, EERs and d-prime values: matching MWIR probes to visible gallery using CMLD.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
LBP	23.60	0.76	39.85	0.48
GLH	23.86	0.76	40.07	53.51
CMLD	**32.49**	**1.02**	**34.26**	**0.78**

heterogeneous periocular regions are extracted from the heterogeneous face images as mentioned in Section 4.3.1. To demonstrate the advantage of our GWLH operator, we compare it with six other operators: (1) LBP, (2) HOG, (3) Gabor filter, (4) HOG+LBP, (5) Gabor+HOG+LBP (or GLH), and (6) Gabor+WLD+LBP+GLBP. The first three operators are applied as single operators, while the latter three are compound operators fused using different combinations of the three simple operators and WLD. The performance of the seven considered schemes is displayed as ROC curves in Figs. 4.6 - Fig. 4.8.

4.6.1 Matching SWIR to visible images

Our first experiment involves matching SWIR periocular images to visible periocular images. Visible light images form the gallery while all SWIR images are involved as test images. Both the SWIR and visible images are encoded using the seven algorithms mentioned above. The results of matching are shown in Fig. 4.13 (a) - (c) for the standoff distances 1.5 m, 50 m and 106 m, respectively. A summary of GARs, EERs and d-prime values are shown in Tables 4.9, 4.10 and 4.11.

The comparison in each of the three cases of standoff distances clearly demonstrates the advantage of using our algorithm, GWLH, over the other six algorithms. However, as the standoff distance increases to 50 m and then to 106 m, the difference in matching performance of our algorithm and the others diminishes too. The performance of all algorithms drops significantly.

Table 4.9: GARs, EERs and d-prime values: matching SWIR probes at 1.5 m to visible gallery.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
LBP	50.26	13.28	27.87	1.22
Gabor	62.50	13.80	22.93	1.38
HOG	64.58	13.93	20.96	1.56
HOG+LBP	52.34	14.06	25.14	1.34
GLH	82.55	40.23	14.07	2.23
Gabor+WLD LBP+GLBP	91.67	57.42	8.99	2.62
GWLH	**93.88**	**64.19**	**7.32**	**2.73**

Table 4.10: GARs, EERs and d-prime values: matching SWIR probes at 50 m to visible gallery.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
LBP	19.82	0.100	40.98	0.46
Gabor	26.05	2.31	40.45	0.44
HOG	28.99	1.26	35.44	0.63
HOG+LBP	22.06	0.35	37.81	0.57
GLH	38.52	6.65	32.01	0.96
Gabor+WLD LBP+GLBP	50.21	**8.61**	26.71	1.21
GWLH	**56.23**	7.56	**24.87**	**1.33**

Table 4.11: GARs, EERs and d-prime values: matching SWIR probes at 106 m to visible gallery.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
LBP	19.24	0.100	42.80	0.32
Gabor	20.49	1.74	45.21	0.27
HOG	22.29	0.69	40.15	0.39
HOG+LBP	18.89	0.21	42.27	0.37
GLH	30.76	3.13	35.49	0.75
Gabor+WLD LBP+GLBP	40.07	4.65	32.70	0.85
GWLH	**44.58**	**5.83**	**31.18**	**0.98**

4.6.2   Matching NIR to visible images

The results of matching NIR periocular images to visible periocular images are shown in Fig. 4.14 (a) - (c). The analyzed standoff distances are 1.5 m, 50 m and 106 m. A summary of the GARs, EERs and d-prime values are shown in Table 4.12, Table 4.13 and Table 4.14 for each of the considered cases. Again, this experiment demonstrates the advantage of GWLH over other operators, in general, although the performance of all algorithms drops considerably in the case of long standoff distances. This requires taking the quality of images into account, since NIR images at large standoff distances are extremely noisy due to insufficient illumination (see Fig. 4.3).

4.6.3   Matching MWIR to visible images

In our next experiment, we match MWIR periocular images to a gallery of visible periocular images. The results shown in Fig. 4.15 are for 200 periocular classes. Color images (one per class) form the gallery while MWIR images (two per class) serve as the probes. It is clear that both our algorithm and other algorithms are not designed to deal with such a large spectral gap which causes dramatically different appearance in MWIR and visible light images due to different imaging mechanism (as explained in Section 4.5.3). However, it is consistant with SWIR and NIR spectra that GWLH has advantage over all other operators.

Table 4.12: GARs, EERs and d-prime values: matching NIR probes at 1.5 m to visible gallery.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
LBP	95.18	45.83	7.03	2.89
Gabor	54.95	13.02	26.04	1.25
HOG	79.95	23.05	14.06	2.17
HOG+LBP	94.40	46.88	7.44	2.89
GLH	89.97	56.12	10.02	2.54
Gabor+WLD LBP+GLBP	95.70	69.53	6.25	2.95
GWLH	**98.05**	**79.04**	**4.42**	**3.18**

Table 4.13: GARs, EERs and d-prime values: matching NIR probes at 50 m to visible gallery.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
LBP	44.12	4.13	28.86	1.14
Gabor	18.07	0.28	43.40	0.19
HOG	12.75	0.21	43.54	0.32
HOG+LBP	30.81	1.12	34.54	0.80
GLH	26.89	2.73	40.23	0.52
Gabor+WLD LBP+GLBP	37.96	3.15	34.51	0.81
GWLH	**54.97**	**9.17**	**25.71**	**1.27**

Table 4.14: GARs, EERs and d-prime values: matching NIR probes at 106 m to visible gallery.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
LBP	**22.46**	0.42	**39.01**	**0.49**
Gabor	10.02	0.10	49.20	0.01
HOG	10.31	0.10	48.38	0.07
HOG+LBP	15.54	0.10	43.97	0.28
GLH	12.43	0.10	46.27	0.18
Gabor+WLD LBP+GLBP	15.18	0.35	45.70	0.23
GWLH	21.82	**0.49**	42.84	0.39

A summary of the GARs, EERs and d-prime values is shown in Table 4.15.

Table 4.15: GARs, EERs and d-prime values: matching MWIR probes to visible gallery.

METHOD	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER (%)	d-prime
LBP	23.60	0.76	39.85	0.48
Gabor	14.47	0.25	47.68	0.16
HOG	21.57	0.51	39.84	0.44
HOG+LBP	24.87	0.51	39.60	0.51
GLH	23.86	0.76	40.07	53.51
Gabor+WLD LBP+GLBP	34.52	2.54	34.98	0.79
GWLH	**42.39**	**3.55**	**30.46**	**0.92**

4.6.4 Matching LWIR to visible images

In our last experiment, we match LWIR periocular images to a gallery of visible periocular images. The results of cross-spectral matching are shown in Fig. 4.16. Visible images form the gallery while LWIR images serve as the probes. As expected, the largest spectral gap considered results in quite degraded performance of matching heterogeneous periocular

regions. LWIR and visible periocular regions do not contain much of edge information in common. A summary of the GARs, EERs and d-prime values is shown in Table 4.16.

Table 4.16: GARs, EERs and d-prime values: matching LWIR probes to visible gallery.

METHOD	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER (%)	d-prime
LBP	8.52	0.37	50.37	0.05
Gabor	15.74	0.19	45.74	0.12
HOG	27.78	0.37	42.22	0.33
HOG+LBP	15.37	0.10	46.20	0.20
GLH	23.70	1.85	41.30	0.32
Gabor+WLD +LBP+GLBP	28.70	9.25	40.00	0.36
GWLH	**29.44**	**9.26**	**39.06**	**0.38**

4.6.5   Impact of Quality on Performance

As anticipated, the quality of active and passive IR (ie., SWIR, NIR, MWIR and LWIR) probes affects the matching performance. In this chapter, the quality of the probes is a function of the standoff distance. We use an adaptive sharpness measure [104] to calculate the image quality of the probes in SWIR, NIR, MWIR and LWIR spectra at all the three standoff distances, as shown in Table 4.17. From the results, the sharpness measure value decreases as standoff distance increases in both cases of SWIR and NIR spectra. This is in consistence with the visual perception of the quality of images in the datasets. The overall sharpness measure values of SWIR images are higher compared to the sharpness measure values of NIR images. It is further observed that although the matching performance of SWIR images at the short standoff distance is lower than that of NIR images in general, the performance of SWIR data degrades with increasing standoff distance slower than that of NIR data does.

After analyzing the matching results, we have made the following observations and conclusions:

Table 4.17: Sharpness measure of the SWIR, NIR, MWIR and LWIR images at different standoff distances

STATISTICS OF SHARPNESS MEASURE	SWIR 1.5 m	SWIR 50 m	SWIR 106 m	NIR 1.5 m	NIR 50 m	NIR 106 m	MWIR	LWIR
Mean	0.5835	0.5112	0.4391	0.4390	0.3910	0.3741	0.3496	0.2273
Standard Deviation	0.0707	0.0732	0.0730	0.0595	0.0461	0.0642	0.0633	0.0656

- The new operator GWLH substantially outperforms three single operators – LBP, HOG and Gabor filters – as well as other three state-of-the-art compound operators for heterogeneous periocular matching between SWIR, NIR, MWIR and LWIR spectra and visible light at both short and long standoff distances (except for NIR at 106 m).

- The performance of all the algorithms on SWIR vs visible and NIR vs visible decreases as the standoff distance increases due to the reduced image quality at longer distances.

- The gap between the matching performance of compound operators and that of simple operators grows proportionally with the increasing quality of heterogeneous data.

- The different imaging principles of MWIR (or LWIR) and visible face images resulted in seriously degraded performance for all involved algorithms. The spectral gap is very large.

4.7   Summary

This chapter raises a new problem of study: periocular recognition in a cross-spectral situation. It reviews recent research work on periocular recognition as well as discussing the advantage of using periocular recognition as a new modality over face recognition. It then addresses the new problem by utilizing two newly proposed operators, CMLD and GWLH, as the tool for feature extraction and compares them with other baseline algorithms. It further generates heterogeneous periocular datasets based on existing heterogeneous face datasets, and presents the results of matching SWIR, NIR and MWIR periocular probes to

a gallery of visible periocular images. Both short (1.5 m) and long (50 m and 106 m) standoff distances were considered. The new operators CMLD and GWLH substantially outperform the baseline algorithms LBP, HOG, Gabor filters and three other compound operators when applied to heterogeneous periocular regions collected at a short standoff distance in the case of the three spectral bands.

As the standoff distance increases (SWIR vs. visible and NIR vs. visible), the matching performance of the heterogeneous periocular images drops for all the algorithms. This drop is attributed to a relatively low quality of heterogeneous images at long standoff distances. When matching MWIR periocular regions to visible regions (only a short standoff distance is considered), our algorithm displayed a relatively low performance. In this case it is due to the different nature of MWIR and visible imagery: MWIR imagery measures the heat of a body while visible imagery measures reflected light.

The results of performance evaluation presented in this work can be used as reference for future study on this new topic of cross-spectral periocular recognition.
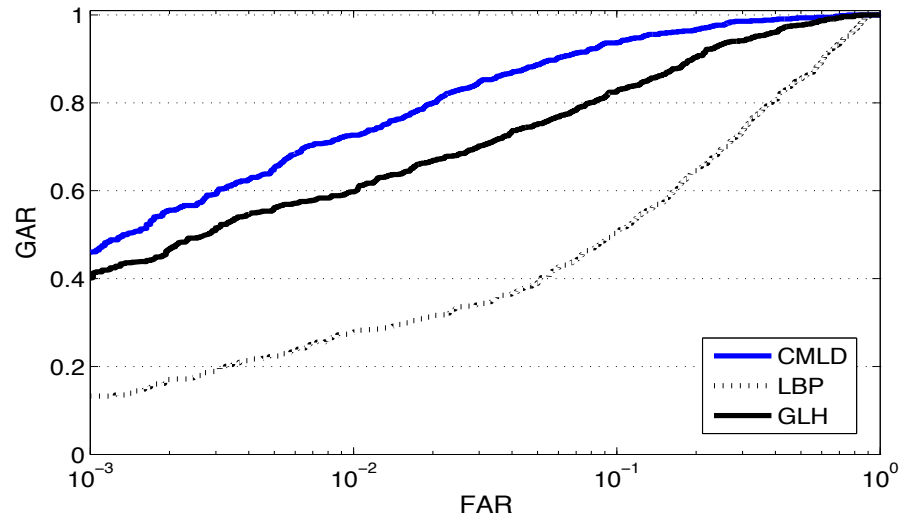
(a)
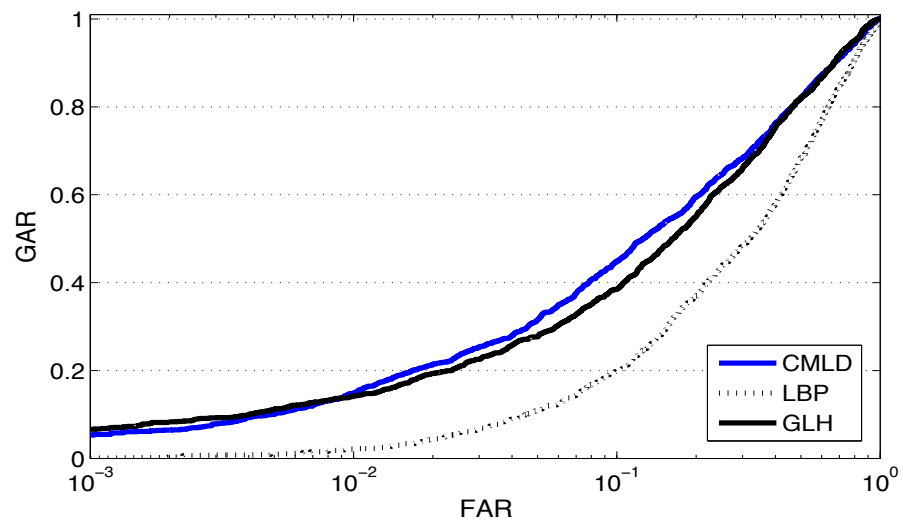
(b)
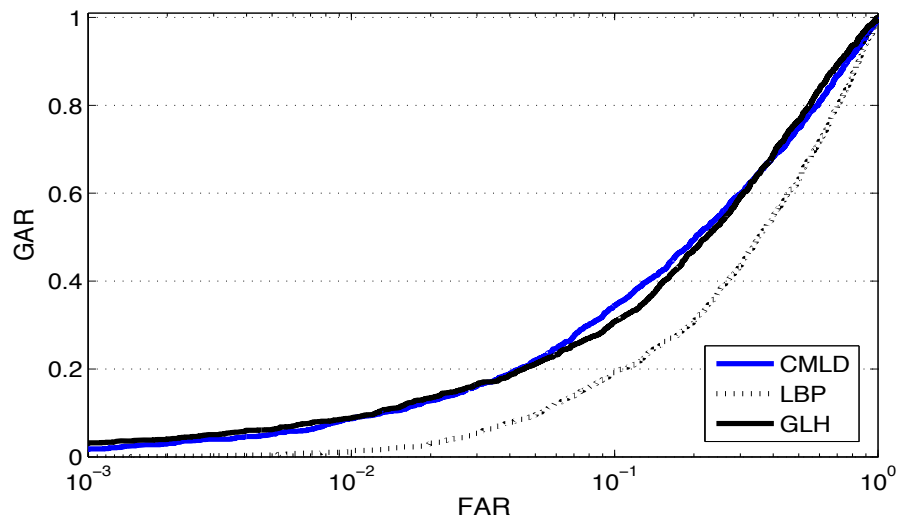
(c)

