Graduate Theses, Dissertations, and Problem Reports

2017

Unifying the Visible and Passive Infrared Bands: Homogeneous and Heterogeneous Multi-Spectral Face Recognition

Nnamdi C. Osia

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Unifying the Visible and Passive Infrared Bands: Homogeneous and Heterogeneous Multi-Spectral Face Recognition

Nnamdi C. Osia

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

Doctor of Philosophy
in
Computer Engineering

Thirimachos Bourlai, Ph.D., Chair
Bojan Cukic, Ph.D.
Lawrence Hornak, Ph.D.
Arun Ross, Ph.D.
Kostas Sierros, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia
2017

Keywords: Intra-Spectral, Cross-Spectral, Matching, MWIR, LWIR, Visible, Face, Recognition, Matching, Image Synthesis, Biometrics

Abstract

Unifying the Visible and Passive Infrared Bands: Homogeneous and Heterogeneous Multi-Spectral Face Recognition

Nnamdi C. Osia

Face biometrics leverages tools and technology in order to automate the identification of individuals. In most cases, biometric face recognition (FR) can be used for forensic purposes, but there remains the issue related to the integration of technology into the legal system of the court. The biggest challenge with the acceptance of the face as a modality used in court is the reliability of such systems under varying pose, illumination and expression, which has been an active and widely explored area of research over the last few decades (e.g. same-spectrum or homogeneous matching). The heterogeneous FR problem, which deals with matching face images from different sensors, should be examined for the benefit of military and law enforcement applications as well. In this work we are concerned primarily with visible band images (380-750 nm) and the infrared (IR) spectrum, which has become an area of growing interest.

For homogeneous FR systems, we formulate and develop an efficient, semi-automated, direct matching-based FR framework, that is designed to operate efficiently when face data is captured using either visible or passive IR sensors. Thus, it can be applied in both daytime and nighttime environments. First, input face images are geometrically normalized using our pre-processing pipeline prior to feature-extraction. Then, face-based features including wrinkles, veins, as well as edges of facial characteristics, are detected and extracted for each operational band (visible, MWIR, and LWIR). Finally, global and local face-based matching is applied, before fusion is performed at the score level. Although this proposed matcher performs well when same-spectrum FR is performed, regardless of spectrum, a challenge exists when cross-spectral FR matching is performed. The second framework is for the heterogeneous FR problem, and deals with the issue of bridging the gap across the visible and passive infrared (MWIR and LWIR) spectrums. Specifically, we investigate the benefits and limitations of using synthesized visible face images from thermal and vice versa, in cross-spectral face recognition systems when utilizing canonical correlation analysis (CCA) and locally linear embedding (LLE), a manifold learning technique for dimensionality reduction. Finally, by conducting an extensive experimental study we establish that the combination of the proposed synthesis and demographic filtering scheme increases system performance in terms of rank-1 identification rate.

This work is dedicated to my family and friends. To my late great uncle Reverend Dr. John Kunirum Osia, thank you for the love and support. To the late great Mr. LaMont Toliver, thank you for your encouragement. For Cameron Edward Jones. You are all deeply missed.

Acknowledgements

First and foremost, I would like to offer my deepest thanks to my advisor and committee Chair, Dr. Thirimachos Bourlai for his knowledge, wisdom and support. He has provided me with many opportunities during and after my time in graduate school. He has gone far and beyond his duties and helped me gain understanding into the workings of academia, industry and life. I am truly appreciative of the support and guidance. I initially met Dr. Bourlai through a summer REU at WVU, which was challenging, yet productive. This was when the Multi-Spectral Imaging Lab was initially created. I am truly proud of the accomplishments of Dr. Bourlai, honored to be a lab member, and one of his students.

To my graduate committee members: Dr. Bojan Cukic, It was a pleasure to work with you, Dr. Dawson, and Dr. Bourlai on some CITeR projects, such as when we presented some work at ManTech in Virginia. That was an enjoyable experience for me, and it gave me confidence as well. Dr. Larry Hornak, thank you for your mentorship and guidance, and allowing me the opportunity to work on CITeR projects as a summer intern. I was able to meet and work with Dr. Bourlai, and some other outstanding faculty because of you. Dr. Arun Ross, your biometrics classes were always very enjoyable, and I think that you are an extraordinary communicator. Thank you for your inspiration, input and recommendations for some of my work. Dr. Kostas Sierros, you have provided me with the greatest piece of advice, as it pertains to the writing and presentation of this work. To simply tell my story, and to make sure that this story flows, connecting ideas and chapters. Thank you all for agreeing to be a part of my committee. Your guidance, analysis, and recommendations have been imperative in improving the quality and completeness of this work.

I would also like to acknowledge my friends and colleagues: Dr. Emanuela Marasco, Dr. Nate Kalka, Dr. Asem Othman, Dr. Cameron Whitelam, Dr. Neeru Narang and Michael Martin for their support and numerous conversations we have had in relation to the science, and graduate studies.

I am also indebted to my parents, Dr. Sisom Francis Osia and Dr. Uzo Grace Ozor Osia; their continual support will be forever appreciated. More than once I had questioned whether years of graduate school would be feasible; their moral support helped overcome my reservations. To my older sister Nebechi, thank you for always being there for me and never giving up on your dreams, you have been inspirational. To my younger sisters Chinye and Uche, thanks

for your support and love as well. My small, but immediate family is easily my single greatest asset and I admire you all.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

As it pertains to multi-spectral sensors, the research gap of this work is passive infrared sensing (e.g. MWIR and LWIR), and the benefits of operation within the passive infrared band. The term passive has been recently coined in literature, to signify that illumination is not necessary, allowing operation in some challenging environment, such as the night-time. This is in contrast to the active infrared band (e.g SWIR and NIR), where illumination is necessary for sensor detection. Other benefits of passive infrared sensing technology are that it is non-contact, eye-safe, non-invasive, and can operate real-time. Passive infrared sensing can be used for a number of applications including geothermal mapping, health and safety, building inspections, maintenance, and in the field of biometrics.

Biometrics is the science and technology of using traits, preferably inherent, for the identification of individuals. There are a number of biometric traits that exist, using either physiological (fingerprint, face, iris, voice and hand geometry) or behavioral characteristics (gait, signature, and keystroke), being researched substantially, and used in academia, industry, and government applications [3], [4]. Face-based recognition (FR) systems, in par-

ticular, are gaining interest because face has several advantages over other biometric traits: it is non-intrusive, understandable, and can be captured in a non-cooperative or covert (when necessary) manner at variable standoff distances.

Face biometrics leverages tools and technology in order to automate the identification of individuals. In most cases, biometric face recognition (FR) can be used for forensic purposes, but there remains the issue related to the integration of technology into the legal system of the court. The biggest challenge with the acceptance of the face as a modality used in court is the reliability of such systems under varying pose, illumination and expression, which has been an active and widely explored area of research over the last few decades (e.g. same-spectrum or homogeneous matching). In real-world scenarios (military and law enforcement) we have to deal with harsh environmental conditions characterized by unfavorable lighting and pronounced shadows [5]. Such an example is a night-time environment [5], where human recognition based solely on visible spectral images may not be feasible. In order to deal with such difficult FR scenarios, multi-spectral camera sensors are very useful because they can image day and night. Thus, recognition of faces across the infrared spectrum has become an area of growing interest ([5–8]).

The heterogeneous FR problem, which deals with matching face images from different sensors, poses a greater challenge than the variation of challenges introduced during image acquisition. This is because it is difficult for many algorithms to compensate for such factors without the use of large training sets. In many heterogeneous FR algorithms, large and readily available datasets need to exist for training, which may be computationally expensive for some algorithms. This is less of an issue with same-spectrum or homogeneous matching systems. Heterogeneous FR applications may include the authentication of an individual at a crime scene using gallery images acquired by CCTV (visible band images) with available probe images of potential suspects (e.g. sketch, composite) [9]. In this work we are concerned primarily with visible band images (380-750 nm) and the passive infrared IR spectrum.

Through our results we find that the homogeneous FR experiments yield high FR accuracy regardless of spectrum of operation, while the heterogeneous FR problem is a bit more challenging. If the heterogeneous FR problem was easily solved, one may argue that there is no additional benefit or need for use of multiple approaches and sensors with similar end results. However, it is important to note that the methods detailed in this work, in conjunction with the research gap, are scenario-dependent. The combination of multiple biometrics, and their corresponding features or the intermediate decisions that are carried out in order to perform an analysis is known as multi-modal fusion. With respect to this work, although information obtained from different sensors of the same biometric are different, they may be fused in order to a achieve a stronger, and more confident biometric. Even at the decision-level, each individual identity decision can be combined via a fusion process.

1.2   Goal

In this dissertation we propose a framework, for both homogeneous and heterogeneous FR systems using multi-spectral sensors. For homogeneous FR systems, we formulate and develop an efficient, semi-automated, direct matching-based FR framework, that is designed to operate efficiently when face data is captured using either visible or passive IR sensors. Thus, it can be applied in both daytime and nighttime environments. First, input face images are geometrically normalized using our pre-processing pipeline prior to feature-extraction. Then, face-based features including wrinkles, veins, as well as edges of facial characteristics, are detected and extracted for each operational band (visible, MWIR, and LWIR). Finally, global and local face-based matching is applied, before fusion is performed at the score level. Although this proposed matcher performs well when intra-spectral FR is performed, regardless of spectrum, a challenge exists when cross-spectral FR matching is performed. The second framework is for the heterogeneous FR problem, and deals with the issue of bridging the

gap across the visible and passive infrared (MWIR and LWIR) spectrums. Specifically, we investigate the benefits and limitations of using synthesized visible face images from thermal and vice versa, in cross-spectral face recognition systems when utilizing canonical correlation analysis (CCA) and locally linear embedding (LLE), a manifold learning technique for dimensionality reduction. Finally, by conducting an extensive experimental study we establish that the combination of the proposed synthesis, restoration, denoising, and demographic filtering scheme increases system performance in terms of rank-1 identification rate

1.3 Contributions

This dissertation details the following set of novel contributions to the science:

1. The proposal of an efficient and automated homogeneous FR framework which is applied on full-frontal face images acquired using visible, MWIR, and LWIR sensors. The pre-processing pipeline and feature extraction methodology ensure that no artificial information is introduced into the face images. The pre-processing pipeline is instrumental in standardizing the face images across our database, which is necessary prior to feature extraction and subsequent experimental evaluation.

2. Designed experiments to quantitatively evaluate the benefits of global vs. local based matchers, specifically the performance of such matchers individually, and after fusion at the score level. Experiments are also designed to evaluate the robustness of the proposed approach with the challenging problem of facial occlusion. Partial face matching is simulated through the partitioning of the whole face into regions.

3. Formulation of a framework for the heterogeneous FR challenge, when matching visible face images to thermal face images, and vice versa. In constructing our cross-spectral

face recognition system, the benefits and limitations of using canonical correlation analysis (CCA) and manifold learning dimensionality reduction (LLE) are investigated.

4. Demonstration of challenges associated with face matching in cross-spectral condition, including utilization of three different methods for pre-processing face images before feature extraction methods are applied. The use of a a cascaded post-synthesis restoration and denoising methodology, which helps demonstrate FR accuracy is thus improved (e.g. practical scenarios). Cohort-specific matching (per gender) instead of blind-based matching (when all images in the gallery are matched against all in the probe set) is explored, in order to increase face recognition accuracy. An extensive experimental study is conducted to establish that the combination of the proposed synthesis, restoration, denoising, and demographic filtering scheme increases system performance in terms of rank-1 identification rate.

1.4 Organization

In Chapter 2 we present background information pertaining to biometrics and FR techniques, multi-spectral face sensing and properties of the active spectrum and passive IR band. In Chapter 3 we present a homogeneous (intra-spectral) framework for face image feature extraction and matching for our assembled database, utilizing skin segmentation and standardization of face images across datasets. We conclude Chapter 3 with empirical evaluation of our multi-spectral matching approach for the homogeneous, intra-spectral FR scenario. In Chapter 4, we detail methodology to deal with heterogeneous, cross-spectral matching, since the physiological and geometric face-based matcher is efficient only in same-band face matching scenarios. The approach attempts to bridge the gap between the visible and passive IR band using image synthesis, image reconstruction, denoising, and cohort filtering (gender classification). Chapter 4 is concluded with experimental evaluation of our

heterogeneous databases. Finally, Chapter 5 concludes the work by providing a summary of the accomplishments as well as future directions for research.

Chapter 2

Biometrics

This chapter presents a succinct review of the related work in the fundamental areas of biometrics, and face recognition specifically. Relevant fields of interest with respect to this work include face image matching techniques and multi-spectral face matching approaches. The intent is to provide the reader with a basic understanding of the field.

2.1 Biometrics

Biometrics utilizes pattern recognition systems that are usually characterized by the following three traits: (1) The system recognizes individuals using verification or identification. (2) The system is either automated or semi-automated. (3) There are biological characteristics used which are either physiological (e.g. fingerprint, face, iris, voice and hand geometry) or behavioral (e.g. gait, signature, and keystroke) (see Fig. 2.1). The *ideal* biometric trait has the following properties:

- permanence - the biometric trait should not change over time.

- uniqueness - the biometric trait should be unique to each individual in the system.

- circumvention (non) - the biometric trait should not easily be spoofed, counterfeited, or imitated.

- collectability - the biometric trait is easily acquired and measured.

- universal - the biometric trait is possessed by all individuals in the system.

- acceptability - individuals in the system accept the trait and are willing to have it collected and assessed.

- performance - speed and accuracy (in terms of matching) of the biometric trait in the system.

Table 2.1: Comparison of Biometric Traits using Characteristics [1]

Biometric Trait	Universality	Uniqueness	Permanence	Collectability	Performance	Acceptablity	Circumvention
Face	High	Low	Medum	High	Low	High	Low
Fingerprint	Medium	High	Hight	Medium	High	Medium	High
Hand Geometry	Medium	Medium	Medium	High	Medium	Medium	Medium
Iris	High	High	Hight	Medium	High	Low	High
Retinal Scan	High	High	Medium	Low	Hight	Low	High
Signature	Low	Low	Low	High	Low	High	Low
Voice Print	Medium	Low	Low	Medium	Low	High	Low
Facial Thermograms	High	High	Low	High	Medium	High	High

Biometric traits are being researched substantially, and used in academia, industry, and government applications [3, 4]. A table comparing some common biometric traits based on characteristics can be seen in Table 2.1.

2.1.1   Design and Performance Evaluation

There are four components that make up a biometric system: sensor, feature extractor, database and matcher. The sensor is used to acquire raw biometric data. The feature
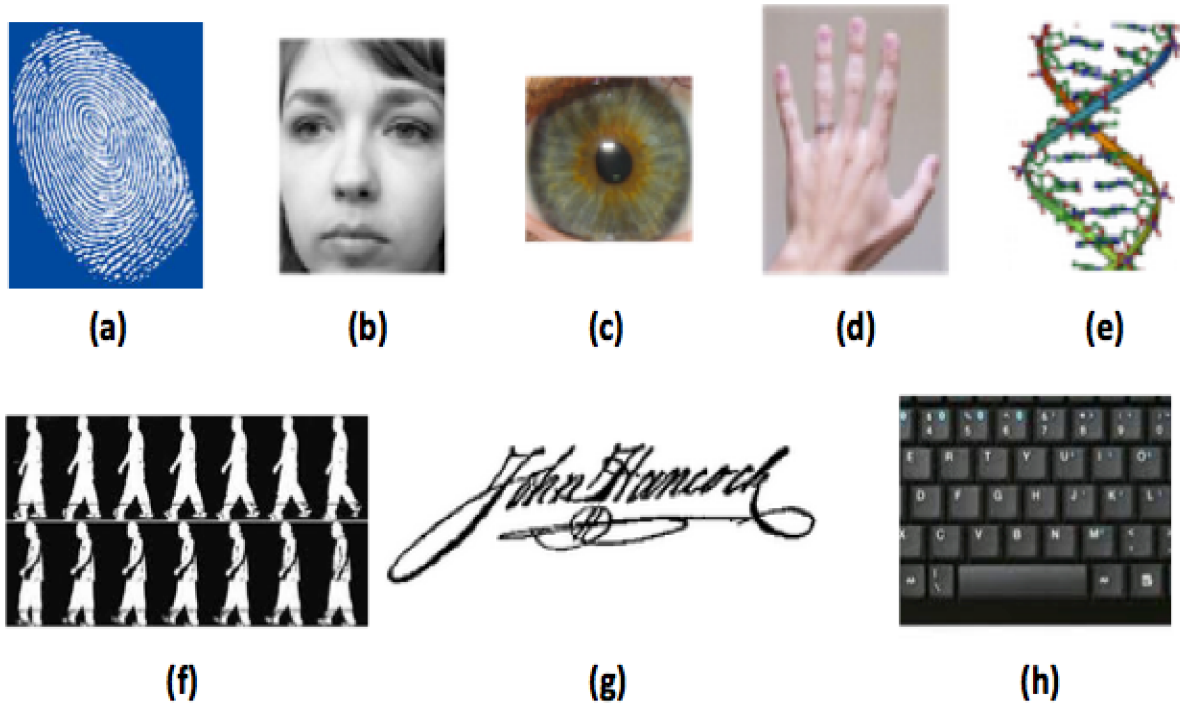
Figure 2.1: Top row represents sample physiological biometric traits: (a) fingerprint, (b) face, (c) iris, (d) hand, and (e) DNA. Bottom row represents sample behavioral biometric traits: (e) gait, (f) signature, (g) keystroke.

extractor processes the biometric data to extract a compact feature set. When an extracted feature is placed into a database, which is used for storage, it becomes a template. The matcher compares two extracted feature sets (or templates) and generates a score. The two stages of a biometric system are enrollment and recognition. Enrollment is the placement of a user's biometric traits into a database, which may be centralized (e.g. saved locally) or decentralized (e.g. saved remotely). During the time of enrollment, templates are usually tagged with the date and time, and updated or aged over time.

2.1.2   Authentication

Biometric systems are designed for authentication, which has two different modes: verification (Am I who I claim to be?) or identification (Who am I?) [10, 11]. Verification is a 2-class problem (e.g. genuine and impostor). On the other hand, identification is a multiclass problem. In recent work, Blackburn [12] has added another task to a biometric system known as a watch list.

- Verification:(Am I who I claim to be?): This mode is utilized when the individual provides an alleged identity. The biometric system then performs a search (one-to-one), comparing the input biometric trait with the biometric template stored in the database. If a match is made, the identity of the person is verified [10] as correct or incorrect. The verification performance of a biometric system is evaluated through the Receiver Operating Characteristic (ROC) curve.