Figure 4.9: Cross-spectral matching of periocular regions using CMLD: (a) SWIR 1.5 m, (b) SWIR 50 m and (c) SWIR 106 m.
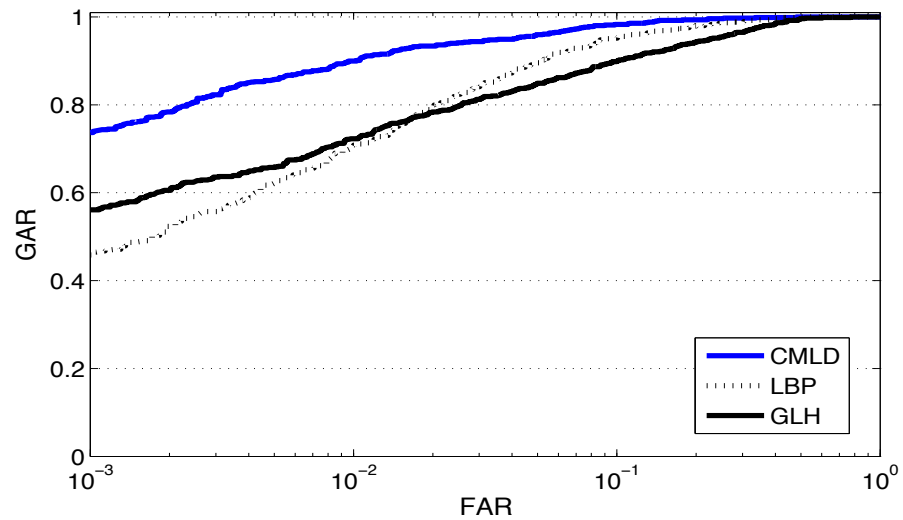
(a)
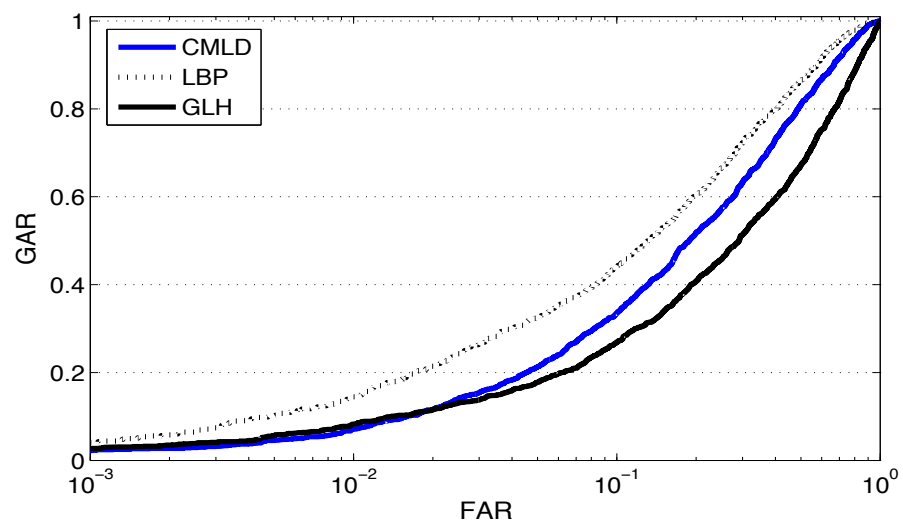
(b)
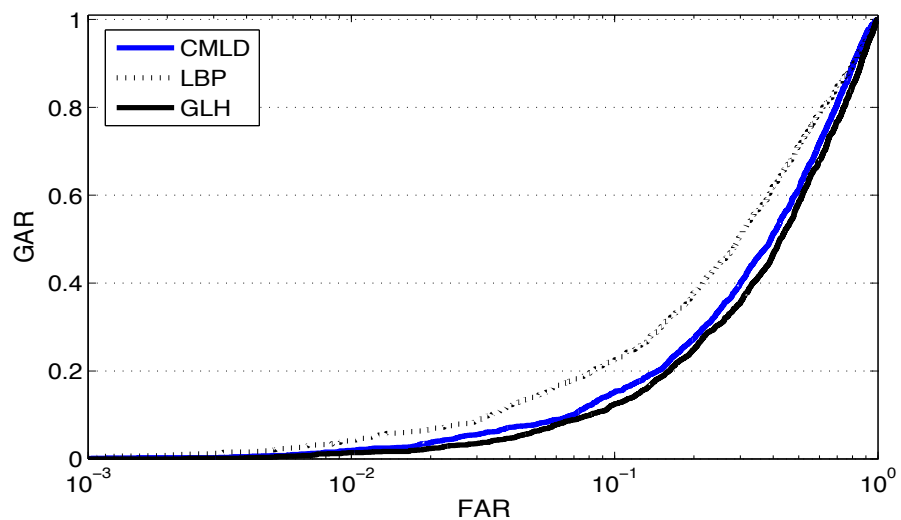
(c)

Figure 4.10: Cross-spectral matching of periocular regions using CMLD: (a) NIR 1.5 m, (b) NIR 50 m and (c) NIR 106 m.
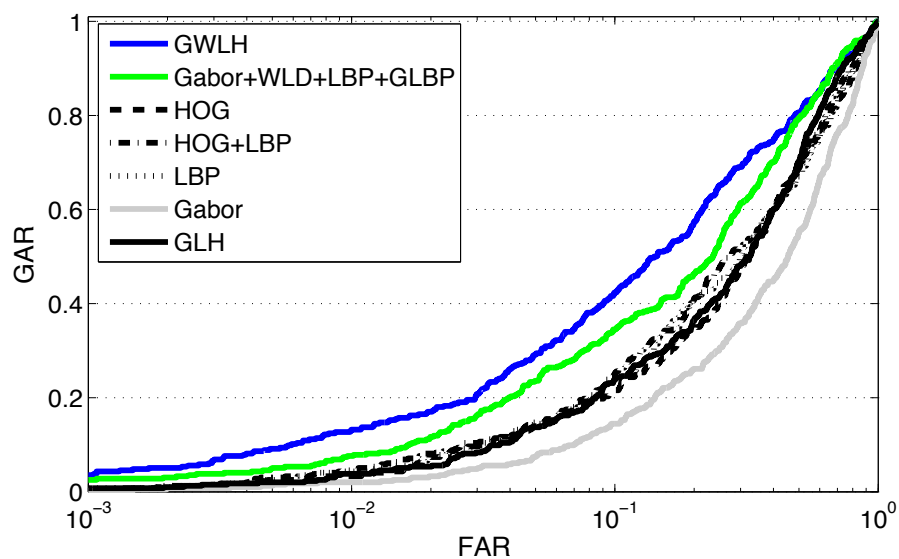
Figure 4.11: Cross-spectral matching of periocular regions using CMLD: MWIR.
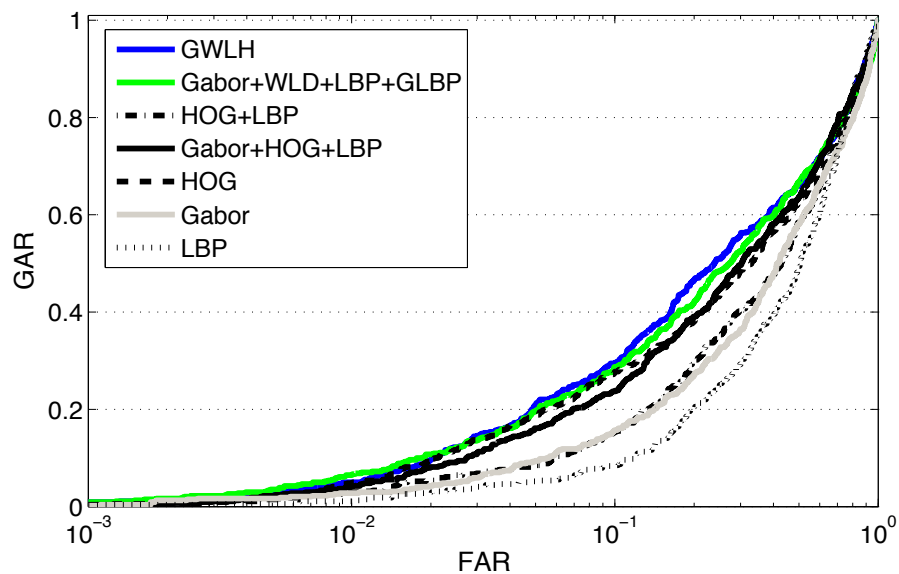
Figure 4.12: Cross-spectral matching of periocular regions using CMLD: LWIR.
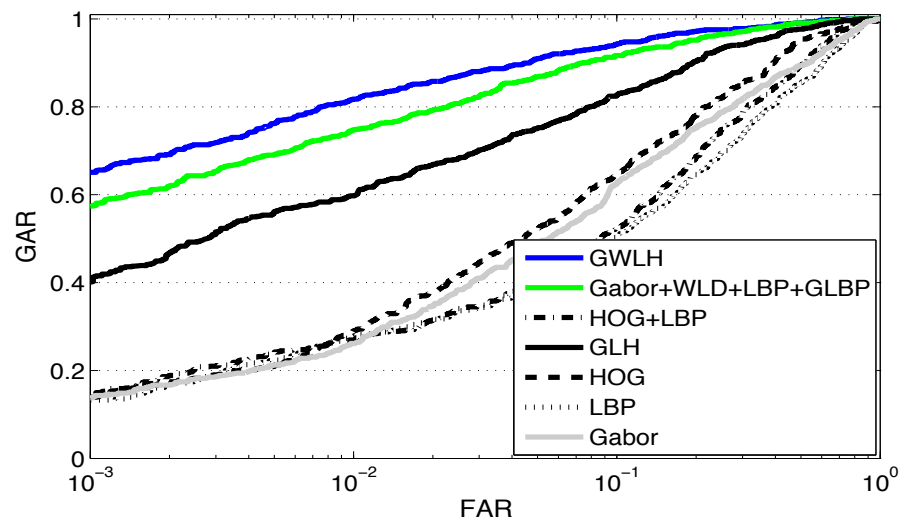
(a)
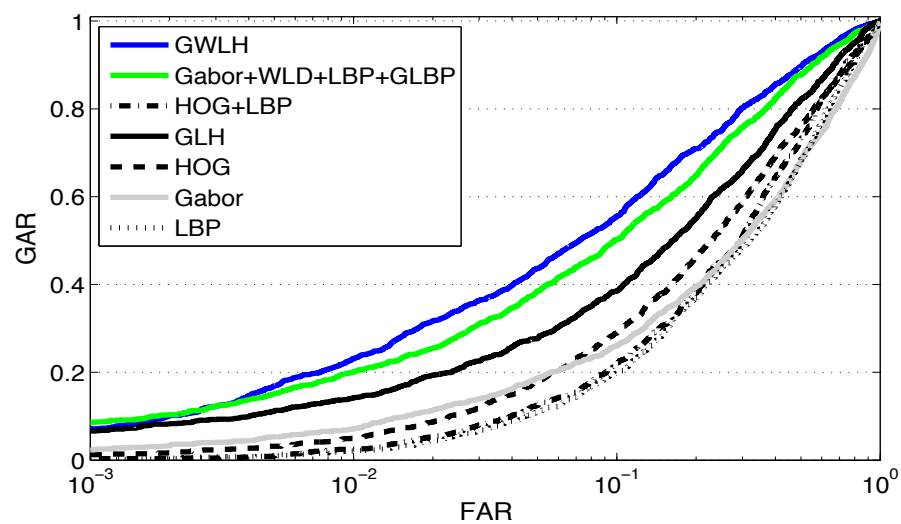
(b)
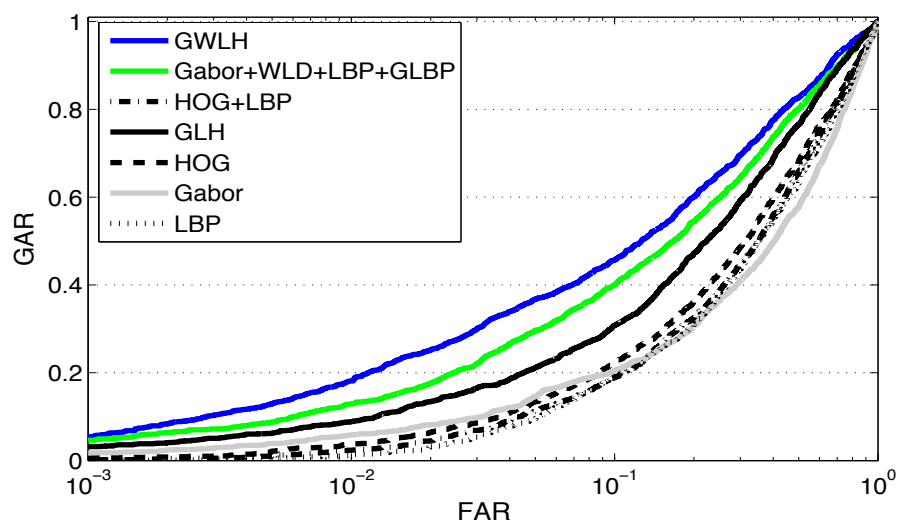
(c)

Figure 4.13: Cross-spectral matching of periocular regions using GWLH: (a) SWIR 1.5 m, (b) SWIR 50 m and (c) SWIR 106 m.
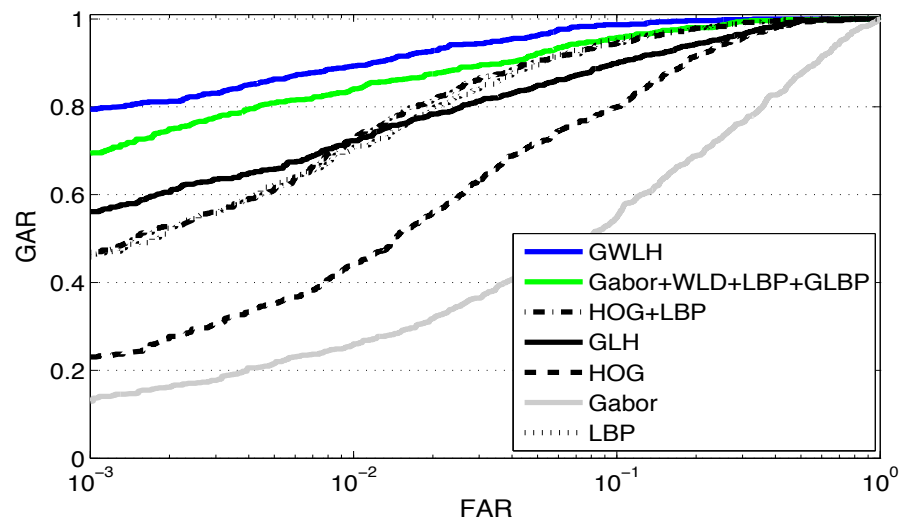
(a)
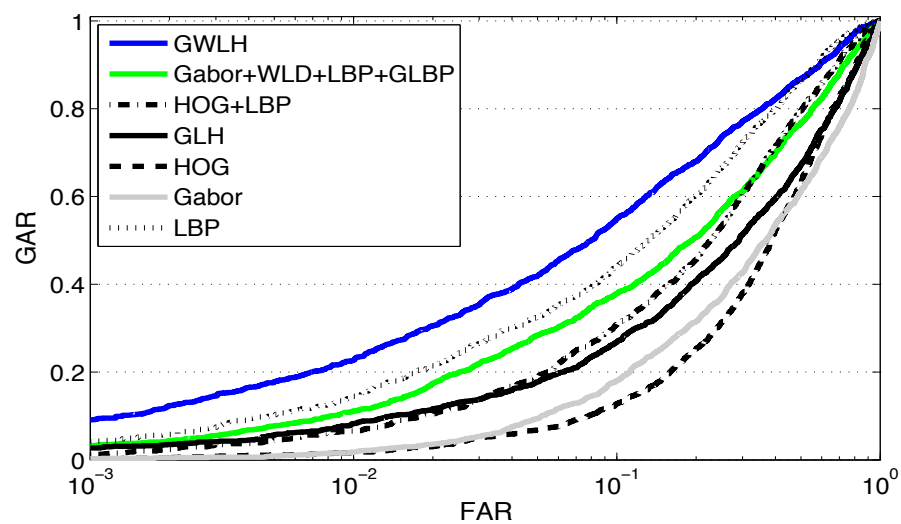
(b)
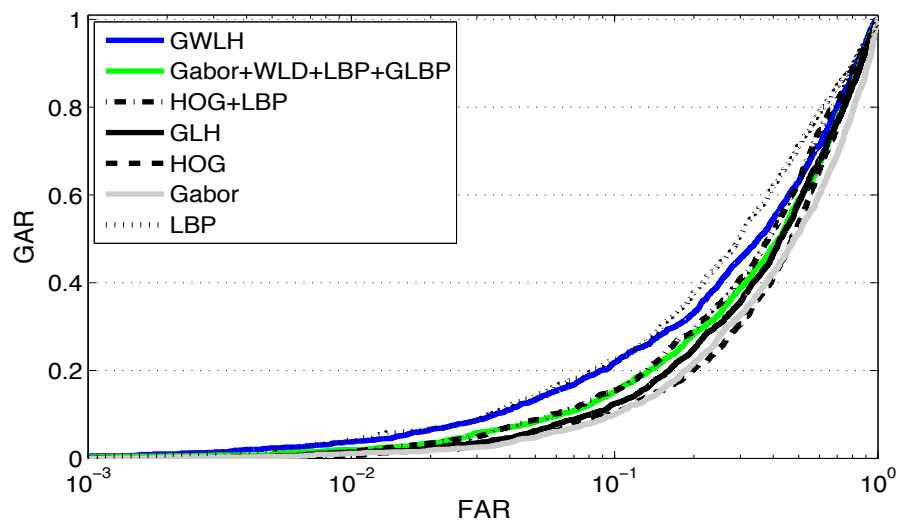
(c)

Figure 4.14: Cross-spectral matching of periocular regions using GWLH: (a) NIR 1.5 m, (b) NIR 50 m and (c) NIR 106 m.
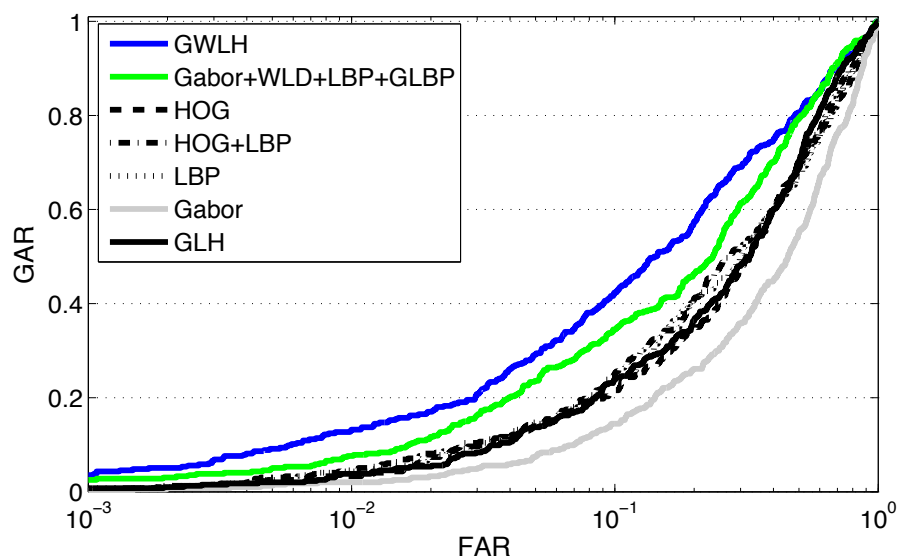
Figure 4.15: Cross-spectral matching of periocular regions using GWLH: MWIR.
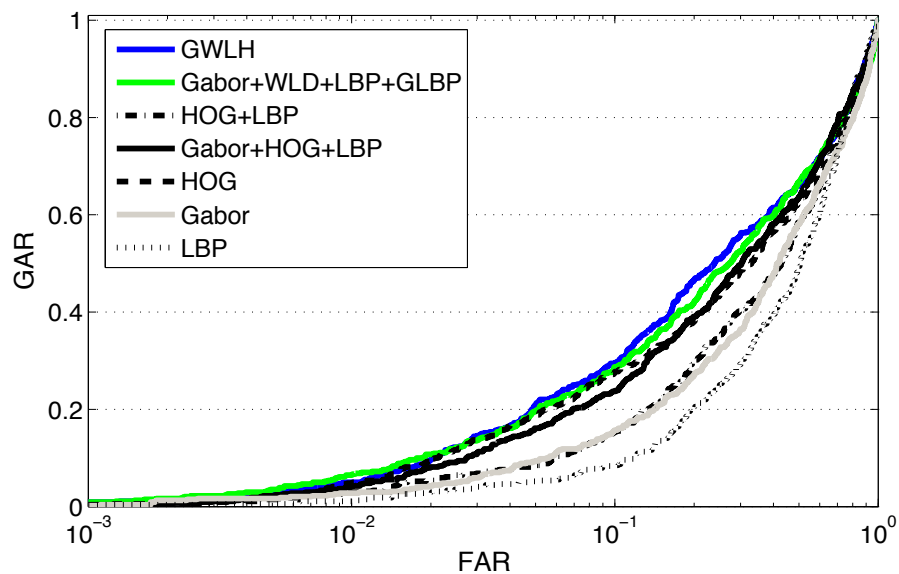
Figure 4.16: Cross-spectral matching of periocular regions using GWLH: LWIR.