- Identification (Who am I?) : This mode is utilized when the identity of the individual is not known in advance. There is an exhaustive search (one-to-many) of the entire template database for a match to the input subject. If a match is found, the individual is identified [10]. It is important to note that a match does not mean that input is identical, but rather falls within a given threshold [11, 13]. The identification perfor-

mance of a biometric system is evaluated through the cumulative match characteristic (CMC) curve.

- Watchlist: In this mode the subject does not claim to be anybody. The biometric trait of the subject is compared with the stored samples in a list to see if the individual concerned is present [12, 14]. Different types of watchlist modes could be comparing a flight passenger towards a database of known terrorists, or comparing a John Doe patient with a list of missing individuals. When an individual is found to be similar to one or more samples in the list given a predetermined threshold, the system will issue an alert and return samples that were found to be similar. When this alert is issued for an individual that is actually present in the list based on score, it is called a correct detection and identification. A false alarm is when an alert is received even though the individual is not present in the list. The frequency at which false alarm encounters occur is called the false alarm rate [12]. In an ideal system we want the false alarm rate to be 0% and the correct detection and identification rate to be 100%.

2.1.3   False Acceptance and False Rejection

When evaluating a verification system, the two types of errors made are false rejection and false acceptance [11]. These are measured as rates, known as the false accept rate (FAR) and false reject rate (FRR). False acceptance is when an impostor is accepted as another individual. The FAR is the proportion of impostors accepted, otherwise known as the false match rate (FMR). False rejection is when a genuine user is rejected. FRR is the proportion of genuine users rejected or false non-match rate (FNMR). Low FAR is required in high security systems (i.e. nuclear facilities), while a low FRR is required in systems where we focus on user convenience. The point where the FRR and FAR are equal is called the equal error rate (EER).

2.1.4   Quantitative Biometric Measures

There are two primary quantitative characteristics that are used in measuring and assessing the performance of a biometric system, namely the Receiver Operating Characteristic (ROC) and Cumulative Match Characteristic (CMC). An ROC curve measures verification performance and relies on aggregate statistics of match scores corresponding to all biometric samples. In contrast, the CMC curve measures identification performance and is based on the ordering of match scores that correspond to each biometric sample in a closed-set.

2.1.5   Receiver Operating Characteristic (ROC)

The Receiver Operating Characteristic (ROC) is used as a metric, mostly for 1:1 biometric systems to assess performance in terms of verifying individuals. The ROC curve is defined as plot of the true positive rate against the false positive rate for a variety of thresholds in a system. In other words, the ROC is a plot of the False Reject Rate (FRR) versus False Acceptance Rate (FAR) for a biometric system. An important point on the ROC curve is the Equal Error Rate (EER), which corresponds to an equal probability of miss-classifying a positive or negative sample.

2.1.6   Cumulative Match Characteristic (CMC)

The Cumulative Match Characteristic (CMC) is used as a metric in 1: m identification systems to assess the accuracy of FR algorithm. The CMC curve is a plot of the identification rate at rank-k. A probe (or test sample) is given rank-k when the actual subject is ranked in position k by an identification system. Thus, a rank-1 outcome is considered a correct identification. The identification rate is an estimate of the probability that a subject is identified correctly at least at rank-k. Hence, identification rate is necessarily an increasing function of k.

2.1.7 Multimodal Systems

There does not seem to be a single biometric trait that is able to be completely accurate and reliable all the time. Fingerprints can be counterfeited [15] and changed through lacerations and bruises[11]. Face recognition systems may may not utilize liveliness detection during sensing and authenticate an impostor. In order to increase the accuracy and performance of biometric systems, we can look to multimodal biometric systems [11, 16, 17]. This is the combination of more than one biometric trait or feature within a system. Hong et al. [18] demonstrated in their manuscript a multimodal biometric system combining face recognition with fingerprint recognition. The system showed huge gains in recognition performance and accuracy. Early applications of face recognition limited the search to the top five people, followed by fingerprint recognition to make a final decision. Recognition along with computation performance was increased with the setup of this system. Also, a combination of different methods of recognition within the same biometric feature could be used to alleviate drawbacks in one system by combining it with another system that does not have the same drawbacks [11]. For example, combining a facial recognition system that has good performance on faces exposed to illumination changes, with a system that has greater performance on face images taken in a controlled environment. This allows for greater flexibility and accommodation within the biometric system. Match scores that are generated by noisy inputs tend to vary largely. By utilizing multiple sensors that capture different biometric traits, much of this variance could be accounted for [17]. A multimodal system may also be useful with circumvention and anti-spoofing [11, 17, 19], making it more challenging for an impostor to provide different features of a genuine user.

2.1.8   Fusion

In biometrics, fusion is the method of combining data in order to help alleviate the problem of dimensionality. Fusion is usually done either prior to matching or after matching. The two types of fusion prior to matching are: (1) image level fusion and (2) feature level fusion. In image level fusion, raw data acquired from biometric sensors are combined. However, it is impossible to design a method for image fusion that is universally applicable due to the differences in images being fused. In recent literature, [20], image fusion methods are broken down into three main stages (e.g image transform, fusion of coefficients, and inverse transform).

In feature level fusion, feature sets from multiple biometric sources are combined. After feature vectors are arranged and grouped accordingly, they are fused and an identity is declared using the joint feature vector. A pre-requisite for fusion at the feature level is the extraction of features given raw input images. These features may take a number of forms, including pixel intensities, or edge and textural information [21]. The features that should be chosen for fusion are scenario dependent, and limited by image quality and resolution. It is also possible that fusion of features such as shapes or image segments are localized to specific areas. To facilitate proper feature extraction, it is important to ensure that the spatial resolution is similar between multiple images if image data is used for feature extraction. Therefore, geometric normalization should be mandated for the input raw data at the feature level stage.

Fusion can also occur after matching in order to increase FR performance in different scenarios. The three types of fusion after matching are: (1) score level fusion, (2) rank even fusion, and (3) decision level fusion. In score level fusion, matching scores from different matching techniques or matchers are combined to increase final score. In rank level fusion (as the name suggests), the output of biometric systems dealing with identification are sorted

with ranks decreasing, and ranks combined. In decision level fusion, decisions on matches are combined to create a final matching decision [3]. Because score normalization is utilized within this work, we will briefly describe score level normalization and fusion techniques [22]. An output matching score (raw) can be denoted as s, from the set S of all scores per matcher, and parallel normalized score can be denoted as n.

- Min-Max: This fusion scheme maps the raw scores to the [0, 1] range. The values of $\max(S)$ and $\min(S)$ specify the end points of the score range:

$$n = \frac{s - \min(S)}{\max(S) - \min(S)}. \tag{2.1}$$

- Z-score: This fusion scheme changes the scores to a 0 mean valued distribution with standard deviation of 1. In the equation below, the mean and std represent the arithmetic mean and standard deviation, respectively:

$$n = \frac{s - \mathrm{mean}(S)}{\mathrm{std}(S)}. \tag{2.2}$$

- Tanh: This fusion approach belongs to the family of robust statistical techniques [23]. It is a way of mapping raw scores to the (0, 1) range:

$$n = \frac{1}{2} \left[tanh\left(.01\frac{s - \mathrm{mean}(S)}{\mathrm{std}(S)} \right) + 1 \right] \tag{2.3}$$

There are a number of approaches which can be used for score fusion, but we briefly review three well known approaches, commonly known as, simple-sum, min-score, and max-score. The quantity n_m^i is representative of a normalized score for matcher m ($m = 1, 2, ..., M$, where M is the number of matchers) applied to the subject $i(i = 1, 2, ..., I)$ where the number of subjects in the database is I. The fused score for subject i can be represented

as f_i.

- Simple-Sum: $f_i = \sum_{m=1}^{M} n_i^m, \forall i$

- Min-Score: $f_i = \min(n_i^1, n_i^2, ..., n_i^M), \forall i$

- Max-Score: $f_i = \max(n_i^1, n_i^2, ..., n_i^M), \forall i$

2.2   Face Recognition

Due to the advantages of FR over other biometric traits, face-based recognition (FR) systems
have gained interest because of the following traits:

- Voluntary action isn't necessary (i.e. placement of fingers for fingerprinting or eye for
  ocular biometrics).

- Ideal for covert (security and surveillance) applications because of non-intrusiveness
  and obscurity.

- Commercially available FR algorithms may be able to compensate for challenging
  factors.

- Greater access to information (i.e. social networks).

- Subjects more willing to be cooperative.

2.3   Properties of the Infrared Band

The infrared band is divided into the active and passive IR bands. The active IR band
consists of the Near IR (0.7 - 0.9 μm) and the Short-Wave IR (0.9 - 2.5 μm) bands. The
NIR band can be represented by remote sensing in the environment, especially with respect

to vegetation, and is frequently utilized in optic telecommunication due to low attenuation losses in the SiO_2 glass (silica) medium. The SWIR band has generally been characterized by the measurement of surface water in soil (absorption) as well as long-distance telecommunications. While data is being acquired in the active IR band, there is usually an external light source which may be easily detectable (e.g. NIR) or not (e.g. SWIR), which aids in illuminating the subject's face. In the passive IR band, the data is acquired using a camera sensor with the ability to detect IR radiation as heat being emitted from the subject's face. The passive IR band can be further sub-divided into the Mid-Wave Infrared (MWIR) and the Long-Wave Infrared (LWIR) band, where MWIR ranges from 3-5 μm, while LWIR ranges from 7-14 μm. The MWIR band has a frequency from 37-100 THz and photon energy of 155-413 meV. The wavelength of the MWIR band exists in the same atmospheric gap in which homing heads of heat seeking missiles were designed to operate, targeting the IR signature of an aircraft, which is usually the exhaust plume of a jet engine. The MWIR band detects temperatures that are somewhat above body temperature, approximately 98.6° F (37 ° C). The LWIR band has a frequency of 20-37 THz and photon energy of 83-155 meV. The LWIR band is characterized by temperatures that are slightly higher than room temperature, approximately 70 ° F (21 ° C) [24].

One of the limitations of the passive IR band, or wavelengths greater than 3 μm, is that glass optics can't be utilized [25]. Almost all of the energy in the MWIR band is blocked by the presence of glass, and energy in the LWIR band appears opaque. Because glass windows aren't transparent in the passive IR band, they may appear darker or brighter depending on the temperature. Due to the constraints imposed by the use of glass in the IR imaging systems, many optics are made of costly materials such as germanium, instead of cheap glass lenses. Another disadvantage of energy radiated in the passive IR bands is that it is not transmitted through water. Imaging of objects or scenes that may be coated with water in the IR passive band may appear duller, due to the washed out thermal contrast.

2.3.1   Face Recognition in Passive Infrared Band vs. Active In-frared Band

Face recognition (FR) has been an active and widely explored area of research over the last few decades, with usage in a wide range of military and law enforcement applications. The majority of FR research is focused primarily on visible band images (380-750 nm). Although visible band FR systems are considered to be efficient when face images are captured under controlled conditions, variation in illumination is still considered to be a challenging problem. Unfortunately, in environmental conditions that are characterized by adverse lighting and conspicuous shadows (such as nighttime environments [26]), FR based solely on visible band images may not be feasible ([27, 28]). Thus, recognition of faces in the infrared spectrum has become an area of growing interest ([6–8]).

Recent research in the area of FR at different bands illustrates that working in the passive IR band is very challenging [29]. However, the main benefit of using thermal-based camera sensors is the capability to use FR systems when operating at difficult environmental conditions, such as low light or complete dark environments. This allows for the potential to have obscure detection and acquisition of face images of different subjects, given that they are within an appropriate (i.e. depending on the sensor capabilities) distance from the camera sensors used. Although the appearance of MWIR and LWIR face images look visually similar (see Fig. 2.5), they both capture different characteristics of the human face. MWIR has both reflective and emissive properties and works better for long-range operations. On the other hand, LWIR consists primarily of the subject's emitted radiation. Face images acquired in the passive band aren't normally affected by external illumination. Another advantage is that unique physiological features for each individual (useful for human recognition), such as subcutaneous information (veins), edges of facial features, and wrinkles, are detectable in this band. Finally, image acquisition in the passive band is unobtrusive, and therefore, FR

in this band is considered very useful for both military and law enforcement applications.

The main benefits of passive IR over the active IR band are: (1) passive IR imagery can be acquired without any external illumination in day or night-time environments; however, in the active IR band, the target may require an external light source in order to be observable; (2) anatomical features (e.g., vein patterns) that may not be observable in the active IR band, may be observable in passive IR band instead; (3) in passive IR images, background clutter is not visible and the tasks of face detection, skin localization, and segmentation can be easier and more reliable than in active IR and visible bands; (4) passive IR imaging is unobtrusive, and therefore FR in this band can provide an operational advantage over other bands for both military and law enforcement applications.

2.3.2   Challenges of Face Recognition

Although face recognition has advantages over other biometric traits, there are still challenges presented within a face recognition system. Face recognition performance significantly deteriorates when there are variations in facial pose, illumination, and expression (collectively known as PIE) are introduced [30] (see Fig. 2.2). Other variables such as the quality of an image (e.g., ratio, resolution, compression, blur), time lapse or facial aging, as well as occlusion contribute to face recognition errors [31, 32]. During the assessment of face recognition within videos, issues such as segmenting the face in varying illuminations [33] and compression artifacts [34] must be considered. One of the most challenging tasks in face recognition is the comparison of two images that have been captured using alternate sensors or imaging modalities (e.g. infrared images and visible images) or with significant time lapses (e.g. face images of the same person over a decade or longer). This area of study is known as heterogeneous face recognition. Some approaches which have received success pertaining to this limitation of face recognition extend the capabilities of the system to covert capture scenar-

Figure 2.2: Face recognition challenges: Pose, Illumination and Expression (PIE).

ios (e.g face recognition at long distances and face recognition in nighttime environments), or situations where face images exhibit changes through the effects of aging (aging-invariant face recognition). While the majority of face recognition research aims to mimic the capabilities of humans, heterogeneous face recognition offers the ability to recognize faces tat humans are not capable of.

2.4   Human vs. Computer Recognition of Faces

The way we perceive things as humans has a lot to do with the way we see them. It is a difficult task to replicate a computer system that recognizes faces in the same manner that humans do. The human visual system model has been used to try and understand biological and physiological processes associated with visual systems. The human visual system is

the primary method of sensing for most humans regardless of our environment. The visual system transmits visible light waves that are detected ocularly, to a region of the brain known as the visual cortex. Upon arrival of the light information at the cortex, it is transmitted through ventral streams for processing. Studies have shown that processing of information regarding a person's location with respect to his or her environment occurs in the mental stream. The mental stream is also responsible for performing object recognition related tasks from the visible light waves that are sensed [35]. There is well supported experimentation that shows that the fusiform face area is a dedicated region of the brain that has the task of associating an identity with that face as it enters the ventral stream [36, 37]. Fig. 2.3 shows the path visible light waves take into the human visual system and to the cortex.

Face recognition research aims to try and mimic an entire region of the brain that is primarily dedicated to one task: extracting information from human faces. However, there are factors that human perception uses for similarity ratings that do not translate to computer similarity, and vice versa. One of these factors is illumination. Experiments for computer based face recognition show that changes in illumination have greater effect on similarity than alternating identity, although humans are less affected by such changes [38]. External cues and other soft biometric traits are inconspicuously used by humans to help identify faces. These cues include broad shoulders and round faces, race of individuals, and on-face based features at different levels, such as eyebrows [39, 40]. Even at a standoff distance, the eyebrows play an vital role in recognition of the face because eyebrows are less affected by shadow and illumination changes. Computer based face recognition techniques should move toward use of different cues and feature levels in order to help mimic the natural way humans perceive faces.
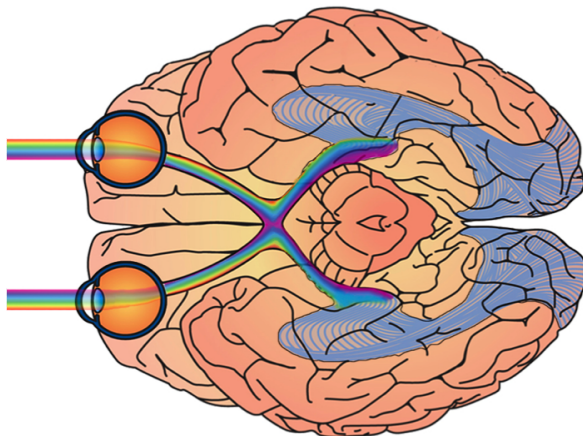
Figure 2.3:   Illustration of the path visible light takes into the human visual system and through the ventral stream to the brain cortex.   This image is taken from http://thewayeyeseesit.com/2013/05/11/the-brains-visual-cortex-corrects-color-casts-naturally/.

2.5   Feature Categorization

Categorization of features is utilized to provide a better understanding of standardization across biometrics. This categorization is beneficial in two ways: (1) helps facilitate metrics for measuring images that may be used legally, and (2) helps improve accuracy of matchers through precise selection of features. Generally speaking, there are usually three levels of features within each biometric trait: Level 1, Level 2, and Level 3. Level 1 features consist of characteristics that are easily observable. With regard to face recognition, these may include skin color, gender and overall appearance or shape of the face. Level 2 features consist of localized face information that needs specialized cortex processing. This may include features relating to structure or relationships among components, and precise shape. Level 3 features consist of characteristics that are not easily seen by the human eye, and may not appear without high definition sensing. Salient features in fingerprints are categorized into three levels: Level 1 features encompass the global structure or ridge pattern (e.g. arch, loop, whorl). Level 2 features consist of minutiae location and orientation, and are
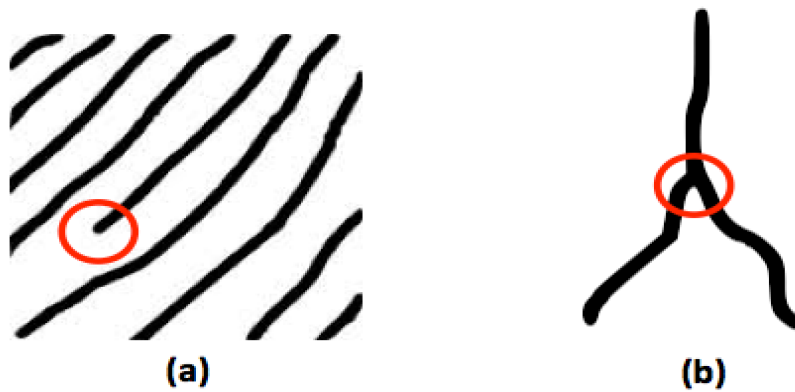
Figure 2.4: Visual representation of fingerprint minutiae (Level 2) points: (a) ridge end point and (b) bifurcation point.

primarily used for matching. Level 3 features consist of information available at higher spatial resolutions, such as dots and ridge width. In comparing fingerprint feature categorization to face, fingerprint features are accepted by both forensic scientists and fingerprint vendors. In this work we utilize Level 1 feature contours (on-face ridges and valleys), and Level 2 fingerprint features, namely minutia set (e.g. bifurcation and end points) for face recognition. Ridge end points and bifurcation points are two basic minutiae points that are frequently used for matching in minutiae based techniques. The ridge end point is the point at which the ridge terminates. The bifurcation point, is the point where a single ridge splits into two. A visual illustration of a ridge end point and bifurcation point can be seen in Fig. 2.4. Fingerprint features have over a century of history and experimentation when compared to traditional face-based features [10]. Furthermore, features used by fingerprint matchers are compact and have physical interpretation in terms of the ridge patterns. Features extracted by humans are not easy to precisely describe, therefore it is difficult for them to be utilized by face recognition systems. A major benefit of this work is that we utilize reliable, dependent and well-defined fingerprint-based features for computer-based face recognition as well as human face examiners.