Chapter 5

Cross-Spectral Partial Face Recognition

The problem of face recognition has been intensively studied for decades. Numerous techniques have been proposed. However, most of these research works has primarily focused on using full frontal or profile facial images, leaving partial face recognition as a relatively unexplored field. Thus, matching partial heterogeneous face images to a gallery of visible images is a new and challenging problem. This problem is motivated by a number of surveillance applications such as recognition of subjects at night or in the presence of severe environmental conditions. Standoff distances may range from a meter to hundreds of meters.

Our latest experiments have shown that face recognition algorithms recently developed in our research group can be adapted to perform cross-spectral matching of partial face images. The images are encoded with Local Binary Patterns and Weber Local Descriptor preceded by Gabor filters or with Composite Multi-lobe Descriptor and then matched by means of a symmetric Kullbuck-Leibler metric. Our analysis has shown that when division of the face into horizontal face regions, such as (1) forehead and eyebrows, (2) eyes and partial nose, and (3) lower part of the nose and mouth, the three regions display similar matching performance. When division of the face into characteristic patches, such as the eye, the nose and the mouth, the nose is shown to be most informative in terms of recognition performance.

5.1 Introduction

Although face recognition as a research area has been intensely studied in the past two decades, many challenges still exist, among which is the problem of partial face recognition. Imagery of partial faces is frequently generated in unconstrained face recognition scenarios, where occlusion of the face is likely to happen. The topic has not been well studied so far and remains an open problem [85, 122–125]. Furthermore, as new practical applications evolve, new challenges offered by the applications arise and hence a need for development of new encoding and matching algorithms to mitigate them. Cross-spectral face recognition is one of these recent applications. Surveillance at night or in harsh environments has appealed to new imaging modalities (such as NIR, SWIR, MWIR and LWIR) and evolved into new applications.

In this chapter we combine the problem of cross-spectral recognition with the problem of partial face recognition, that is, cross-spectral partial face recognition. The periocular region can be seen as a special case of partial face and periocular recognition has been recently explored (see for example, [107, 110–112, 125]). We have already spent efforts on studying the problem of periocular recognition in the cross-spectral setting in Chapter 4. In this chapter, we will address the problem of cross-spectral partial face recognition in a more general way by considering multiple facial regions. To be specific, we choose three facial regions in two different ways: (a) the top, the middle and the bottom regions in the case of division of the face into horizontal strips, and (b) the eye, the nose and the mouth in the case of division of the face into characteristic patches. We are especially interested in two questions. The first question is whether a facial part is sufficient for a face recognition task when facial occlusion is present. The second question is which facial part is the most informative in terms of recognition performance for conducting such a face recognition task.

The literature contains a few references on partial face matching of NIR face images and visible images [85]. However, results of heterogeneous matching of a partial face or periocular region in SWIR and MWIR spectra have not been previously reported, and this motivates the research described in our work. If addressed, heterogeneous partial face matching and heterogeneous matching of periocular regions will provide a baseline for development of new

improved algorithms for heterogeneous face recognition at night or in challenging environments. In addition to heterogeneous partial face matching we explore the effect of varying standoff distances on the recognition performance of the selected cross-spectral matching algorithm.

In this work, we use a method based on Gabor Generalized LBP combined with Gabor Weber descriptors. The algorithm was originally developed in our lab for cross-spectral face recognition [38, 116]. We use this algorithm to conduct two different types experiments corresponding to the two ways of division of the face, in order to find out the aforementioned questions raised by us.

The remainder of this chapter is organized as follows. Section 5.1 provides an introduction to the new problem of cross-spectral partial face recognition and a review of related topics and works. Section 5.2 studies the problem of cross-spectral partial face recognition using the first division scheme (horizontal strips). It gives the definition of partial face and describes the datasets used in the experiments. The experimental results for partial face recognition in two ways of covering face are presented. Section 5.3 explores the problem of cross-spectral partial face recognition using the other division scheme (characteristic patches). A summary of this chapter is provided in Section 5.4.

5.2 Cross-Spectral Partial Face Recognition with Horizontal Strips

5.2.1 Partial Face

In this section we study the problem of partial face recognition using the first face division scheme – horizontal strips. The face is divided into facial parts in the shape of rectangular strips vertically from the top to the bottom. We consider three facial parts: (1) the top which involves eyes and nasal bridge, (2) the middle which involves cheeks and nasal tip, and (3) the bottom which involves mouth and a part of the chin. An example of a face partitioned into the three parts is shown in Figure 5.1. The three regions shown in Fig. 5.1 have the following dimensions: 160×56 for the eyes and nose bridge, 160×46 for the nose

and the area of cheeks, and 160×52 for the mouth and a part of the chin in the case of larger original images, while 112×40, 112×34 and 112×38, respectively, in the case of smaller original images.

Figure 5.1: The first scheme of division of the face into three horizontal strips.

5.2.2   Informative Region

Since there are three different parts after the division of a face, we are interested in finding out which part contributes more to the overall recognition performance, as well as how much recognition performance is gained when including a new facial part. Therefore, we perform two experiments for each cross-spectral matching (i.e., NIR vs visible, SWIR vs visible and MWIR vs visible). The first experiment assumes covering two out of the three areas and keeps the third area in cross-spectral matching. The second experiment assumes a sequential face covering, in which we decrease the area of the face involved in the cross-spectral matching from all 3 regions exposed to a single region exposed. By conducting these two experiments we intend to identify the most informative part of the face for the purpose of cross-spectral matching. Clearly, this analysis is bound to the encoding algorithm that we use to produce results. Thus, all conclusions are relative to this algorithm.

5.2.3   Recognition System

The recognition framework follows the one we have used in the previous chapters: prepro-cessing, feature extraction and matching (see Section 2.2 in Chapter 2). For preprocessing and matching we use image alignment, cropping and a simple intensity normalization for prepeocessing while sum of two I-divergence distances as the metric to compare the feature

(a)

(b)
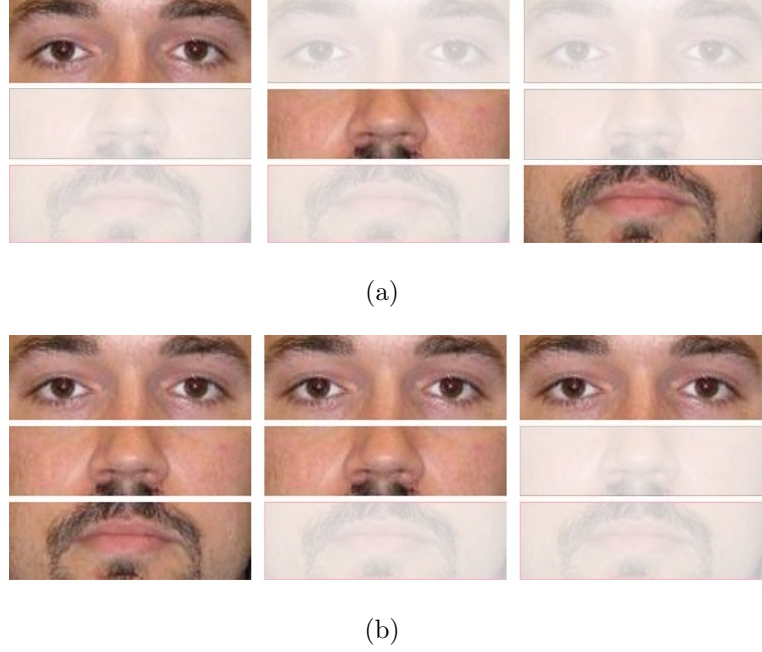
Figure 5.2: Two ways of covering the face: (a) Covering Two out of Three, (b) Sequential Covering.

vectors for matching. For feature extraction, we use a recent algorithm developed in our lab: Gabor+WLD+LBP+GLBP. Details of this algorithm can be found in Chapter 2. Here we only give a brief description.

The composite operator Gabor+WLD+LBP+GLBP adopts a Gabor filter-based approach at the initial encoding stage, followed by an encoding scheme that involves three operators – WLD, LBP and GLBP to extract robust features across different spectral bands. These three operators are designed to encode both magnitude and phase of filtered images. Each normalized image in any spectrum is processed with a bank of Gabor filters at 2 different scales and 8 orientations generating 16 responses in total. A normalized and preprocessed face image is convolved with the Gabor filter at orientation θ and scale s resulting in the filtered image $Y(z, \theta, s)$. Both magnitude and phase responses are used to effectively encode heterogeneous data block by block. The magnitude response is encoded using two distinct operators: SWLD [71] and uniform LBP [65]. For encoding the phase response we adopt a uniform generalized LBP operator (GLBP). All operators consider the relationship among 12 neighbors at both radii $r = 1$ and $r = 2$.

5.2.4   Datasets

In our experiments for the first face division scheme, we use three datasets as the original datasets (before division of the face). The first two datasets, Pre-TINDERS (Tactical Imager for Night/Day Extended-Range Surveillance) and TINDERS are collected by the Advanced Technologies Group, West Virginia High Tech Consortium (WVHTC) Foundation. The third dataset is collected by Pinellas County Sheriff's Office (PCSO). A detailed description of the datasets and sample images (see Figure 4.3) are provided in Section 2.3 of Chapter 2. Sample images from the datasets using the first division scheme are shown in Figure 5.3.
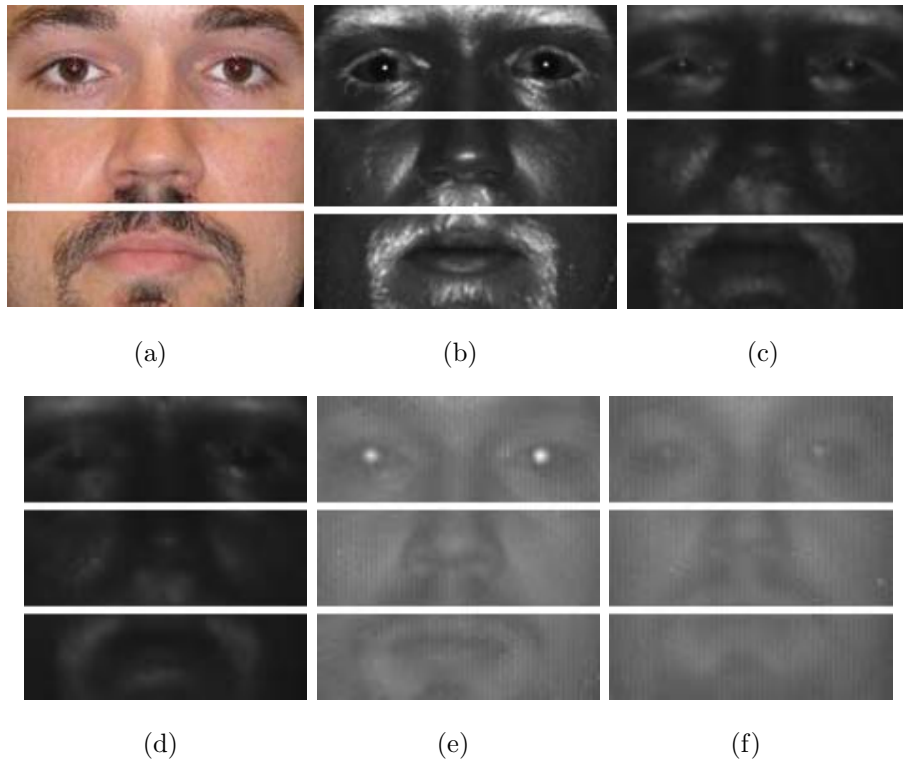
Figure 5.3: Partial face datasets: (a) visible light, (b) SWIR 1.5 m , (c) SWIR 50 m, (d) SWIR 106 m, (e) NIR 50 m and (f) NIR 106 m.

5.2.5   Covering Two out of Three Facial Regions

In this section, we discuss the first experiment: Face Recognition when Covering Two out of Three Facial Regions. Figure 5.4 (a) shows the performance of the matching algorithm described in Section 5.2.3 when query SWIR images are compared to a gallery of high quality

visible images. Three ROCs demonstrate performance for SWIR data collected at standoff distances of 1.5 m, 50 m and 106 m. It is interesting to note that each face sub-region contributes approximately the same amount of information to matching performance. The performance of these sub-regions is very similar for each given standoff distance. Note the fair performance of the cross-spectral matcher. Even for a large stand off distance such as 106 m the value of GAR is about 0.6 for the FAR set to 0.1.

Figure 5.4 (b) shows the performance of the cross-spectral matcher with MWIR probe images from the PCSO database. Again, three different regions show similar matching performance. However, the performance of the algorithm on MWIR data is significantly lower compared to the performance on SWIR data. The GAR is about 0.4 when the value of FAR is set to 0.1. Similar to the case with periocular regions, this performance drop can be attributed to the attempt of cross matching data of very different origin - thermal vs. reflective.

Figure 5.4 (c) show the performance of the cross-spectral matcher with NIR probe images from TINDERS dataset at long stand off distances of 50 m and 106 m, respectively. Again, the gallery is formed from high quality visible images collected at a short standoff distance. Similar to the case of SWIR and MWIR probes, the performance of each individual region (one of the three) is quite similar. Note that the ROC curves for these two cases are positioned between the pair of ROC curves - MWIR and SWIR at 106 m - and the other pair of ROC curves - SWIR at 1.5 m and SWIR at 50 m. This is consistent with the observations for the experiments with heterogeneous periocular regions.

5.2.6   Sequential Covering

The second experiment is performed with the purpose to determine how much of the face could be covered to satisfy a minimal performance requirement in practical applications. The results of cross-spectral matching for this case are shown in Figure 5.5.