2.5.1   Face Detection and Alignment

The initial step in many face recognition systems is the detection and alignment of images in the face. This is often viewed as a preprocessing step, and this stage is critical in order to detect the presence of a face in a digital image, and align that detected face with the spatial coordinate system. The face detector proposed by Viola and Jones [41], uses a cascaded classifier in conjunction with images represented using a set of Haar-like features. While many methods have been proposed to improve upon Viola and Jones detector, it serves as an excellent baseline for state of the art performance [42] within the visible spectrum. Face alignment is usually done by first detecting the location of some fixed set of anthropometric landmarks on the face. In most face recognition systems, these landmarks are the center of the two eyes. Using the two eye locations, a 2D affine transformation is performed to fix the angle and distance between the two eyes. More advanced approaches may utilize 3D affine transformations, or other methods of alignment, on a more permanent set of landmarks (e.g., mouth, nose and chin).

2.5.2   Photometric Normalization

Variations in illumination is one of major limiting factors and challenges within a face recognition system. Over time several methods have been proposed to compensate for illumination changes [43]. The representation of the photometric normalization technique after it is applied to a face image depends on spectrum of operation. The application of photo-metric normalization approaches which include contrast limited adaptive histogram equalization (CLAHE), my be used to facilitate both intra-spectral and cross-spectral matching [44]. Other effetive methods for photometric normalization include single-scale self quotient image (SQI) and the Difference-of-Gaussian (DoG) methods [2].

- Contrast limited adaptive histogram equalization (CLAHE): For experiments in this

work, a local region of 8×8 is used in the image along with histogram equilization, as seen in equation 2.4, on each sub-region.

$$f(n) = \frac{(N-1)}{M} \times \sum_{k=0}^{n} h(k) \tag{2.4}$$

The amount of pixels and gray level bins in each sub-region, are depicted here as M and N, and the histogram of each sub-region is h. Each histogram is normalized by CLAHE so that the height falls under the .01 clip limit threshold, in order to better the contrast without increasing noise. More precisely, gray level counts underneath the predetermined clip limit are uniformly redistributed among gray levels underneath the clip threshold. Finally, bilinear interpolation is utilized to combine each sub-region.

- Self Quotient Image (Single-Scale SQI): According to the model known as the Lambertian, which is named after Johann Lambert who introduced perfect diffusion, the process for formulating an image is described as:

$$I(x, y) = \rho_w(x, y)n(x, y)s, \tag{2.5}$$

where $\rho_w(x, y)$ is the albedo of the facial surface, n is the surface normal, and s is the reflection of the lighting. It is important that the extrinsic factor s is separated from ρ and n in order to reduce the impact of illumination. The self-quotient image, Q, of I is defined as [45, 46],

$$Q = \frac{I(x, y)}{I(\hat{x}, y)} = \frac{\rho_w(x, y)n(x, y)s}{G * [\rho_w(x, y)n(x, y)s]}, \tag{2.6}$$

where $\hat{I}$ is the flattened version of I and G is the smoothing kernel.

- Difference-of-Gaussian (DoG) Filtering: A type of normalization that is usually used in IR images in order to maximize the visibility of edges and other features. This normalization is highlighted by the subtraction of one blurred version of the first gray scale image from another, that is less blurred than the first version. The blurred images can be acquired through the convolution of the first gray scale image with two Gaussian kernels having different standard deviation, e.g., σ_0 and σ_1:

$$D(x, y | \sigma_0, \sigma_1) = [G(x, y, \sigma_0) - G(x, y, \sigma_1)] * I(x, y). \tag{2.7}$$

  The symbol $*$ is the convolution operator, and the Gaussian kernel function based on σ is defined as:

$$G(x, y, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x^2+y^2)/2\sigma^2}. \tag{2.8}$$

2.6 Face Image Matching Techniques

Face recognition systems utilize independent and unique face characteristics from individuals, but system performance may be influenced by approach and design of system [47]. There are a good amount of face recognition techniques (e.g. appearance, model and texture), as well as matching approaches that may be utilized and combined for optimal system performance.

- Appearance-based methods (i.e. PCA and LDA [48], which use Eigenfaces) derive feature vectors based on pixel values of face images and transform the features to low-dimensional salient features. A technique developed by [49] for efficiently representing pictures of faces using principal component analysis (PCA) served as the motivation for eigenfaces, which was developed at Massachusetts Institute of Technology (MIT) [50–52]. Variations of eigenfaces are often used as the foundation for a number of different

face recognition methods. However, it has been challenged that this technique does not bear any similarity to the way humans recognize and measure similarity between faces. According to Woodward et al. [53] the mathematical properties of the eigenface representation and matching process have been shown to achieve favorable results in certain minimally controlled environments. The eigenface recognition method is best utilized in well-lit, frontal image capture situations [54] like most face recognition technology.

- Model-based methods (i.e. Elastic Bunch Graph Matching [55]) encodes aspects of the face by using a face model that is created using the subject's face appearance and shape. The main advantage of this methodology is that it can provide face recognition that is invariant to affine transformations and localized changes in facial expressions [55–58]. In Elastic Graph Matching, local features are extracted at controlled and pre-determined locations of the face. The distances between these nodes are measured as well. Some of the features are more reliable and important for recognition than others in the set. In literature, there has been an approach with the use of weights [59] due to the difference in importance of the nodes. An extension of the Elastic Graph Matching approach has been introduced that utilizes several images of the same individual, usually from different angles and poses. Each node on the graph then contains a set of values. This improves the recognition because it will be more robust to differences in posture, poses and facial expressions. Elastic Graph Matching is called elastic because the match is preserved approximately instead of rigidly [57].

- Local feature Analysis (LFA) [51, 54] is one of the most commonly used facial biometric technology today, and can accommodate for some changes in facial expression and aging. Local feature analysis refers to a subset of algorithms that extract a set of geometrical features and distances from facial images, and uses those features as

the basis for representation, comparison and matching. The features used are typically on-face(e.g., eyes, mouth, nose, jaw line, eyebrows and cheeks). These features are represented with their position, size and general outline. The good performance compared with some other approaches, are among the factors that has made this technique popular. One disadvantage for this approach is that it is dependent on a relatively stable acquisition environment and image quality.

- Texture-based (i.e. LBP and SIFT) approaches extract textural features from images and try to find patches or relationships between pixels.

2.6.1   Geometric Features

The length and width of a nose, position of a mouth, and shape of a chin are all types of geometric features. They are reliable face traits that can be used to identify an individual. The Euclidean distance (finding nearest vector) is a metric which can be utilized to find the closest match. Some of the most renowned work in face recognition was achieved using features that are geometric [60, 61]. A benefit of using these type of geometric features as a basis for face recognition is that even with low quality spatial resolution and noisy images, matching is feasible. Although the face image may not have some of the Level-2 and Level-3 features we are accustomed too, its overall geometrical shape and configuration can be extracted for face recognition. Automating the extraction of these type of features is challenging because the feature extraction will be extremely sensitive to the scaling and rotation of a face in the image plane [60].

2.6.2   Template Matching

Template matching is the extraction of facial features or regions and the comparison of these with stored facial features or regions from known subjects. An example is the straightforward

approach of comparing grey-scale intensity values for face recognition used by [62]. However, there are other approaches for template matching that are far more complex. A number of these methods my involve extensive pre-processing and transformation of the extracted facial features or regions. For example, Turk et al. [50] used PCA, also known as the eigenface method, for pre-processing. Wiskott et al. [55] pre-processed extracted regions using Gabor filters to encode Elastic Graphs. Template based approaches offer good recognition accuracy in addition to the usual high recognition speed and small memory requirements offered by many feature based strategies [60].

2.6.3 Global vs. Local Matching

There are trade-offs to the type of features (e.g. global or local) that are used in matching face templates regardless of the approach chosen. Global-based features come from the entire face and are characterized by vectors fixed in length, therefore they are compared in a time efficient manner. Local-based features are detected sets of points which may vary depending on the local information of the subject. Local-based features are less susceptible to variations in the face, but computation may be exhaustive. The computational complexity of a face matching system may also be increased using a training-based approach instead of utilizing direct matching. Although, under certain conditions, a larger training data set may increase performance accuracy, performance may be sensitive to the alternate training sets available and used especially when the size of the training set is not sufficient. With the use of direct matching approaches, computational overload can be reduced and the designed and developed FR system can provide an attractive alternative to training-based ones. Our intra-spectral FR approach for the homogeneous framework follows such a concept.

2.6.4 Similarity Metrics

Four measures are frequently used in order to compute the similarity between features, i.e., chi-squared distance (Chi) [63], distance transform (DT) [63], Euclidean distance (L2) and City-block distance (L1). The chi-squared distance is defined as follows:

$$\chi^2(n,m) = \frac{1}{2} \sum_1^l \frac{h_n(k) - h_m(k)}{h_n(k) + h_m(k)} \tag{2.9}$$

where h_n and h_m are the two feature vectors, l is the length of the feature vector, and n and m are two sample vectors which have been extracted from an image of the gallery and probe sets respectively. The distance or similarity metric from image X to image Y) is defined as follows:

$$D(X,Y) = \sum_{Y(i,j)} w(d_X^{k_{Y(i,j)}}(i,j)), \tag{2.10}$$

where $k_{Y(i,j)}$ is the code value of pixel (i,j) of image Y, and w is a user-controlled penalty function [63]. The Euclidean distance is defined as follows:

$$d(h_n, h_m) = \sqrt{\sum_{k=1}^l (h_n(k) - h_m(k))}, \tag{2.11}$$

where h_n and h_m are the two histogram feature vectors. Finally, the distance (city-block) is defined as:

$$d(h_n, h_m) = \sum_{k=1}^l |h_n(k) - h_m(k)|. \tag{2.12}$$

2.7 Multi-Spectral Face Fusion Literature Review

Fusion of image content from sensors is an approach that has received some merit in past, notably when face content is fused for FR. In [64], Singh et al. fuse both image and score information obtained from visible and infrared images. First, registration is done using applied mutual information and then fusion is done with a proposed 2-granular SVM. Next, an efficient algorithm based on the theory of evidence is proposed to fuse match scores that are generated from multiple classifiers. The match score received from image fusion algorithm is then integrated to improve FR performance, and validated using Notre Dame and Equinox face database. In [65], Kong et al. propose a software-based registration system for fusion of face images, where the subject may be wearing eyeglasses in thermal images. The motivation of their work was the lack of information around eyes in thermal images due to glass blocking localized areas of thermal energy. Therefore, eyeglass regions were detected using a fitting method that is ellipse-based, and those regions were replaced by generalized eye template patterns to preserve detail for FR. Two visible-thermal face image databases (the NIST/Equinox and the UTK-IRIS databases) are used for evaluation. Gyaourova et al. [66] also examine limitations and variations in IR images from eyeglass interference. Gyaourova et al. discuss sensitivity as a result of occlusion from sunglasses, and discover recognition performance degrades when eyeglasses are present in probe images but not gallery image and vice versa. To remedy this, [66] propose fusing both modalities through a pixel-based scheme that operates in the wavelet domain, and utilizes genetic algorithms (GAs) to decide how to combine IR with visible content.

2.8   Face Matching Approaches for Passive Infrared Band

Visible images are sometimes fused with images captured at another band (such as the IR band) in order to boost recognition accuracy [67]. However, to the best of our knowledge, there are no face-based matching algorithms designed to work efficiently when operating across different bands, visible and passive IR bands. In [68] the authors reported that FR in the LWIR band achieves a rank-1 accuracy of 97.3% when using local binary patterns (LBP), while no cropping or geometrical normalization step is required. In *Socolinsky et al.* [69], the authors used two standard FR algorithms to show that, under variable illumination conditions, the usage of LWIR face images yield a higher recognition performance than visible ones. However, the drawback of the approaches is that LWIR and visible images were divided into multiple training and testing sets, resulting in an increase of the FR system design time. In addition, [69] was performed using co-registered images that were captured simultaneously by both visible and LWIR cameras - this is not usually possible in operational environments. In other IR-based FR approaches, such as *Trujillo's et al.* and [70], the authors proposed an unsupervised local and global feature extraction paradigm to classify different facial expressions. In *Chen et al.* and [71] the authors combined visible and thermal-based images and compared them using Principle Component Analysis. However, neither *Trujillo* [70] nor *Chen's* work [71] focused on the MWIR band. Over the last couple of years, MWIR and LWIR sensing technology has grown in terms of resolution, pixel size and advances in methodological approaches.

Recent advances in appearance based IR FR has closely reflected research in visible spectrum based recognition. Progress in comparison with some aforementioned earlier works is rooted mainly in the use of more sophisticated statistical techniques. For example, *Elguebaly et al.* [72] recently described a method based on a generalized Gaussian mixture model, where parameters are learned from a training image set using a Bayesian approach. Although

substantially more complex, this approach did not demonstrate a statistically significant improvement in recognition of database used, achieving rank-1 rate of approximately 95%. The wavelet transform has been studied extensively as a means of representing a wide range of 1D and 2D signals, including face appearance in the visible spectrum, because of its ability to capture both frequency and spatial information. *Srivastava et al.* [73] were the first to investigate use of wavelet transform based on a bank of Gabor filters for extracting robust features from face appearance images in the IR spectrum. The marginal density functions of the filtered features are then modelled using *Bessel K forms* which are matched using the simple L_2-norm. *Srivastava et al.* reported a remarkable fit between the observed and the estimated marginals across a large set of filtered images. The curvelet transform is an extension of the wavelet transform, in which the degree of orientational localization is dependent on the scale of the curvelet [74]. The curvelet transform facilitates a sparser representation than wavelet transforms with effective spatial and directional localization of edge-like structures in natural images. *Xie et al.* [75] described the first IR based FR system which utilizes the curvelet transform for feature extraction. The method utilized a simple nearest neighbor classifier which demonstrated a slight advantage (of approximately 1-2 %) over simple linear discriminant based approaches, but with a significant improvement in computational and storage demands. *Wu et al.* [76] and *Xie et al.* [77] proposed to exploit temperature differential between vascular and non-vascular tissues, extracting invariant features in IR imagery. *Wu et al.* formulated the model governing blood perfusion, which is based on differential equations by using a series of assumptions on relative temperatures of body's deep and superficial tissues, and the ambient temperature. The model is then used to compute a "blood perfusion image" from the original segmented thermogram of a face. Finally, blood perfusion images are matched using a standard linear discriminant and an network of radial basis functions. These works are summarized in Table 2.2.

Table 2.2: Literature Review Matrix for Passive Infrared Band

Author(s)	Spectrum(s)	Approach/Features	Pros	Cons
H. Mendez, C. Martin, J. Kittler, Y. Plascencia, and E. Reyes[68]	LWIR	Local Binary Patterns (LBP)	rank-1 accuracy of 97.3%	No cropping or geometric normalization
D. Socolinsky, L. Wolff, J. Neuhesel, and C. Eveland[69]	Visible & LWIR	Eigen-faces and ARENA (appearance based/l-nearest neighbor)	Co-registered visible and LWIR images captured simultaneously	LWIR and visible divided into multiple training and testing sets
L. Trujillo, G. Olague, R. Hammond, and B. Hernandez[70]	LWIR	Local and global feature extraction paradigm	Unsupervised (no learning necessary)	Facial expressions were classified, no FR
X. Chen, P. Flynn, and K. Bowyer [71]	Visible & LWIR	Principle Component Analysis (PCA)	Visible and thermal-based images fused	FR accuracy is lower in time-lapse (different session scenarios)
P. Buddharaju, I. Pavlidis, P. Tsiamyrtzis, and M. Bazakos[78]	MWIR	Physiological features	Robust and invariant to pose	No cropping or geometric normalization
T. Elguebaly, and N. Bouguila[72]	Visible & LWIR	Gaussian mixture model	Unsupervised algorithm with rank-1 accuracy of 95%	Substantially more complex
A. Srivastana, and X. Liu [73]	Visible & LWIR	Wavelet transform based on Gabor filters	Effective with low-resolution images	Bessel parameters significantly reduce representation
Z. Xie, S. Wu, G. Liu and Z. Fang[75]	LWIR	Curvelet transform	Improved computational and storage demands	Better performance with radiant energy instead of thermal
S. Wu, W. Song, L.J. Jiang, S. Xie, F. Pan, W.Y. Yau, and S. Ranganath[76]	LWIR	Blood perfusion	Less sensitive to ambient temperature	Time and storage efficiency
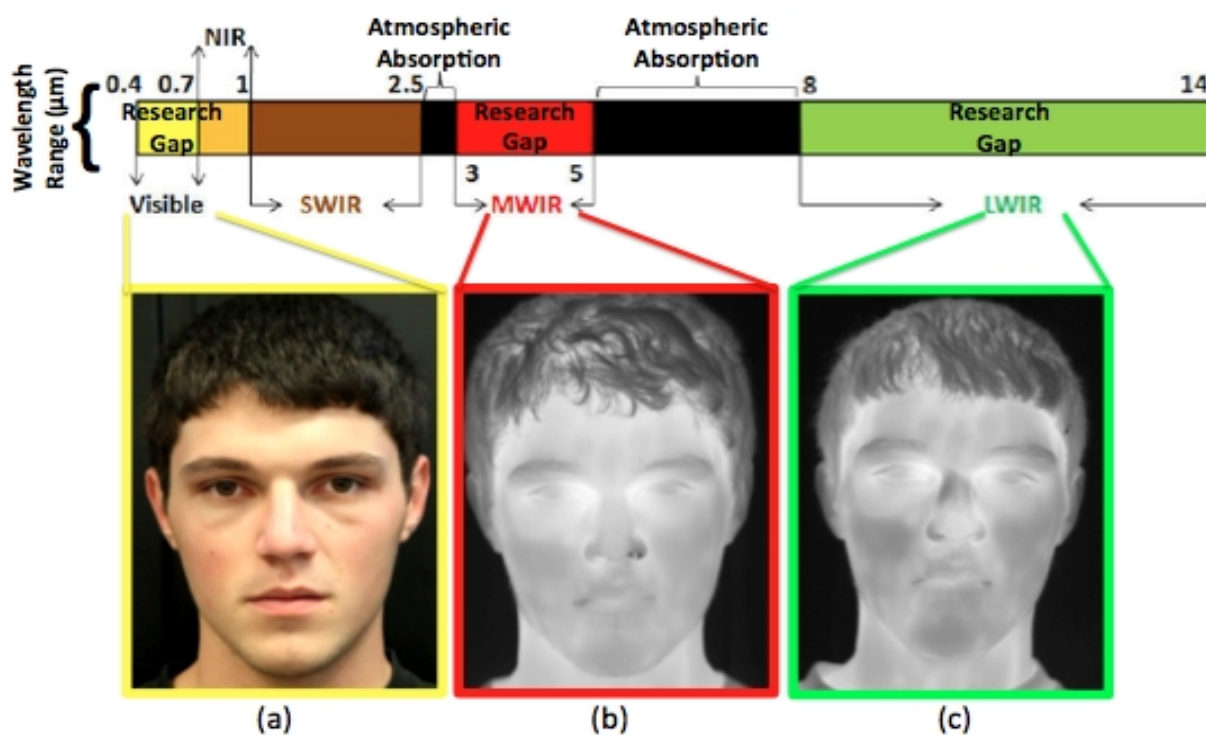
Figure 2.5: Sample face images of a randomly selected subject are also illustrated that correspond to the three individual bands of interest, i.e. (a) Visible (yellow), (b) MWIR (red), (c) LWIR (green). Different bands of the electromagnetic spectrum (visible and passive IR) are highlighted to illustrate an existing research gap in terms of having a unified FR algorithm that can operate in all discussed bands, utilizing facial characteristics that are extracted and exploited by similar feature extraction and face-based matching techniques.