The ROC curves displayed in Figure 5.5 (a) indicate that covering the area of mouth and chin followed by removal of the area of nose and cheeks uniformly degrades performance. Note that increasing standoff distance reduces quality of SWIR images (see Figure 5.3 (c)
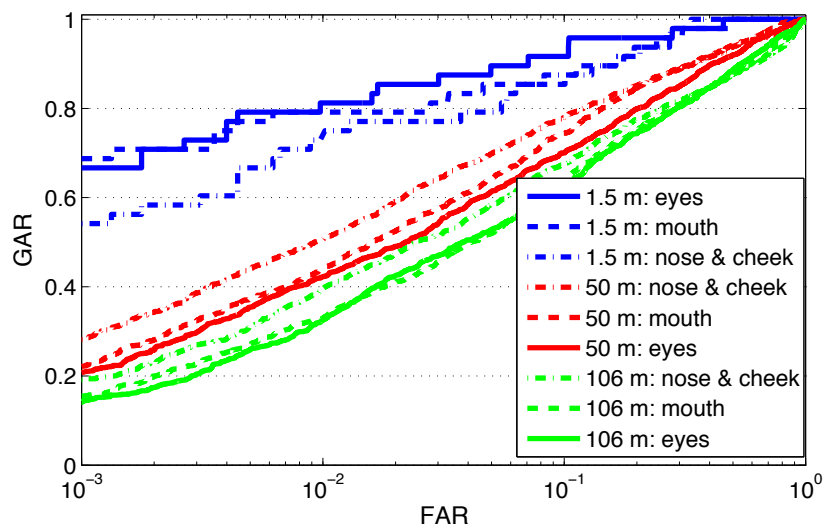
(a)
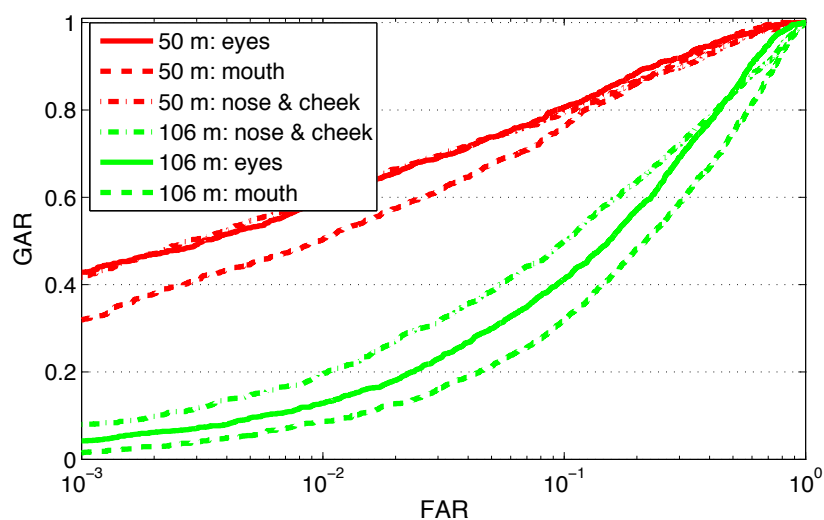
(b)
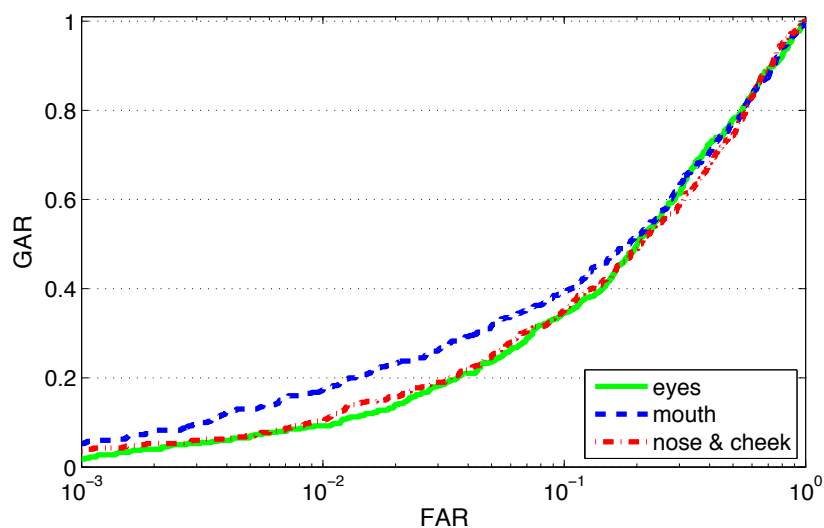
(c)

Figure 5.4: Covering two out of three facial parts: (a) SWIR, (b) NIR and (c) MWIR.
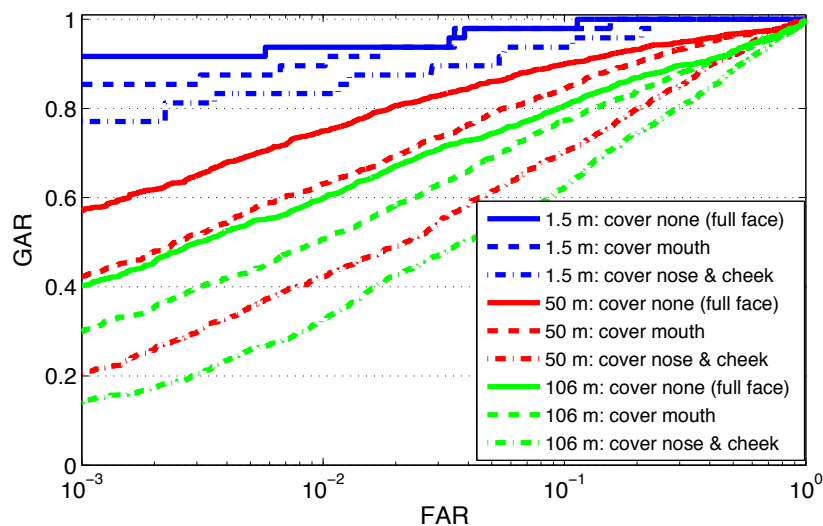
(a)
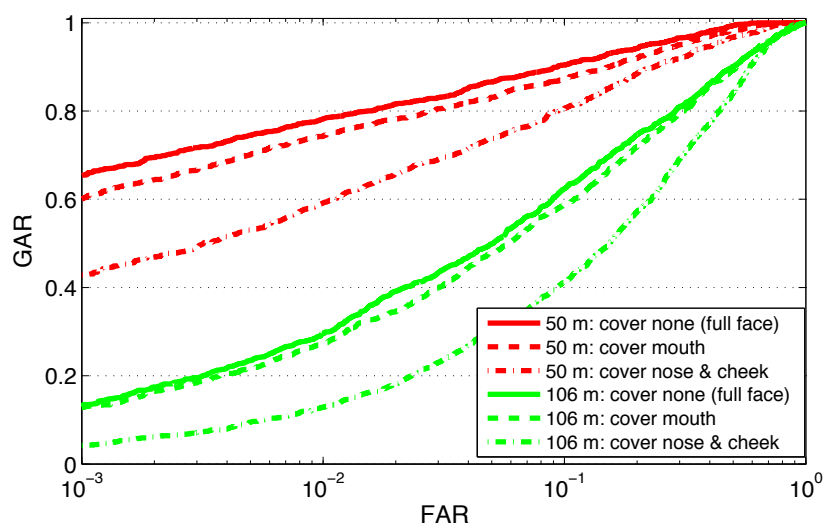
(b)
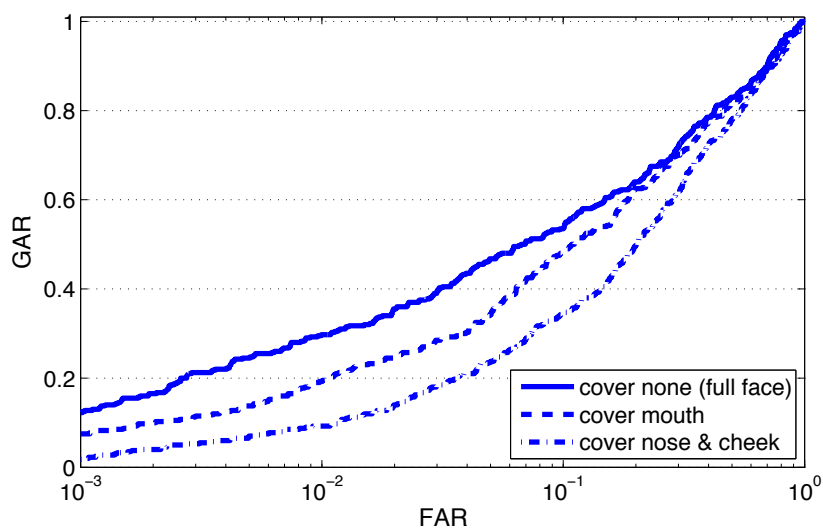
(c)

Figure 5.5: Incremental face covering: (a) SWIR, (b) NIR and (c) MWIR.

and (d)) and thus degrades performance. The performance of MWIR probes (see Figure 5.5 (b)) is lower compared to the performance of SWIR probes.

NIR probes at 50 m standoff distance demonstrate relatively high performance compared to NIR probes at 106 m. standoff distance (shown in Figure 5.5 (c)). One potential explanation for this performance drop is the weakness of NIR source of illumination, which is insufficient for quality imaging at this distance (see Figure 5.3 (e) and (f)). Another interesting fact about the performance of NIR probes is that covering the area of mouth and chin does not degrade performance.

5.3   Cross-Spectral Partial Face Recognition with Characteristic Patches

This section studies the problem of cross-spectral matching of partial faces using the second scheme of face division – characteristic patches. The regions are selected such that different facial regions are of different characteristics, (i.e., different organs) rather than being "mixed up" as in the first scheme.

5.3.1   The Second Face Division Scheme

After cropping and normalization of the face image (see Section 2.2.1 in Chapter 2 for more details), three regions of the face are selected and segmented: (a) the eye, (b) the nose, and (c) the mouth. An example of a face partitioned into the three regions is shown in Fig. 5.6. The three regions are chosen to have an area as close as possible for the sake of a fair comparison. The exact dimensions of each region are as follows : 40×38 for the eye, 24×62 for the nose, and 50×30 for the mouth. Partial face datasets are generated from the same datasets employed in previous cross-spectral face experiments which encompass the whole IR spectrum – SWIR, NIR, MWIR and LWIR.

For each cross-spectral comparison we perform an experiment where the three face regions are compared using the CMLD-I operator. By conducting these experiments we intend to identify the most informative part of the face for the purpose of cross-spectral matching.
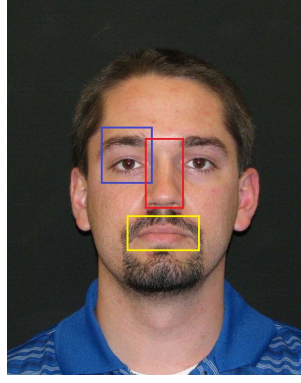
Figure 5.6: The second scheme of division of the face into three characteristic patches.

Clearly, this analysis is bound to the encoding algorithm that we use to produce results. Thus, all conclusions are relative to this algorithm.

5.3.2 Informativity Experiments

Matching SWIR Probes against Visible Gallery

Figure 5.7 (a) - (c) show the performance of the three different face regions encoded with CMLD-I when query SWIR images at standoff distance of 1.5 m, 50 m and 106 m are compared to a gallery of high quality visible images, respectively. As observed from the results the region of the eye outperforms the other two regions of nose and mouth, which suggests that it is more preferable to use the eye region than the other two to conduct a partial face recognition task. This observation is especially pronounced at shorter standoff distances. As the distance increases, the performance gap between the eye and other two regions narrows.

It is worth noting that even with a significant reduction of area compared to the full face, the eye alone is able achieve a GAR of 0.9362 at FAR = 0.1. This validates the usage of the eye region in partial face recognition. Metrics of matching performance such as EERs, d-primes, GARs with FAR set to 0.1 and 0.001, and rank-1 identification rates are provided in the first three rows of Table 5.1.
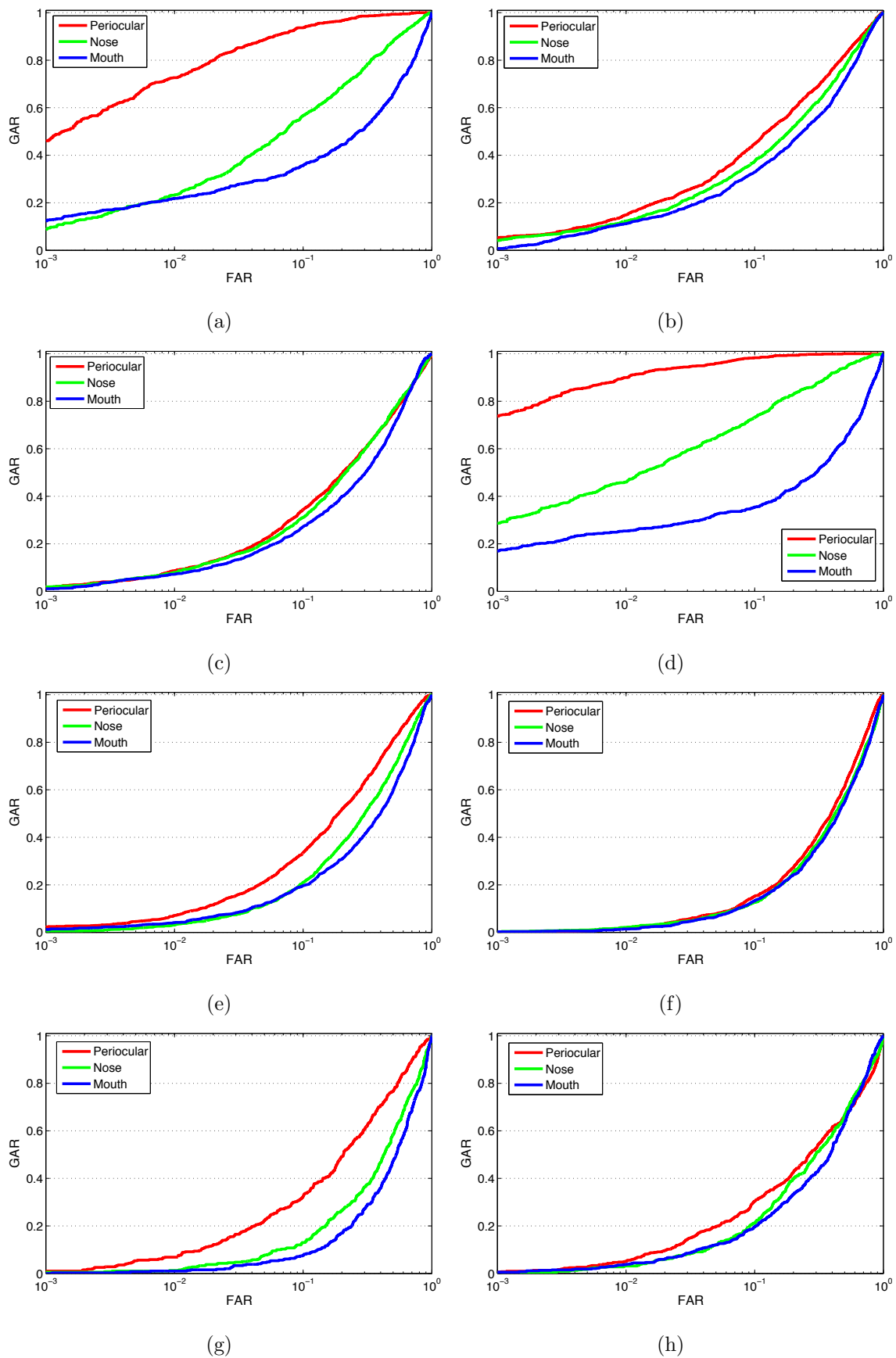
Figure 5.7: Matching IR against visible light facial regions: (a)-(c) SWIR 1.5 m, 50 m and 106 m; (d)-(f) NIR 1.5 m, 50 m and 106 m; (g) MWIR; (e) LWIR.

Table 5.1: EERs, d-prime and GAR values: matching IR to visible light facial regions.

Spectrum & Distance	Facial Region	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER (%)	d-prime	Rank-1 Id. Rate
SWIR 1.5 m	Eye	**93.62**	**46.09**	**7.83**	**2.61**	**65.63**
	Nose	56.64	8.85	26.41	1.29	32.29
	Mouth	35.68	12.37	40.63	0.65	36.46
SWIR 50 m	Eye	**44.89**	**5.32**	**31.08**	**1.01**	29.51
	Nose	37.61	4.13	34.44	0.84	**30.21**
	Mouth	32.98	0.77	38.55	0.65	22.22
SWIR 106 m	Eye	**34.44**	**1.74**	**35.4**	**0.7**	**17.36**
	Nose	31.11	1.81	35.48	0.71	17.01
	Mouth	27.29	1.11	40.37	0.58	12.15
NIR 1.5 m	Eye	**98.31**	**73.7**	**4.52**	**3.22**	**72.39**
	Nose	73.18	28.65	18.33	1.81	46.88
	Mouth	35.42	16.93	41.54	0.66	37.50
NIR 50 m	Eye	**33.54**	**2.24**	**33.52**	**0.87**	**23.26**
	Nose	21.15	0.56	40.14	0.51	6.25
	Mouth	19.68	1.19	45.19	0.32	8.33
NIR 106 m	Eye	**15.25**	**0.07**	**44.12**	**0.31**	**4.86**
	Nose	12.64	0.28	46.19	0.15	2.78
	Mouth	13.35	0.28	47.28	0.1	4.17
MWIR	Eye	**32.49**	**1.02**	**34.26**	**0.78**	**3.30**
	Nose	13.00	0.50	46.25	0.17	1.38
	Mouth	7.61	0.01	51.53	0.08	1.13
LWIR	Eye	**30.37**	**0.56**	**39.58**	**0.36**	**7.51**
	Nose	21.11	0.19	40.59	0.39	2.31
	Mouth	19.63	0.56	43.50	0.38	3.47

Matching NIR Probes against Visible Gallery

Figure 5.7 (d) - (f) display the performance of matching NIR probe images at standoff distances of 1.5 m, 50 m and 106 m, respectively, to visible light images. Again, the gallery is formed of high quality visible images collected at a short standoff distance. Similar to the case of cross-spectral matching between SWIR and visible light, the performance of the eye region is at the top. Again, as standoff distance increases, the advantage of using the eye over other regions diminishes. This is consistent with observations in the experiment with the SWIR band. A summary of the EERs, d-primes, GARs with FAR set to 0.1 and 0.001, and the rank-1 identification rates is included in Table 5.1.

Matching MWIR Probes against Visible Gallery

Figure 5.7 (g) demonstrates the performance of the cross-spectral matching between MWIR probe images and visible light images. The three different face regions have lower matching performance compared to the case of SWIR at a short distance due to degraded imagery. For the eye region, the GAR is now 0.3249 with FAR set to 0.1 compared with GAR of 0.9362 for the SWIR 1.5 m case. The performance drop can be attributed to the attempt of cross-matching data of very different origin – thermal vs. reflective. Nevertheless, the eye region is substantially better than the other two. Detailed matching results can be found in Table 5.1.

Matching LWIR Probes against Visible Gallery

The last experiment for the partial face recognition study is matching LWIR partial face probes against visible ones. Figure 5.7 (h) displays the matching results for this case. Due to the same reason as for the MWIR case, the three different face regions show substantially degraded performance compared to the case of SWIR at 1.5 m. For the eye region, the GAR is now 0.3037 at FAR = 0.1 dropping from the GAR of 0.9362 in the SWIR 1.5 m case. Once again the conclusion that the eye region is more preferable than the other two regions still holds. Detailed matching results can be found in the last three rows of Table 5.1.

5.4　Summary

This chapter deals with a new problem within the area of face recognition: cross-spectral partial face recognition. We presented the performance analysis of matching heterogeneous partial faces using two different schemes of face division. The encoding and matching algorithms were previously developed in our lab for the purpose of heterogeneous face recognition and adapted to work with partial face.

For the first scheme, we presented the results of partial face matching with probes being SWIR, NIR, and MWIR data and galleries composed of visible face images. Heterogeneous faces were partitioned into three non-overlapping horizontal strips from the top of the face to the bottom. We conducted two experiments: (1) covering of two out of three regions and (2) sequential covering of the face. The numerical results demonstrated that each region of face contributes almost an equal amount of information in terms of matching performance. This conclusion is valid for all three cross-spectral comparisons and for varying standoff distances, but under the condition that the heterogeneous images are encoded using Gabor Weber and Gabor Generalized LBP algorithm.

For the second scheme, the face was partitioned into three characteristic patches. We presented the results of partial face matching with probes being NIR, SWIR, MWIR, and LWIR data and galleries composed of visible face images. A newly developed operator CMLD is further utilized to conduct a study on cross-spectral partial face recognition where different facial regions are compared to find out the best one in terms of informativity. The experimental results show that for all IR bands and all standoff distances the eye region is the most advantageous among the three facial regions in terms of informativity for conducting a cross-spectral partial face recognition task.

Chapter 6

Image Quality Parity

This chapter addresses the problem of image quality disparity in cross-spectral face recognition. The problem is common when the heterogeneous images are acquired at different standoff distances. A technique called image quality parity is proposed for cross-spectral face recognition when there is a quality disparity between the probes and the gallery. It is achieved in two approaches — either to blur the images of higher quality (visible light images in our case) or to enhance the images of lower quality (infrared images). For blurring, we utilize a Gaussian smoothing kernel on the images with higher quality. For enhancement, a BM3D-based denoising step and a Laplacian-based sharpening step are combined. A quality measure tool called Adaptive Sharpness Measure is used for guiding the blurring and enhancement processes.

Heterogeneous face images are encoded using Gabor+WLD+LBP+GLBP as described in Chapter 2. Matching scores are generated by means of a Kullback-Leibler distance between two feature vectors. The two datasets employed in the experiments are Pre-TINDERS and TINDERS composed of heterogeneous face images acquired in NIR, SWIR and visible light spectra. To demonstrate the advantages of the proposed technique, we compare the performance of the cases in which SWIR and NIR spectra are matched against visible light at long distances of 50 and 106 meters, both before and after blurring and enhancement.

The outline of this chapter is as follows. Section 6.1 provides an introduction to the role of image quality and different factors contributing to quality degradation. A review of quality measure methods and utilizing quality for biometric systems is given subsequently.

Section 6.2 describes the details of the quality measure we adopt in this work – Adaptive Sharpness Measure (ASM). In Section 6.3 we introduce the techique of image quality parity for heterogeneous images by blurring. We explain the techique of image quality parity by enhancement in Section 6.4. Section 6.5 evaluates and presents the recognition performance with or without the usage of the blurring and enhancement approaches. The final section 6.6 summarizes the work and discoveries of the chapter.

6.1   Introduction

Image quality plays an important role in a face recognition system [126, 127]. The performance of the system is always affected by the quality of the input data despite having a good recognition algorithm. Common factors leading to a degraded image are poor lighting, defocus blur, camera noise, atmospheric conditions (such as fog, snow and rain), off-angle, occlusion, and so on.

There are two categories of quality measures: generic and biometric modality specific [128]. The former can be used for any biometric modality while the latter is designed to address issues related to a specific modality such as iris, fingerprints and faces [129–132]. Examples of generic quality measures that are used to quantify the degradation of a perceived image – typically compared to a reference good quality image – include image contrast, brightness, illumination and sharpness. Since factors like contrast, brightness and illumination are already dealt with to some degree at the normalization stage in our recognition system (see Chapter 2 for details), we hereby in this chapter focus on sharpness which can be a very common degradation factor for a surveillance scenario at a long standoff distance under severe atmospheric conditions. An illustration of SWIR images degrading with increasing standoff distance due to atmospheric noise is shown in Figure 6.1, where standoff distance varies from 1.5 m to 50 m and 106 m.

A lot of work has been done on image quality assessment or guidance for the biometric recognition task. Grother and Tabassi in their paper [133], relate the NIST Fingerprint Image Quality (NFIQ) to the matching performance of a fingerprint matcher. A Normalized Matching Score (NMS) measure is introduced and they show that quality of the finger-
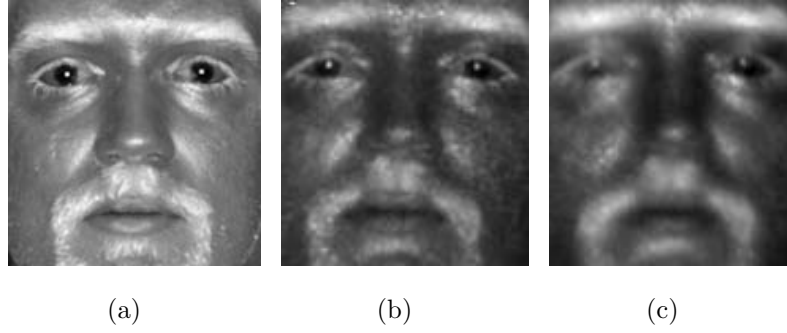
Figure 6.1: Quality degradation with increasing standoff distance due to atmospheric noise: (a) SWIR 1.5 m, (b) SWIR 50 m and (c) SWIR 106 m.

print samples is related to NMS (especially genuine scores) in the fingerprint recognition problem. The papers of Nandakumar et al. [134, 135] estimate the joint densities of quality and matching scores (genuine and impostor distributions). Then a likelihood ratio test between the estimated genuine and impostor distribution is adopted to evaluate the verification performance. In the work of Nandakumar et al. [134] the method is demonstrated on fingerprint and iris biometrics; for each modality a quality-based density is evaluated; hence a multi-modal distribution is obtained as a product of the individual density modalities. An improvement is obtained with respect to the case when the modalities are combined without the quality measures. In the work of Nandakumar et al. [135], the joint densities of single biometrics are described by a Gaussian Mixture Model (GMM). The model parameters are estimated with an Expectation Maximization (EM) algorithm. The works of Kryszczuk and Drygajlo [136, 137], involve biometric sample quality at the matching stage by concatenating matching scores due to the original matcher and quality measures. These quality-based matchers are called Q-stack classifiers; the method is demonstrated on face and fingerprint biometrics by adopting Support Vector Machines (SVM), Bayes classifiers and Linear Discriminant-based classifiers. In spite of the fundamental theory presented in these works, the obtained improvement of performance on individual face and fingerprint modalities is marginal. The work of Zuo and Schmid [138, 139] studied the assessment and fusion of quality factors in iris images and videos. They further used nonlinear mappings on iris quality measures and verification scores to predict and boost the performance [140].

As a summary, all these works either use the image quality measure (or score) to either

accept/discard a biometric sample or to use it as a weak feature for the recognition task. In this work, we utilize quality measure from a different perspective – to use it as an indicator for image blurring or sharpening. After this image processing stage, a degraded probe image and the reference image of high quality (and vice versa) will be brought to a similar level of quality, which yields higher recognition performance than the case without this stage.

6.2   Quality Measure

As mentioned earlier, in this chapter we focus on sharpness as the factor contributing to image quality. The question then goes to what sharpness measure to choose. As a cost function in real-time applications, robustness to noise and computational complexity are two primary concerns for choosing a sharpness measure. Gradient-based sharpness measures [141], especially the famous Tenengrad measure [142, 143], are known for their effectiveness and computational efficiency. Moreover, their pixel-based computations facilitate the differentiation between edge and noise pixels. Indeed, the differentiation reduces to simply assigning different weights to these pixels instead of tedious edge detection.

In this work, we utilize a gradient-based sharpness measure called Adaptive Sharpness Measure (ASM) by Yao et al. [104]. The main idea behind is to apply weight allocation schemes for adaptive unsharp masking (AUM) to sharpness measures based on the argument that actual visual perception is more sensitive to the transitions in the vicinity of edges, whose responses should be enhanced by allocating larger weights. The weights for the horizontal and the vertical directions, $L_x(x, y)$ and $L_y(x, y)$ respectively, are given by

$$\begin{cases} L_x(x, y) \ = [I(x+1, y) - I(x-1, y)]^2, \\ L_y(x, y) \ = [I(x, y+1) - I(x, y-1)]^2, \end{cases} \tag{6.1}$$

where $I(x, y)$ denotes the image intensity. The Adaptive Sharpness Measure then becomes

$$S = \sum_M \sum_N (L_x I_x^2 + L_y I_y^2), \tag{6.2}$$

where I_x and I_y are the horizontal and vertical components obtained using the Sobel operator, respectively. M and N represent the total number of image rows and columns, respectively.

As an illustration, the ASM values of the sample images in Figure 6.1 are calculated and listed in Table 6.1.

Measure	SWIR 1.5 m	SWIR 50 m	SWIR 106 m
Adaptive Sharpness Measures	0.8592	0.5323	0.3998

Table 6.1: Adaptive Sharpness Measure values for the sample images in Figure 6.1

6.3   Quality Parity by Blurring

After the choice of Adaptive Sharpness Measure as the quality measure tool, we then use it as an indicator for image blurring or sharpening. For example, during blurring we use it as an indicator of how much blurring should be applied to an image to lower its quality to the level of its heterogeneous counterpart. To degrade the quality of an image by blurring, we propose to use a simple smoothing function – a Gaussian kernel – to fulfill this goal. Its mathematical description is given by:

$$I_s(x,y) = I(x,y) * \frac{1}{\sigma\sqrt{2\pi}} e^{-(x^2+y^2)/2\sigma^2}, \tag{6.3}$$

where $I(x,y)$ is an input image and $I_s(x,y)$ is the output result after blurring, σ is the standard deviation and the symbol $*$ denotes the convolution operation.

An illustration of the blurring function is shown in Figure 6.3. The effect of blurring using Gaussian-based blurring function is shown in Figure 6.3, where the original visible light image and the blurred outputs are compared. The sharpness measure values are also calculated and listed in Table 6.2.

Measure	Original visible light image	Blurring with $\sigma = 0.1$	Blurring with $\sigma = 0.2$
Adaptive Sharpness Measures	0.6476	0.5871	0.5861

Table 6.2: Adaptive Sharpness Measure values for the sample images in Figure 6.3

.
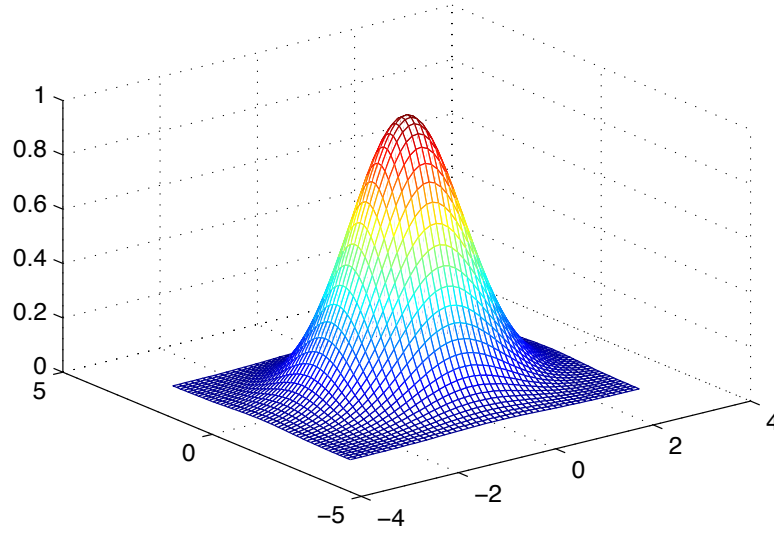
Figure 6.2: The Gaussian smoothing kernel.

(a)   (b)   (c)
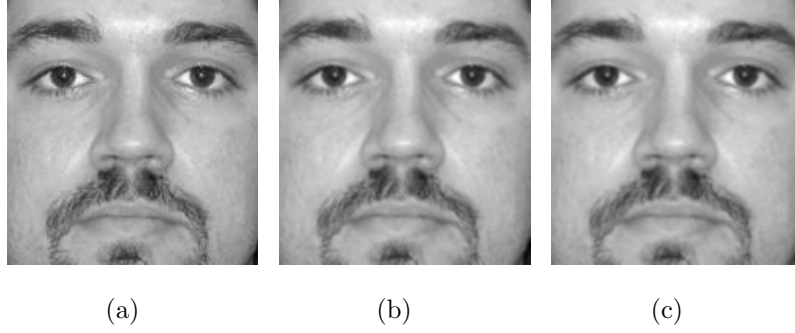
Figure 6.3: The effect of blurring: (a) original visible light image (grayscale converted), (b) blurring with $\sigma = 0.1$ and (c) blurring with $\sigma = 0.2$.

6.4 Quality Parity by Enhancement

The second approach proposed in this chapter for quality parity of heterogeneous face images is image enhancement by combination of denoising and sharpening. The combination of the two enhancement stages has the advantage of retaining the useful facial details while suppressing the noise. For denoising we utilize a technique called BM3D which is based on sparse representation in a 3D transform-domain [144]. For sharpening we use a method based on the Laplacian operator.

6.4.1 Denoising

The first part of BM3D is grouping where 2-dimensional fragments of a given image (i.e., image blocks) are collected according to similarity and then a 3D array is constructed by stacking the similar image neighborhoods together. A block-matching method which has been extensively used for motion estimation in video compression is used to realize the grouping task. The importance of grouping is to enable the usage of a higher-dimensional filtering of each group, that is, the second part of the BM3D technique: collaborative filtering. The collaborative filtering of the group of image blocks produces estimates in a way such that each group of blocks collaborates for the filtering of all others, and vice versa. Given a group of n image blocks, a total of n estimates will be produced: one for each of the grouped fragments. An effective collaborative filtering is realized as shrinkage in the transform domain consisting of the following: (a) apply a 3-dimensional linear transform to the group, (b) shrink (e.g. by soft- and hard-thresholding or Wiener filtering) the transform coefficients to attenuate the noise, and (c) invert the linear transform to produce estimates of all grouped image blocks. For better denoising performance, an improved grouping and collaborative filtering step with Wiener filtering is further added after the first step of basic grouping and collaborative filtering with hard-thresholding. A flowchart illustrating the two-step BM3D algorithm is presented in Figure 6.4. Denoising of a face image using the BM3D algorithm is displayed in Figure 6.5 (b) as an example.

Table 6.3: Adaptive Sharpness Measure values before and after denoising or sharpening.

Quality Measure	Original Image	After Denoising	After Further Sharpening
Adaptive Sharpness Measure	0.5850	0.3859	0.6479

6.4.2 Sharpening

After denoising, a sharpening stage is performed to recover the useful facial details partially attenuated during the denoising stage. In this chapter we use a Laplacian-based sharpening technique. It starts with finding the second order derivatives with the Laplacian
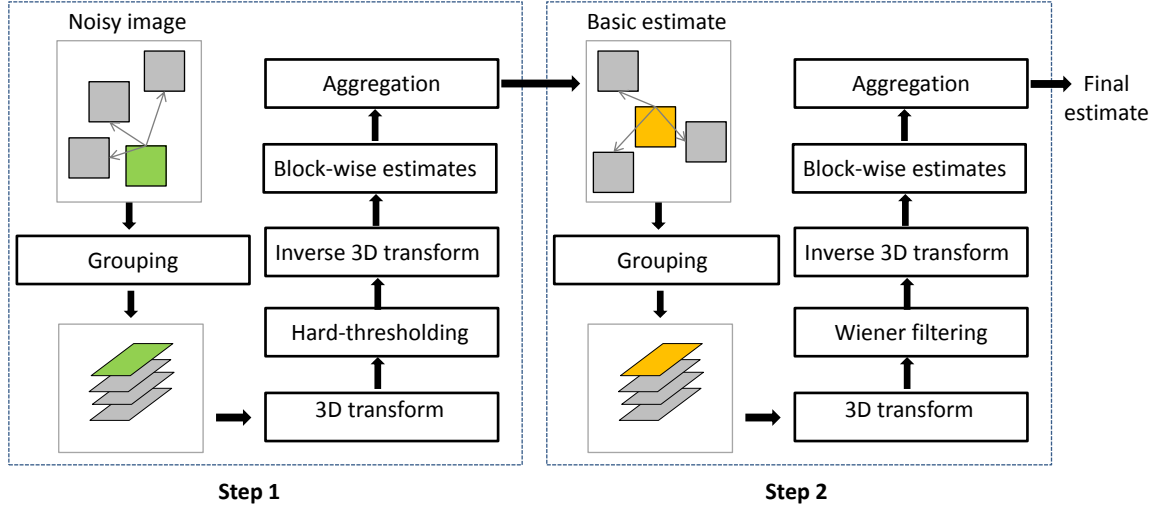
Figure 6.4: Flowchart of the BM3D denosing algorithm.