Chapter 3

Homogeneous Face Recognition

3.1 Introduction

The goals of this chapter include 1) the proposal of a new semi-automated pre-processing FR approach that is applied on full-frontal face images acquired using visible, MWIR and LWIR camera, and 2) the design of experiments to quantitatively illustrate the benefits of global vs. local based matchers, and finally, the performance of such matchers when fused at the score level. In regards to the semi-automated pre-processing approach, human eyes and eye centers are detected. Then, eye coordinates are used to geometrically normalize faces. Also, the inter-ocular distance is fixed by setting the dimensions of the input face images at a specific spatial resolution and the eye centers of these images at predetermined (x, y) locations. Prior to the feature extraction stage, the parameters of the image diffusion and face segmentation algorithms are enhanced. Anisotropic image diffusion is also used to smooth each face image, prior to applying top-hat segmentation.

In our experiments, we use the segmented features to demonstrate that our face recognition system performance improves when fusing global and local-based matchers. The global matcher uses the feature segmented image for matching, i.e. this matcher analyzes the mor-
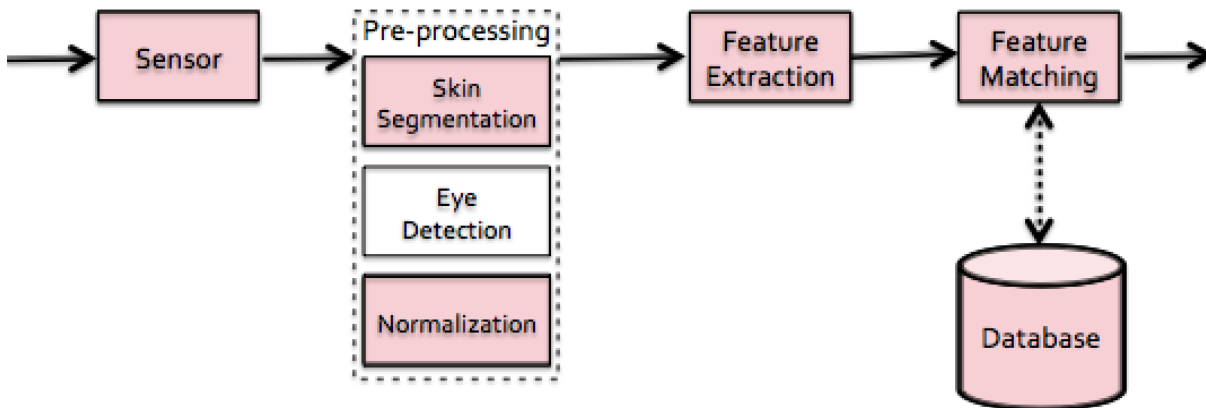
Figure 3.1: Main components typically found in a FR system. The shaded pink objects illustrate areas in which contributions were made. There aren't any contributions that were made to the area of eye detection in this work. All eye detection described in this work is manually done.

phology of the face through the convolution of face images, producing a match ratio [79]. Although the global matching algorithm is based on overlap of neighborhood pixels, the accuracy of the algorithm increases when unique segments are matched. The local matcher requires fiducial points that we obtain using the extracted feature segment, and matches all points to one another (point to point). These fiducial points are minutiae (level 2 features) extracted from physiologically-based (when using subcutaneous facial characteristics) and geometrically-based face features (e.g. eye edges and eyelashes), which are unique for each individual (see Fig. 3.2). The use of minutia points helps facilitate a metric for measuring similarity using our local matcher and helps improve accuracy through precise selection of features. The efficiency of our local matcher is tested using automated methods for extraction of correspondence points, such as *Scale-Invariant Feature Transform* (SIFT) [80] and *Speeded Up Robust Feature* (SURF) [81]. In the matching and decision making step for both matchers, after our feature segmentation scheme is applied to each image, input images (probes) are matched with a stored template (gallery). Both matchers perform well individually, but optimal performance is achieved fusion. At the matcher level, we used score-sum
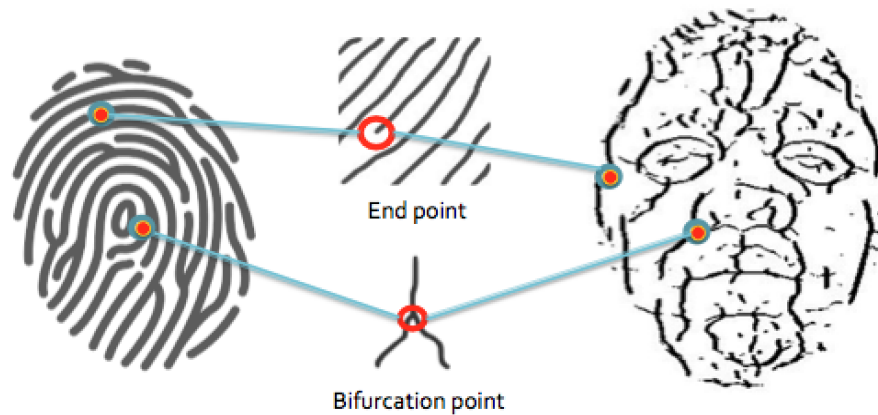
Figure 3.2: Minutiae points (Level 2 features) commonly used in fingerprint recognition are utilized by our local matcher for face recognition.

with min-max normalization for fusion. Fig. 3.1 illustrates main components typically found in a FR system and the contributions to those respective components made in this work.

3.2 Intra-Spectral Face Image Database

In this section, we will describe the hardware used for: 1) the acquisition of the high-quality face images, and 2) MWIR and LWIR face images. We will also describe the live subject-capture setup used during the data collection process and the three spectral face image databases utilized in this paper. Finally, the IR sensitivity and calibration necessary for the passive IR sensors is discussed.

3.2.1 Subject-Capture Setup

Three different spectral cameras were employed for data collection in order to assemble our multi-spectral face image database. The live face capture configuration we used had a standoff distance of 5 feet ($\sim$ 1.5m) between the subject and our cameras. The database was collected indoors over multiple sessions spanning over a time period of 3 months. At

the first session, the subjects were briefed in regards to the data collection process, after which they signed a consent document. In total, 138 subjects were used in the construction of the databases. Each database consists of only full frontal face images with a neutral facial expression for every subject. Of the 15 samples that exist for each subject, only 4 samples were used for our matching experiments. The first 2 samples were the gallery images, while the remaining 2 samples were the probe images.

Experimental Sensors

A Canon EOS 5D Mark III digital camera was used for the acquisition of high-quality visible face images (1920 × 1080). A FLIR SC8000 MWIR camera was used for the acquisition of high-quality MWIR face images (1920 × 1080). A FLIR SC600 LWIR camera was used for the acquisition of LWIR face images (640 × 480). Although the passive infrared sensors utilized here provide a 14-bit dynamic range, images that were extracted from our video sensors were saved as 8-bit images.

- **Visible Dataset:** In this work, the Canon SLR is used to obtain standard RGB, ultra-high resolution fontal pose face images in the **visible** spectrum (2.5 (a)). This digital camera has a 21.1-megapixel full-frame CMOS sensor with DIGIC 4 Image Processor and a vast ISO Range of 100-6400. It also has *Auto Lighting Optimizer* and *Peripheral Illumination Correction* that enhances its capability. The visible images are extracted from the movie files in JPEG format. The JPEG images extracted had an approximate file size of 100 KB per image. This dataset consists of 408 images (204 for probe and 204 for gallery) with four images per subject (102 subjects).