(a)   (b)   (c)
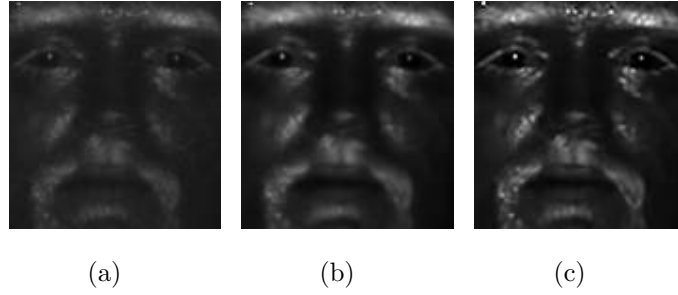
Figure 6.5: Denoising and sharpening: (a) original SWIR 50 m face image, (b) denoised output, and (c) further sharpened output.

operator:

$$\nabla^2 I = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}, \tag{6.4}$$

with the components in the x- and y-directions as:

$$\begin{cases} \frac{\partial^2 I}{\partial x^2} = I(x+1,y) + I(x-1,y) - 2I(x,y), \\[2mm] \frac{\partial^2 I}{\partial y^2} = I(x,y+1) + I(x,y-1) - 2I(x,y), \end{cases} \tag{6.5}$$

where $I(x,y)$ is an input image (the output of denoising). $\frac{\partial^2 I}{\partial x^2}$ and $\frac{\partial^2 I}{\partial y^2}$ are the directional derivatives along the x- and y-axis, respectively. To finally obtain the sharpened image, the output after processing with the Laplacian operator is added to the original input image:

$$I_{sh}(x,y) = I(x,y) + c\nabla^2 I, \tag{6.6}$$

where $I_{sh}(x,y)$ is the sharpened output image and c is a weight which adjusts the degree of sharpening needed. An example of the effect of sharpening is provided in Figure 6.5 (c). The ASM values for the original SWIR image, the image after denoising and the image after further sharpening are listed in Table 6.3.

6.5 Evaluation

In this section, we present the recognition performance for image quality parity via blurring or enhancement and compare them with the cases where blurring or enhancement are not applied. The recognition framework follows the same one that is described in Chapter 2. The heterogeneous images are encoded using a composite operator named Gabor+WLD+LBP+GLBP which combines Gabor filters, LBP, GLBP and WLD [38, 39]. In our experiments we use the TINDERS dataset (for detailed information, refer to Section 2.3 in Chapter 2). Results are presented for both 50 m and 106 m standoff distances.

6.5.1 Blurring of Visible Images

Matching SWIR Probes against Visible Gallery

The first experiment involves matching low quality SWIR face images acquired at 50 m and 106 m to visible face images of high quality acquired at a short standoff distance of 1.5 m. Image blurring is applied to the visible face images. The value of the blurring parameter, σ, is set to be 0.1 when matching visible images to SWIR images at 50 m and is set to be $\sigma = 0.2$ when matching visible to SWIR images at 106 m. The heterogeneous images are encoded using the composite operator Gabor+WLD+LBP+GLBP. The results of matching with and without blurring are shown in Figure 6.6. Values of EER, d-prime and GAR at FAR set to 0.1 and 0.001 are summarized in Table 6.4.

SWIR images at a long standoff distance experience some loss of quality due to air turbulence, insufficient illumination, and optical effects during data acquisition, which reflects on the matching performance [105, 121]. As the distance increases the impact of image quality becomes more pronounced. The results clearly show an improvement of performance after

Figure 6.6: ROC curves: matching SWIR probes to visible gallery with blurring.

Table 6.4: GAR, EER and d-prime values: matching SWIR probes to visible gallery with blurring.

Spectrum & Distance	Method	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER(%)	d-prime
SWIR 50 m	Original	91.88	62.11	8.90	2.57
	Blurring	92.93	67.09	7.92	2.74
SWIR 106 m	Original	82.50	44.79	14.17	2.00
	Blurring	86.74	51.67	11.75	2.27

using the blurring approach for both standoff distances of 50 m and 106 m.

Matching NIR Probes against Visible Gallery

In the second experiment, NIR face images (the probes) are matched to short standoff visible face images (the gallery). Again, the results of matching are parameterized by the standoff distances of 50 m and 106 m, which are shown in Figure 6.7. The values of σ are set to be 0.1 and 0.2 for matching visible images to NIR images at 50 m and 106 m, respectively. The encoding algorithm is also chosen to be Gabor+WLD+LBP+GLBP. Similar to the case of matching SWIR images to visible images, camera and atmospheric effects have significant impact on the recognition performance, especially when the distance is set to 106 m. Once

again, the blurring technique is proven to be beneficial for matching heterogeneous images with different quality. EER, d-prime and GAR at FAR set to 0.1 and 0.001 are summarized in Table 6.5 for both 50 m and 106 m distances, before and after blurring.
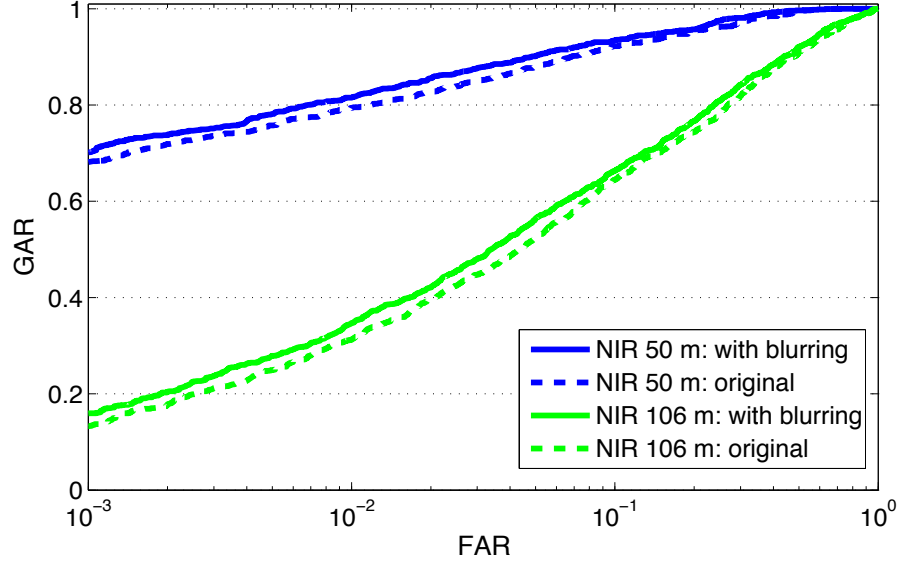
Figure 6.7: ROC curves: matching NIR probes to visible gallery with blurring.

Table 6.5: GAR, EER and d-prime values: matching NIR probes to visible gallery with blurring.

Spectrum & Distance	Method	GAR (%) at FAR = 10^{-1}	GAR (%) at FAR = 10^{-3}	EER(%)	d-prime
NIR 50 m	Original	92.23	68.21	8.71	2.66
	Blurring	93.42	70.24	7.63	2.77
NIR 106 m	Original	64.48	13.28	23.24	1.49
	Blurring	66.38	15.96	21.73	1.57

6.5.2 Enhancement of Infrared Images

Matching SWIR Probes against Visible Gallery

In the first experiment for the enhancement approach to image quality parity, SWIR face images (probes) at the standoff of 50 m and 106 m are denoised, sharpened and then matched to short range visible face images (gallery). The results of matching parameterized

by the stand-off distance are shown in Figure 6.8. EER, d-prime and GAR at FAR set to 0.1 and 0.001 are summarized in Table 6.6. By comparing the matching performance before and after the application of image enhancement, we clearly see the benefit of using this approach – a substantial performance improvement is observed.
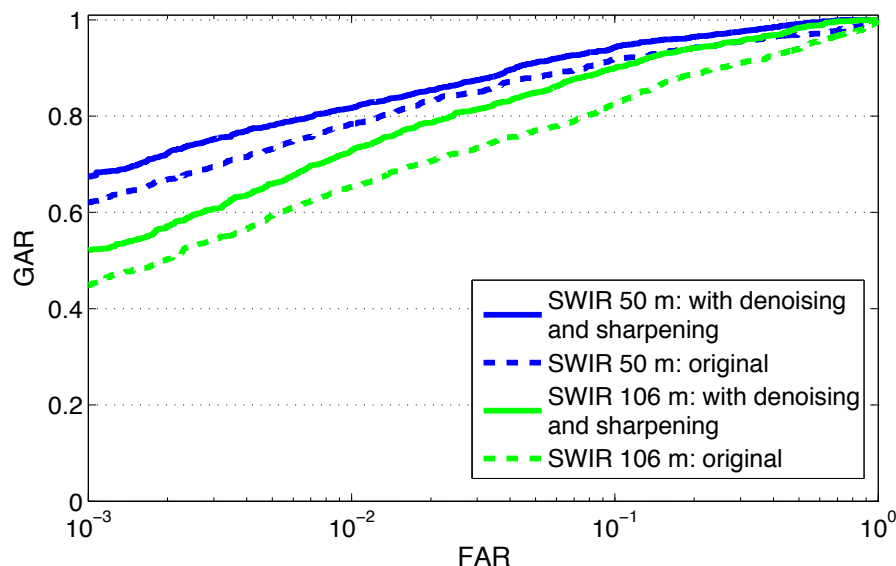
Figure 6.8: ROC curves: matching SWIR probes to visible gallery with enhancement.

Table 6.6: GAR, EER and d-prime values: matching SWIR probes to visible gallery with enhancement by denoising and sharpening.

Spectrum & Distance	Method	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER(%)	d-prime
SWIR 50 m	Original	91.88	62.11	8.90	2.57
	Enhancement	94.33	67.44	7.29	2.83
SWIR 106 m	Original	82.50	44.79	14.17	2.00
	Enhancement	90.00	52.15	10.00	2.55

Matching NIR Probes against Visible Gallery

In the next experiment using the enhancement approach, NIR face images at long standoff distances are matched against short standoff visible face images. The matching results parameterized by the standoff distances of 50 m and 106 m are displayed in Figure 6.9. EER,

d-prime and GAR at FAR set to 0.1 and 0.001 are summarized in Table 6.7 for both 50 m and 106 distances, before and after enhancement. Once again, the enhancement approach is proven to be beneficial for matching heterogeneous images with different quality.
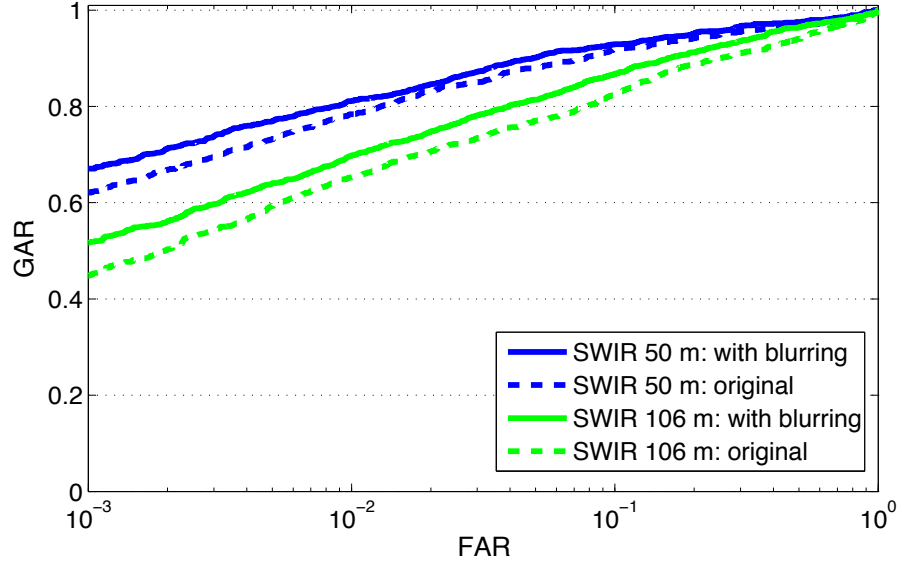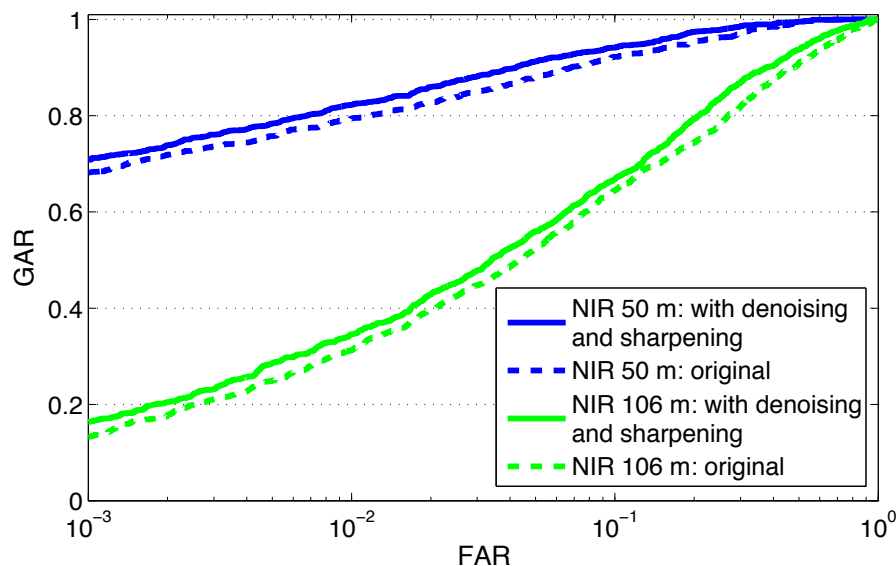

Figure 6.9: ROC curves: matching NIR probes to visible gallery with enhancement.

Table 6.7: GAR, EER and d-prime values: matching NIR probes to visible gallery with enhancement by denoising and sharpening.

Spectrum & Distance	Method	GAR (%) at FAR $= 10^{-1}$	GAR (%) at FAR $= 10^{-3}$	EER(%)	d-prime
NIR 50 m	Original	92.23	68.21	8.71	2.66
	Enhancement	94.12	70.87	7.12	2.78
NIR 106 m	Original	64.48	13.28	23.24	1.49
	Enhancement	66.81	16.38	20.39	1.64

The following paragraph summarizes the observations in the experiments:

- Infrared images acquired at a long standoff distance suffer from quality degradation due to atmospheric and camera effects which leads to a serious drop in the recognition performance.

- For both the SWIR and NIR spectra at both 50 meters and 106 meters, usage of image quality parity via blurring or enhancement leads to an improvement in the recognition performance, especially in the case of SWIR spectrum.

- As the quality disparity between the heterogeneous images increases – such as when the standoff distance increases from 50 m to 106 m, both methods are shown to be more beneficial (a larger performance improvement), especially for the SWIR band.

- In all cases of infrared bands and standoff distances, the enhancement approach appears to be slightly more advantageous than the blurring approach.

6.6   Summary

This chapter proposes a technique called image quality parity for cross-spectral face recognition when there is a quality disparity between the probes and the gallery. We consider the case where the gallery is composed of high quality face images collected at a short standoff distance and the probes are long range NIR or SWIR face images of low quality. We propose two approaches to the realization of image quality parity: (a) blurring the images of higher quality (visible light images in our case); and (b) enhancing of the images of lower quality (infrared images in our case). For blurring, a Gaussian smoothing kernel is utilized and applied to the visible light images. For enhancement, a BM3D-based denoising step and a Laplacian-based sharpening step are combined together. Adaptive Sharpness Measure is used as the tool for quantifying the amount of blurring and enhancement. An overview of related research work on image quality for biometric systems is also provided.

We present the evaluation of matching SWIR and NIR facial images to visible facial images. Long standoff distances at both 50 m and 106 m are considered. Both cases of recognition with and without blurring or enhancement are considered and compared. In all cases the proposed technique of quality parity (both approaches) is proven to be beneficial for the cross-spectral recognition task. The results are documented in figures and tables.

Chapter 7

Conclusion and Future Work

The final chapter summarizes the work and contributions made in the dissertation as well as envisioning possible future research problems that can be further explored. The work that has been done includes: a thorough comparative study on cross-spectral face recognition using some of the currently available or well-known operators; a detailed description of three operators newly proposed by us for the problem of cross-spectral face recognition; proposal and implementation of the technique of image quality parity for cross-spectral face recognition; a new topic of cross-spectral partial face recognition (cross-spectral periocular recognition as a special case). Potential topics for the future work include: to convert some special facial areas in infrared (especially SWIR) images due to the distinct imaging nature of infrared from visible light; to fuse infrared imagery acquired at different standoff distances to improve the recognition performance; to propose a general quality measure for heterogeneous images (i.e., independent of the electromagnetic wavelength); to design a new operator working at the nominal level of measurement; to study the problem of cross-spectral iris recognition.

7.1 Contribution and Conclusion

7.1.1 Cross-Spectral Face Recognition

Chapter 2 presented an overview of recent advances in the field of heterogeneous face recognition, emphasizing the topic of local operators developed for matching IR face probes to

a gallery composed of high quality visible face images. A brief description of each individual and composite operator (10 in total) was provided. The list of individual operators included LBP, GLBP, WLD, HOG and Gabor filters. Composite operators included Gabor+LBP, Gabor+GLBP, Gabor+WLD, GOM, and Gabor+LBP+GLBP+WLD.

The results of matching SWIR and NIR facial images to visible facial images were presented as ROC curves as well as in tables with GARs at two specific levels of FAR, EERs and d-prime values. Both short (1.5 m) and long (50 m and 106 m) standoff distances were considered. Conclusions from the experimental results are made as follows:

- The combination of Gabor filters followed by other local operators substantially outperformed the original LBP and the other individual operators. Among all the methods, Gabor+WLD+LBP+GLBP and GOM are the best in terms of recognition performance.

- As the standoff distance increased, the matching performance of all the methods dropped. This drop was attributed to a relatively low quality of imagery at long standoff distances (SWIR vs. visible and NIR vs. visible).

7.1.2 New Operators for Feature Extraction

Chapter 3 proposed three new operators – CMLD, GWLH and GMLM to extract and encode face features for the task of cross-spectral face recognition. After implementing the three operators, we present and analyze the experimental results of matching SWIR, NIR, MWIR or LWIR facial images to visible light images. Different standoff distances varying from short (1.5 m) to intermediate (50 m) and long (106 m) for SWIR and NIR are considered.

We made the following conclusions from the experimental results:

- The three operators substantially outperformed several popular simple operators such as LBP, HOG and WLD. They also outperformed the other composite operators such as Gabor followed by LBP (Gabor+LBP), and GOM.

- As the standoff distance increased, the matching performance of the heterogeneous face images dropped. This drop is attributed to a relatively low quality of SWIR and

NIR face images at long standoff distances.

7.1.3   Cross-Spectral Periocular Recognition

Chapter 4 introduced a new topic of study: periocular recognition in a cross-spectral context. It reviewed recent research work on periocular recognition as well as discussed the advantage of using periocular recognition as a new modality over face recognition. It then addressed the new problem by utilizing two newly proposed operators–CMLD and GWLH as the tool for feature extraction and compared them with other baseline algorithms. It further generated heterogeneous periocular datasets based on existing heterogeneous face datasets, and presented the results of matching SWIR, NIR and MWIR periocular probes to a gallery of visible periocular images. Both short (1.5 m) and long (50 m and 106 m) standoff distances were considered.

The experimental results led us to the following conclusions:

- The new operators CMLD and GWLH substantially outperformed baseline algorithms such as LBP, HOG, Gabor filters and three other compound operators when applied to heterogeneous periocular regions collected at a short standoff distance in the case of the three spectral bands.

- As the standoff distance increased (SWIR vs. visible and NIR vs. visible), the matching performance of the heterogeneous periocular images dropped for all the algorithms. This drop is attributed to a relatively low quality of heterogeneous images at long standoff distances. When matching MWIR periocular regions to visible regions (only a short standoff distance is considered), our algorithm displayed a relatively low performance. In this case it was due to the different nature of MWIR and visible imagery: MWIR imagery measures the heat of a body while visible imagery measures reflected light.

7.1.4   Cross-Spectral Partial Face Recognition

Chapter 5 deals with the new problem of cross-spectral partial face recognition. We presented the analysis of matching performance of partial matching of heterogeneous face.

The encoding and matching algorithms have been previously developed in our lab for the purpose of heterogeneous face recognition and adapted to work with partial face.

We presented the results of partial face matching with probes being SWIR, NIR, and MWIR data and a gallery composed of visible face images. The heterogeneous face was partitioned into three non-overlapping regions, either in a way of dividing them into horizontal strips or in a way of dividing them into characteristic patches. For the first way of division, we conducted two experiments: (1) covering of two out of three regions and (2) sequential covering of face. For the second way, we conducted an experiment to find out the most informative region in terms of recognition performance.

Conclusions from the experimental results are made as follows:

- The experimental results demonstrated that for the first way of division the three regions of the face contributed almost an equal amount of information in terms of matching performance. This conclusion was valid for all three cross-spectral comparisons and for varying standoff distances, but under the condition that the heterogeneous images were encoded using the Gabor WLD and Gabor Generalized LBP algorithms. When dividing the face using the second way, the periocular region was shown to be the most informative for all infrared bands and at all standoff distances.

- Increasing standoff distance substantially degraded performance of probes in SWIR and NIR spectra. The different nature of MWIR and visible face images resulted in seriously degraded performance of partial face matching when MWIR images were used as probes.

7.1.5   Image Quality Parity

Chapter 6 proposed to use a technique called image quality parity for cross-spectral face recognition when there is a quality disparity between the probes and the gallery. This problem is common when the heterogeneous images are acquired at different standoff distances. Image quality parity was achieved in two approaches: (a) utilizing a Gaussian blurring function on the images of higher quality (the visible light images in our case); (b) enhancing the images of lower quality (the infrared images in our case) by a combination of denoising and

sharpening. Adaptive Sharpness Measure was used as a guidance tool during the blurring process. An overview of related research work on image quality for biometric systems was also provided.

Conclusions from the experimental results were as follows:

- For both the SWIR and NIR spectra at both 50 meters and 106 meters, the usage of image quality parity by the two approaches led to a substantial improvement in the recognition performance.

- As the quality disparity between the heterogeneous images increases – i.e, as the stand-off distance increases from 50 m to 106 m, both methods were shown to be more beneficial (a more substantial performance improvement), especially for the SWIR band.

- In all cases of infrared bands and standoff distances, the enhancement approach appeared to be slightly more advantageous than the blurring approach.

7.2 Future Work

In Chapter 3 we addressed the problem of cross-spectral face recognition with a focus on designing new high-performance operators. However, the same problem can be approached by making the heterogeneous images look more "similar" to each other. For example, there are some special facial areas within the infrared images which look quite distinct from the areas within the visible light images due to different imaging nature. We can convert the special areas of the infrared images to be more similar to the areas of the visible light images by learning a relationship between these areas. By doing so, we can see if it helps improving the final recognition performance.

In Chapter 6, we addressed the problem of image quality disparity by blurring or enhancement. However, we can improve infrared images of lower quality by fusing the images collected at different stand-off distances at the image level. We can further test whether this imagery-level fusion idea is beneficial for improving the final recognition performance.

In Chapter 6, we use the tool ASM for quality measure (sharpness measure to be specific). This measure, like many other measures, is not designed for a cross-spectral context, that

is, it is not suitable for measuring the quality of infrared and visible light images at the same time. A general quality measure for heterogeneous images (i.e., independent of the electromagnetic wavelength) would be better if designed.

We designed a novel operator acting at the interval level of measurement and fused it with other operators at the ratio and ordinal levels in Chapter 3. We can further design a new operator working at the nominal level of measurement to have a complete set of levels of measurement-based operators.

In addition to applying our novel operators to the biometric modalities of face and partial face (including the periocular region) as discussed in Chapter 3, Chapter 4 and Chapter 5, we can also apply them to the modality of iris in a cross-spectral matching context.

Bibliography

[1] 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*, vol. 101, no. 1, pp. 1–15, 2006.

[2] A. K. Jain, P. Flynn, and A. A. Ross, eds., *Handbook of Biometrics.* Springer, 1 ed., 2008.

[3] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM computing surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.

[4] D. Maltoni, D. Maio, A. Jain, and S. Prabhakar, *Handbook of fingerprint recognition.* Springer, 2009.

[5] R. Y. F. Ng, Y. H. Tay, and K. M. Mok, "A review of iris recognition algorithms," in *2008 International Symposium on Information Technology*, vol. 2, pp. 1–7, IEEE, 2008.

[6] J. Martinez, H. Perez, E. Escamilla, and M. M. Suzuki, "Speaker recognition using mel frequency cepstral coefficients (mfcc) and vector quantization (vq) techniques," in *22nd Annual International Conference on Electronics, Communications and Computers, CONIELECOMP 2012*, 2012.

[7] S. Motiian, P. Pergami, K. Guffey, C. A. Mancinelli, and G. Doretto, "Automated extraction and validation of children's gait parameters with the kinect," *Biomedical engineering online*, vol. 14, no. 1, p. 1, 2015.

[8] F. Galton, "Personal identification and description," *Nature*, vol. 38, no. 5, pp. 173–177, 201–202, 1888.

[9] A. J. Goldstein, L. D. Harmon, and A. B. Lesk, "Identification of human faces," *Proc. IEEE*, vol. 59, no. 5, pp. 748–760, 1971.

[10] T. Kanade, "Picture processing system by computer complex and recognition of human faces," *Doctoral dissertation, Kyoto University*, vol. 3952, pp. 83–97, 1973.

[11] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Journal of the Optical Society of America A*, vol. 4, no. 3, pp. 519–524, 1987.

[12] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 586–591, 1991.

[13] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[14] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, pp. 775–779, July 1997.

[15] F. Samaria and S. Young, "Hmm-based architecture for face identification," *Image and vision computing*, vol. 12, no. 8, pp. 537–543, 1994.

[16] A. V. Nefian and M. H. Hayes III, "Hidden Markov models for face recognition," *Choice*, vol. 1, p. 6, 1998.

[17] R. Mehrotra, K. R. Namuduri, and N. Ranganathan, "Gabor filter-based edge detection," *Pattern Recognition*, vol. 25, no. 12, p. 1479–1494, 1992.

[18] T. Ojala, M. Pietikäinen, and D. Harwood, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions," in *Proc. of IAPR Int. Conf. on Pattern Recognition*, pp. 582–556, 1994.

[19] P. J. Phillips, H. Moon, S. Rizvi, P. J. Rauss, *et al.*, "The FERET evaluation methodology for face-recognition algorithms," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 10, pp. 1090–1104, 2000.

[20] J. Phillips, P. Grother, R. Micheals, D. Blackburn, E. Tabassi, and M. Bone, "FRVT 2002," *Evaluation Report, March*, 2003.

[21] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference on*, vol. 1, pp. 947–954, IEEE, 2005.

[22] S. Gong, S. J. McKenna, and A. Psarrou, *Dynamic vision: from images to face recognition.* Imperial College Press, 2000.

[23] 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*, vol. 97, no. 1, pp. 103–135, 2005.

[24] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 9, pp. 1063–1074, 2003.

[25] K. Chang, K. Bowyer, and P. Flynn, "Face recognition using 2D and 3D facial data," in *ACM Workshop on Multimodal User Authentication*, pp. 25–32, 2003.

[26] K. Ricanek Jr and T. Tesafaye, "Morph: A longitudinal image database of normal adult age-progression," in *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, pp. 341–345, IEEE, 2006.

[27] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Expression-invariant 3D face recognition," in *Audio-and Video-Based Biometric Person Authentication*, pp. 62–70, Springer, 2003.

[28] B. Moghaddam and M.-H. Yang, "Learning gender with support faces," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 5, pp. 707–711, 2002.

[29] F. J. Prokoski and R. B. Riedel, "Infrared identification of faces and body parts," in *Biometrics*, pp. 191–212, Springer, 1996.

[30] M. P. Hansen and D. S. Malchow, "Overview of SWIR detectors, cameras, and applications," in *Proc. of SPIE: Thermosense XXX*, pp. 69390I–69390I–11, 2008.

[31] Goodrich, "Surveillance Using SWIR Night Vision Cameras." `http://www.sensorsinc.com/facilitysecurity.html`. [Online; accessed on Feb. 01, 2015].

[32] D. Socolinsky, L. Wolff, J. Neuheisel, and C. Eveland, "Illumination invariant face recognition using thermal infrared imagery," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 527–534, 2001.

[33] X. Chen, P. Flynn, and K. Bowyer, "PCA-based face recognition in infrared imagery: Baseline and comparative studies," in *Proc. of IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 127 – 134, 2003.

[34] P. Buddharaju, I. T. Pavlidis, P. Tsiamyrtzis, and M. Bazakos, "Physiology-based face recognition in the thermal infrared spectrum," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 29, no. 4, pp. 613 – 626, 2007.

[35] T. Bourlai, N. Kalka, A. Ross, B. Cukic, and L. Hornak, "Cross-spectral face verification in the short wave infrared (SWIR) band," in *Proc. of Int. Conf. on Patt. Recog.*, pp. 1343 – 1347, 2010.

[36] J. Kirschner, "SWIR for Target Detection, Recognition, And Identification." `http://www.photonicsonline.com/doc.mvc/SWIR-For-Target-Detection-Recognition-And-0002`, 2011. [Online; accessed on Jan. 04, 2015].

[37] F. Nicolo and N. A. Schmid, "A method for robust multispectral face recognition," in *Proc. of the International Conference on Image Analysis and Recognition*, pp. 180–190, 2011.

[38] F. Nicolo and N. A. Schmid, "Long range cross-spectral face recognition: Matching SWIR against visible light images," *IEEE Trans. on Inf. Forensics and Security*, vol. 7, no. 6, pp. 1717–1726, 2012.

[39] Z. Cao and N. A. Schmid, "Matching heterogeneous periocular regions: Short and long standoff distances," in *Proc. of the IEEE International Conference on Image Processing*, pp. 4967–4971, 2014.

[40] Z. Cao and N. A. Schmid, "Recognition performance of cross-spectral periocular biometrics and partial face at short and long standoff distance," *Open Transactions on Information Processing*, vol. 1, no. 2, pp. 20–32, 2014.

[41] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1063–1074, Sept 2003.

[42] M. Williams, "Better face-recognition software," *Technology Review*, vol. 30, 2007.

[43] K. Bonsor and R. Johnson, "How facial recognition systems work," *HowStuffWorks. com. Np*, vol. 4, 2001.

[44] M. Crawford, "Facial recognition progress report," *SPIE Newsroom*, 2011.

[45] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705–741, 1995.

[46] K. J. Snell, A. Parent, M. Levesque, and P. Galarneau, "Active range-gated near-IR TV system for all-weather surveillance," in *Enabling Technologies for Law Enforcement and Security*, pp. 171–181, International Society for Optics and Photonics, 1997.

[47] M. Tamir, "Systems and methods for night time surveillance," Jan 2006. US Patent App. 11/325,147.

[48] S. Koyama, Y. Inaba, M. Kasano, and T. Murata, "A day and night vision mos imager with robust photonic-crystal-based RGB-and-IR," *Electron Devices, IEEE Transactions on*, vol. 55, no. 3, pp. 754–759, 2008.

[49] Equinox. `http://www.equinoxsensors.com`. [Online; accessed on July. 09, 2015].

[50] F. Prokoski, "History, current status, and future of infrared identification," in *Computer Vision Beyond the Visible Spectrum: Methods and Applications, 2000. Proceedings. IEEE Workshop on*, pp. 5–14, IEEE, 2000.

[51] M. A. Turk and A. P. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 13, no. 1, pp. 71–86, 1991.

[52] C. Jutten and J. Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, no. 1, p. 1–10, 1991.

[53] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[54] M. S. Bartlet and T. J. Sejnowski, "Independent components of face images: A representation for face recognition," in *Proc. 4th Annual J. Symp. Neural Computation*, 1997.

[55] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Transactions on Neural Networks*, vol. 13, no. 6, p. 1450–1464, 2002.

[56] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3, pp. 321–377, 1936.

[57] T. Melzera, M. Reitera, and H. Bisch, "Appearance models based on kernel canonical correlation analysis," *Pattern Recogniton*, vol. 36, no. 9, pp. 1961–1971, 2003.

[58] S. Marčelja, "Mathematical description of the responses of simple cortical cells," *Journal of the Optical Society of America*, vol. 70, no. 11, p. 1297–1300, 1980.

[59] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *Journal of the Optical Society of America A*, vol. 2, no. 7, p. 1160–1169, 1985.

[60] J. G. Daugman, "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 36, no. 7, pp. 1169–1179, 1988.

[61] T. S. Lee, "Image representation using 2D Gabor wavelets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, p. 959–971, 1996.

[62] A. K. Jain, N. K. Ratha, and S. Lakshmanan, "Object detection using Gabor filters," *Pattern Recognition*, vol. 30, no. 2, p. 295–309, 1997.

[63] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," in *Proc. of IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp. 200–205, 1998.

[64] L. Wang and D. C. He, "Texture classification using texture spectrum," *IEEE Trans. on Pattern Analysis and Mach. Int.*, vol. 23, no. 8, pp. 905 – 910, 1990.

[65] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Inf. Forensics and Security*, vol. 24, no. 7, pp. 971–987, 2002.

[66] T. Ahonen, A. Hadid, and M. Pietikainen, "Face recognition with local binary patterns," in *Proc. of European Conference on Comuputer Vision (ECCV)*, pp. 469–481, 2004.

[67] M. Heikkilä and M. Pietikäinen, "A texture-based method for modeling the background and detecting moving objects," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 28, no. 4, pp. 657–662, 2006.

[68] G. Zhao and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 915–928, 2007.

[69] L. Zhang, R. Chu, S. Xiang, S. Liao, and S. Z. Li, "Face detection based on multi-block LBP representation," in *Advances in Biometrics*, vol. 4642 of *Lecture Notes in Computer Science*, pp. 11–18, Springer Berlin Heidelberg, 2007.

[70] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 886 – 893, 2005.

[71] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikeinen, X. Chen, and W. Gao, "WLD: A robust local image descriptor.," *IEEE Trans. on Pattern Analysis and Mach. Int.*, vol. 32, no. 9, pp. 1705–1720, 2010.

[72] H. E. Ross and D. J. Murray, *E. H. Weber on the Tactile Senses, 2nd Edition*. Erlbaum (UK) Taylor and Francis, 1996.

[73] X. Chen, P. J. Flynn, and K. W. Bowyer, "IR and visible light face recognition," *Comput. Vis. Image Understand.*, vol. 99, no. 3, pp. 332–358, 2005.

[74] Z. Pan, G. Healey, M. Prasad, and B. Tromberg, "Face recognition in hyperspectral images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 12, pp. 1552–1560, 2003.

[75] S. G. Kong, J. Heo, F. Boughorbel, Y. Zheng, B. R. Abidi, A. Koschan, M. Yi, and M. A. Abidi, "Multiscale fusion of visible and thermal IR images for illumination-invariant face recognition," *International Journal of Computer Vision*, vol. 71, no. 2, pp. 215–233, 2007.

[76] "UTK-IRIS." http://imaging.utk.edu. [Online; accessed on July. 09, 2015].

[77] S. Z. Li, R. Chu, S. Liao, and L. Zhang, "Illumination invariant face recognition using near-infrared images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 29, no. 4, pp. 627–639, 2007.

[78] M. A. Akhloufi, A. Bendada, and J.-C. Batsale, "Multispectral face recognition using non linear dimensionality reduction," in *SPIE Defense, Security, and Sensing*, pp. 73410J–73410J, International Society for Optics and Photonics, 2009.

[79] M. Akhloufi and A. H. Bendada, "Multispectral infrared face recognition: a comparative study," in *Proc. of Quantitative InfraRed Thermography*, 2010.

[80] M. Akhloufi and A. H. Bendada, "A new fusion framework for multispectral IR face recognition in the texture space," in *Proc. of Quantitative InfraRed Thermography*, 2010.

[81] D. Lin and X. Tang, "Inter-modality face recognition," in *Computer Vision–ECCV 2006*, pp. 13–26, Springer, 2006.

[82] J. Yang, A. F. Frangi, J.-Y. Yang, D. Zhang, and Z. Jin, "KPCA plus LDA: a complete kernel Fisher discriminant framework for feature extraction and recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 2, pp. 230–244, 2005.

[83] S. Liao, D. Yi, Z. Lei, R. Qin, and S. Z. Li, "Heterogeneous face recognition from local structures of normalized appearance," in *Proc. of IAPR/IEEE Int. Conf. on Biometrics*, pp. 209–218, 2009.

[84] D. Yi, R. Liu, R. Chu, Z. Lei, and S. Z. Li, "Face matching between near infrared and visible light images," in *Advances in Biometrics*, pp. 523–530, Springer, 2007.

[85] D. Yi, S. Liao, Z. Lei, J. Sang, and S. Z. Li, "Partial face matching between near infrared and visual images in MBGC portal challenge," in *Proc. of IAPR/IEEE International Conference on Biometrics*, pp. 733–742, 2009.

[86] "Multiple biometric grand challenge (MBGC)." `http://www.nist.gov/itl/iad/ig/mbgc.cfm`. [Online; accessed on July. 09, 2015].

[87] P. J. Phillips, P. J. Flynn, J. R. Beveridge, W. T. Scruggs, A. J. O'toole, D. Bolme, K. W. Bowyer, B. A. Draper, G. H. Givens, Y. M. Lui, *et al.*, "Overview of the multiple biometrics grand challenge," in *Advances in Biometrics*, pp. 705–714, Springer, 2009.

[88] B. Klare and A. K. Jain, "Heterogeneous face recognition: Matching nir to visible light images," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, pp. 1513–1516, IEEE, 2010.

[89] Cognitec, "Facevacs Software Developer Kit Cognitec Systems." `http://www.cognitec-systems.de`. [Online; accessed on Jan. 04, 2015].

[90] H. Maeng, H.-C. Choi, U. P. S.-W. Lee, and A. K. Jain, "NFRAD: Near-infrared face recognition at a distance," in *Proc. of the International Joint Conference on Biometrics*, pp. 1–7, 2011.

[91] Z. Chai, Z. Sun, H. Mendez-Vazquez, R. He, and T. Tan, "Gabor ordinal measures for face recognition," *IEEE Trans. on Inf. Forensics and Security*, vol. 9, no. 1, pp. 14–26, 2014.

[92] WVHTCF, "Tactical Imager for Night/Day Extended-Range Surveillance." `http://www.wvhtf.org/programs/advancedtech/ONR%20Fact%20Sheet%20TINDERS.pdf`. [Online; accessed on Feb. 01, 2015].

[93] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. of IEEE Conference. on Computer Vision and Pattern Recognition*, pp. 511– 518, 2001.

[94] X. Tan and B. Triggs, "Fusing Gabor and LBP feature sets for kernel-based face recognition," in *Proceedings of the 3rd International Conference on Analysis and Modeling of Faces and Gestures*, AMFG'07, pp. 235–249, 2007.

[95] S. Xie, S. Shan, X. Chen, and J. Chen, "Fusing local patterns of Gabor magnitude and phase for face recognition," *Image Processing, IEEE Transactions on*, vol. 19, no. 5, pp. 1349–1361, 2010.

[96] Y. Guo and Z. Xu, "Local Gabor phase difference pattern for face recognition.," in *Proc. of Int. Conf. on Pattern Recognition*, pp. 1–4, 2008.

[97] F. Nicolo, *Homogeneous and Heterogeneous Face Recognition: Enhancing, Encoding and Matching for Practical Applications*. PhD thesis, West Virginia University, 2012.

[98] S. S. Stevens, "On the theory of scales of measurement," *Science*, vol. 103, no. 2684, pp. 677–680, 1946.

[99] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[100] R. B. Martin, K. M. Kafka, and B. E. Lemoff, "Active-SWIR signatures for long-range night/day human detection and identification," in *Proc. of the SPIE Symp. on DSS*, pp. 209–218, 2013.

[101] B. Klare and A. K. Jain, "Heterogeneous face recognition: Matching NIR to visible light images," in *Proc. of Int. Conf. on Pattern Recognition*, pp. 1513–1516, 2010.

[102] X. Wang, T. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," in *Computer Vision, 2009 IEEE 12th International Conference on*, pp. 32–39, Sept 2009.

[103] A. Sinha, S. Banerji, and C. Liu, "Gabor-based novel local, shape and color features for image classification," in *Int. Conf. on Image Processing, Computer Vision and Pattern Recognition*, pp. 299–306, 2012.

[104] Y. Yao, B. Abidi, and M. Abidi, "Digital imaging with extreme zoom: System design and image restoration," in *Proc. of the IEEE Int. Conf. on Computer Vision Systems*, 2006.

[105] Z. Cao and N. A. Schmid, "Composite multi-lobe descriptor for cross-spectral face recognition: Matching active IR to visible light images," in *Proc. of Automatic Target Recognition XXV Conference in the 2015 SPIE DSS Symposium*, pp. 94760T–94760T–13, 2015.

[106] Z. Cao and N. A. Schmid, "Novel composite operator for cross-spectral face recognition," *SPIE Newsroom*, April 2015.

[107] U. Park, A. Ross, and A. K. Jain, "Periocular biometrics in the visible spectrum: a feasibility study," in *Proc. of IEEE 3rd Int. Conf. on Biometrics: Theory, Applications and Systems*, pp. 1–6, 2009.

[108] D. L. Woodard, S. J. Pundlik, J. R. Lyle, and P. E. Miller, "Periocular region appearance cues for biometric identification," in *Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops*, pp. 162–169, 2010.

[109] S. Bharadwaj, H. S. Bhatt, M. Vatsa, and R. Singh, "Periocular biometrics: When iris recognition fails," in *Proc. of IEEE Int. Conf. on Biometrics: Theory, Applications and Systems*, pp. 27–29, 2010.

[110] K. P. Hollingsworth, S. S. Darnell, P. E. Miller, D. L. Woodard, K. W. Bowyer, and P. J. Flynn, "Human and machine performance on periocular biometrics under near-infrared light and visible light," *IEEE Trans. on Inf. Forensics and Security*, vol. 7, no. 2, pp. 588–601, 2012.

[111] U. Park, R. R. Jillela, A. Ross, and A. K. Jain, "Periocular biometrics in the visible spectrum," *IEEE Trans. on Inf. Forensics and Security*, vol. 6, no. 1, pp. 96–106, 2011.

[112] C. W. Tan and A. Kumar, "Human identification from at-a-distance images by simultaneously exploiting iris and periocular feature," in *Proc. of Int. Conf. on Pattern Recognition*, pp. 553–556, 2012.

[113] Z. Cao, N. A. Schmid, and T. Bourlai, "Local operators and measures for heterogeneous face recognition," in *Face Recognition Across the Electromagnetic Spectrum* (T. Bourlai, ed.), Springer, 2015.

[114] A. Sharma, S. Verma, M. Vatsa, and R. Singh, "On cross spectral periocular recognition," in *Proc. of the IEEE International Conference on Image Processing*, pp. 5007–5011, 2014.

[115] N. D. Kalka, T. Bourlai, B. Cukic, and L. Hornak, "Crossspectral face recognition in heterogeneous environments: A case study on matching visible to short-wave infrared imagery," in *Proc. of Int. Joint Conf. on Biometrics*, pp. 1–8, 2011.

[116] J. Zuo, F. Nicolo, N. Schmid, and S. Boothapati, "Encoding, matching and score normalization for cross spectral face recognition: Matching SWIR versus visible data," in *Proc. of IEEE 5th Int. Conf. on Biometrics: Theory, Applications and Systems*, pp. 203 – 208, 2012.

[117] B. F. Klare and A. K. Jain, "Heterogeneous face recognition using kernel prototype similarities," *IEEE Trans. on Pattern Analysis and Mach. Int.*, vol. 35, no. 6, pp. 1410–1422, 2013.

[118] J. Xu, M. Cha, J. L. Heyman, S. Venugopalan, R. Abiantun, and M. Savvides, "Robust local binary pattern feature sets for periocular biometric identification," in *Proc. of IEEE Int. Conf. on Biometrics: Theory, Applications and Systems*, pp. 1–8, 2010.

[119] S. McCallum, "Pinellas county sheriff's office dhs future opportunitie."

[120] S. Schuckers, P. J. Paulo Meyer Lopez, N. Sazonova, F. Hua, R. Lazarick, C. Miles, E. Talbassi, E. Sazonov, A. Ross, and L. Hornak, "Quality–Face / Iris Research Ensemble (Q-FIRE) Dataset Overview." `http://www.clarkson.edu/citer/pdf/53287.pdf`, 2010. [Online].

[121] Z. Cao and N. A. Schmid, "Fusion of operators for heterogeneous periocular recognition," *Pattern Recognition Letters*, 2015 (in press).

[122] J. Kim, J. Choi, J. Yi, and M. Turk, "Effective representation using ica for face recognition robust to local distortion and partial occlusion," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 12, pp. 1977–1981, 2005.

[123] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 2, pp. 210–227, 2009.

[124] K. Reddy, G. Babu, and L. Kishore, "Partial face recognition," *IUP Journal of Electrical & Electronics Engineering*, vol. 3, no. 4, 2010.

[125] S. Liao, A. K. Jain, and S. Z. Li, "Partial face recognition: Alignment-free approach.," *IEEE Trans. on Pattern Analysis and Mach. Int.*, vol. 35, no. 5, pp. 1193–1205, 2013.

[126] A. M. Burton, S. Wilson, M. Cowan, and V. Bruce, "Face recognition in poor-quality video: Evidence from security surveillance," *Psychological Science*, vol. 10, no. 3, pp. 243–248, 1999.

[127] H. Sellahewa, S. Jassim, *et al.*, "Image-quality-based adaptive face recognition," *Instrumentation and Measurement, IEEE Transactions on*, vol. 59, no. 4, pp. 805–813, 2010.

[128] A. Abaza, M. A. Harrison, and T. Bourlai, "Quality metrics for practical face recognition," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, pp. 3103–3107, IEEE, 2012.

[129] H. Fronthaler, K. Kollreider, and J. Bigun, "Automatic image quality assessment with application in biometrics," in *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*, pp. 30–30, IEEE, 2006.

[130] E. Tabassi and P. Grother, *Fingerprint image quality.* Springer, 2009.

[131] K. Chang, K. W. Bowyer, S. Sarkar, and B. Victor, "Comparison and combination of ear and face images in appearance-based biometrics," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 9, pp. 1160–1165, 2003.

[132] N. D. Kalka, J. Zuo, N. A. Schmid, and B. Cukic, "Image quality assessment for iris biometric," in *Defense and Security Symposium*, pp. 62020D–62020D, International Society for Optics and Photonics, 2006.

[133] P. Grother and E. Tabassi, "Performance of biometric quality measures," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 4, pp. 531–543, 2007.

[134] K. Nandakumar, Y. Chen, A. K. Jain, and S. C. Dass, "Quality-based score level fusion in multibiometric systems," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 4, pp. 473–476, IEEE, 2006.

[135] K. Nandakumar, Y. Chen, S. C. Dass, and A. K. Jain, "Likelihood ratio-based biometric score fusion," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 2, pp. 342–347, 2008.

[136] K. Kryszczuk and A. Drygajlo, "Improving classification with class-independent quality measures: Q-stack in face verification," in *Advances in Biometrics*, pp. 1124–1133, Springer, 2007.

[137] K. Kryszczuk and A. Drygajlo, "Improving biometric verification with class-independent quality information," *Signal Processing, IET*, vol. 3, no. 4, pp. 310–321, 2009.

[138] J. Zuo and N. A. Schmid, "Global and local quality measures for NIR iris video," in *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 120–125, 2009.

[139] N. D. Kalka, J. Zuo, N. A. Schmid, and B. Cukic, "Estimating and fusing quality factors for iris biometric images," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 40, no. 3, pp. 509–524, 2010.

[140] J. Zuo and N. A. Schmid, "Adaptive quality-based performance prediction and boosting for iris authentication: Methodology and its illustration," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 6, pp. 1051–1060, 2013.

[141] A. Santos, C. Ortiz de Solorzano, J. J. Vaquero, J. Pena, N. Malpica, and F. Del Pozo, "Evaluation of autofocus functions in molecular cytogenetic analysis," *Journal of microscopy*, vol. 188, no. 3, pp. 264–272, 1997.

[142] J. M. Tenenbaum, "Accommodation in computer vision.," tech. rep., DTIC Document, 1970.

[143] E. Krotkov, "Focusing," *International Journal of Computer Vision*, vol. 1, no. 3, pp. 223–237, 1988.

[144] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *Image Processing, IEEE Transactions on*, vol. 16, no. 8, pp. 2080–2095, 2007.

List of Publications

[1] Z. Cao and N. A. Schmid, "Matching heterogeneous periocular regions: Short and long standoff distances," in *Proc. of the IEEE International Conference on Image Processing*, pp. 4967–4971, 2014.

[2] Z. Cao and N. A. Schmid, "Recognition performance of cross-spectral periocular biometrics and partial face at short and long standoff distance," *Open Transactions on Information Processing*, vol. 1, no. 2, pp. 20–32, 2014.

[3] Z. Cao and N. A. Schmid, "Composite multi-lobe descriptor for cross-spectral face recognition: Matching active IR to visible light images," in *Proc. of Automatic Target Recognition XXV Conference in the 2015 SPIE DSS Symposium*, pp. 94760T–94760T–13, 2015.

[4] M. Alkhweldi, Z. Cao, and N. A. Schmid, "Distributed estimation of a parametric field with random sensor placements," in *Proc. of Automatic Target Recognition XXV Conference in the 2015 SPIE DSS Symposium*, pp. 94760R–94760R–10, 2015.

[5] Z. Cao and N. A. Schmid, "Fusion of operators for heterogeneous periocular recognition," *Pattern Recognition Letters*, 2015 (in press).

[6] Z. Cao and N. A. Schmid, "Novel composite operator for cross-spectral face recognition," *SPIE Newsroom*, April 2015.

[7] Z. Cao, N. A. Schmid, and T. Bourlai, "Local operators and measures for heterogeneous face recognition," in *Face Recognition Across the Electromagnetic Spectrum* (T. Bourlai, ed.), Springer, 2015.

[8] Z. Cao and N. A. Schmid, "Composite multi-lobe descriptors for cross-spectral recognition of full and partial face," *Optical Engineering*, 2016 (accepted).

[9] Z. Cao, N. A. Schmid, and X. Li, "Image disparity in cross-spectral face recognition: Mitigating camera and atmospheric effects," in *Proc. of Automatic Target Recognition XXV Conference in the 2016 SPIE DSS Symposium*, 2016.