- **MWIR Dataset:** A FLIR SC8000 camera served as the imager for obtaining **MWIR** images (see 2.5 (b)). The camera features an Indium Antimonite (InSB) Focal Plane

Array (FPA) achieving mega-pixel image resolution in a single MWIR image. The spectral range of the camera is 3-5 μm, and it has a 14-bit dynamic range. The MWIR images were extracted from the thermal sequential files in BMP format. The BMP images extracted had an approximate file size of 619 KB per image. This dataset consists of 408 images (204 for probe and 204 for gallery) with four images per subject (102 subjects).

- **LWIR Dataset:** A FLIR SC600 camera served as the imager for obtaining **LWIR** images (see 2.5 (c)). The camera features Indium Antimonite (InSB) Focal Plane Array (FPA) achieving mega-pixel image resolution in a single LWIR image. The spectral range of the camera is 7-14 μm, and provides imaging performance up to 14-bits. The LWIR images were extracted from the thermal sequential files in BMP format. The BMP images extracted had an approximate file size of 922 KB per image. This dataset consists of 404 images (202 for probe and 202 for gallery) with four images per subject (101 subjects).

3.2.2 Passive Infrared Sensitivity and Calibration

The MWIR camera used in this work has a Noise Equivalent Temperature Difference (NETD) of less than 25mK. NETD indicates the sensitivity of a detector of thermal radiation. FLIR software was used: (a) to perform regular calibration, before acquiring each set of thermal face images. We used a black body and a two-point non-uniform correction process that performs both gain and offset normalization of pixel-to-pixel non-uniformity; (b) to remove noise (e.g., dead pixels) from face images and control the temperature scale limits (e.g., setting the temperature range from 28°to 40°Celsius that is the typical range of human body temperature) during data collection. The LWIR camera used in this work has a NETD of less

than 50mK. The FLIR software automatically calibrates the sensor regularly and removes noise from the face images.

3.3   Pre-processing

It is imperative the input face images are geometrically normalized using our pre-processing pipeline prior to feature-extraction. Our pre-processing pipeline utilizes skin segmentation and normalization (e.g. interocular and geometric) in order to eliminate possible artificial information from being introduced into our system. The methodology used to perform pre-processing for each spectrum can be seen in Fig. 3.3.

3.3.1   Skin Segmentation

The first step of the face image matching pipeline is segmenting the face from the subject's body, clothing and background noise to eliminate any artificial features.

- *Visible:* Skin segmentation is carried out by first converting the color (RGB) image to a hue saturated value (HSV) image. However, we are only interested in the hue of the HSV image, along with chroma components C_B and C_R of the original color image. We traverse the image and if the chroma components and hue value both fall within acceptable thresholds, then the pixel is set white, representing skin. Due to variations in illumination among visible face images, there may be shadowing effects on the skin. Therefore, a range of thresholds is considered in order to avoid over/under segmentation of skin in the visible spectrum. The optimal segmented face image is visually chosen and used for experimentation. Before using the binary mask created to mask the original image, the mask should be filled for holes. It is important to note that the more you decrease the max value or range for C_R threshold, the more skin pixels

are chosen. Mustafa Ucak originally implemented the implemented skin segmentation method described above, however it was adapted for this work.

- *Passive Infrared:* A blob detection algorithm is used to detect the bright portions of the thermal face using a predetermined threshold value. The threshold value for the blobs is the numerically larger pixel of the two intensity values that produce the greatest intensity difference. Using this threshold, a binary image is created, where pixel intensities above the threshold are assigned a value of one and pixel intensities below the threshold receive a value of zero. All blobs, which represent human skin, are analyzed and the largest one is selected. The human face is the largest uncovered area of skin in our images due to clothing, so the largest blob represents the subject's face and neck. Background pixels (holes) inside the blob are then removed so that the mask is a solid. The face blob is then used as a mask against the original infrared image, setting the background and clothing black.

3.3.2   Normalization

There are two types of normalization done, the first of which is inter-ocular normalization, followed by geometric normalization. This is valuable because shape and size of the head varies amongst subjects. A canonized image is especially important because both of our matchers depend on location and alignment of extracted features.

- *Inter-ocular Normalization:* It standardizes all face images so head sizes are relatively similar. Once the eyes are detected on the face images, they are used to normalize all images so that the inter-ocular distance is fixed to 60 pixels. This is accomplished by resizing the image acquired after skin segmentation using a ratio computed from the desired inter-ocular distance (60 px) and the actual inter-ocular distance, i.e. the
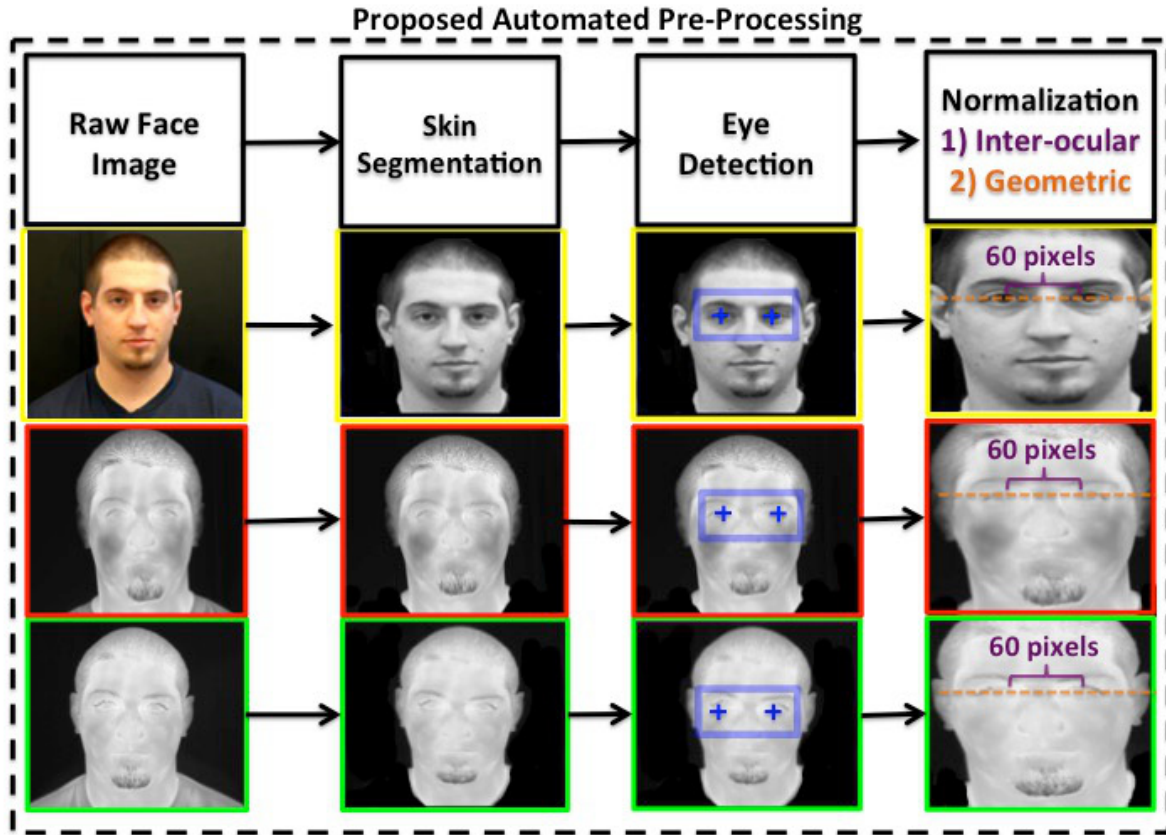
Figure 3.3: Methodology used to perform pre-processing over visible and passive infrared bands for visible spectrum (yellow), MWIR spectrum (red), and LWIR spectrum (green) images.

one computed when using the image after skin segmentation.

- **Geometric Normalization:** A geometric normalization scheme is applied to images acquired after inter-ocular normalization. The normalization scheme compensates for slight perturbations in the frontal pose, and consists of eye detection and affine transformation. The canonical faces are automatically constructed by applying a similarity transformation. Finally, all faces are canonicalized to the same dimension of 320×256.

3.4 Feature Extraction and Matching

The proposed methodology is composed of a feature extraction and matching process. The feature extraction process has two steps. First, anisotropic diffusion is used to reduce noise while preserving vital image content. Next, top hat segmentation is carried out in order to segment and extract face-based features. The features extracted include: (a) veins, (b) edges, (c) wrinkles, and (d) face perimeter outlines (see Fig. 3.7). After features are extracted, different matchers (e.g. global and local) are utilized, before finally, they are fused together at the score level in an effort to achieve increased rank-1 identification performance. The methodology used to perform feature extraction and matching for each spectrum can be seen in Fig. 3.4. The aforementioned methodological steps are described below.

3.4.1 Anisotropic Diffusion & Top Hat Segmentation

All face images are processed to remove background noise added during video acquisition, and enhance edges, the use of Perona-Malik anisotropic diffusion [82]. This is important because noise is reduced without the removal of significant image content, such as edges and lines. The mathematical representation for this process is described as follows:

$$\frac{\partial I(\bar{x}, t)}{\partial} = \nabla(c(\bar{x}, t)\nabla I(\bar{x}, t)) \tag{3.1}$$

where $I_{N,t} = I_t(x, y + 1) - I_t(x, y)$. The diffusion operator inherently behaves differently, depending on which spectrum we are operating in. For images in the visible spectrum, diffusion is used in edge detection between lines of dissimilar contrast. The face-based features (wrinkles, veins, edges, and perimeters) are segmented the use of image morphology. For the passive infrared band, heat diffusion generally produces weak sigmoid edges in the thermal band during heat conduction, which in turn creates smooth temperature gradients
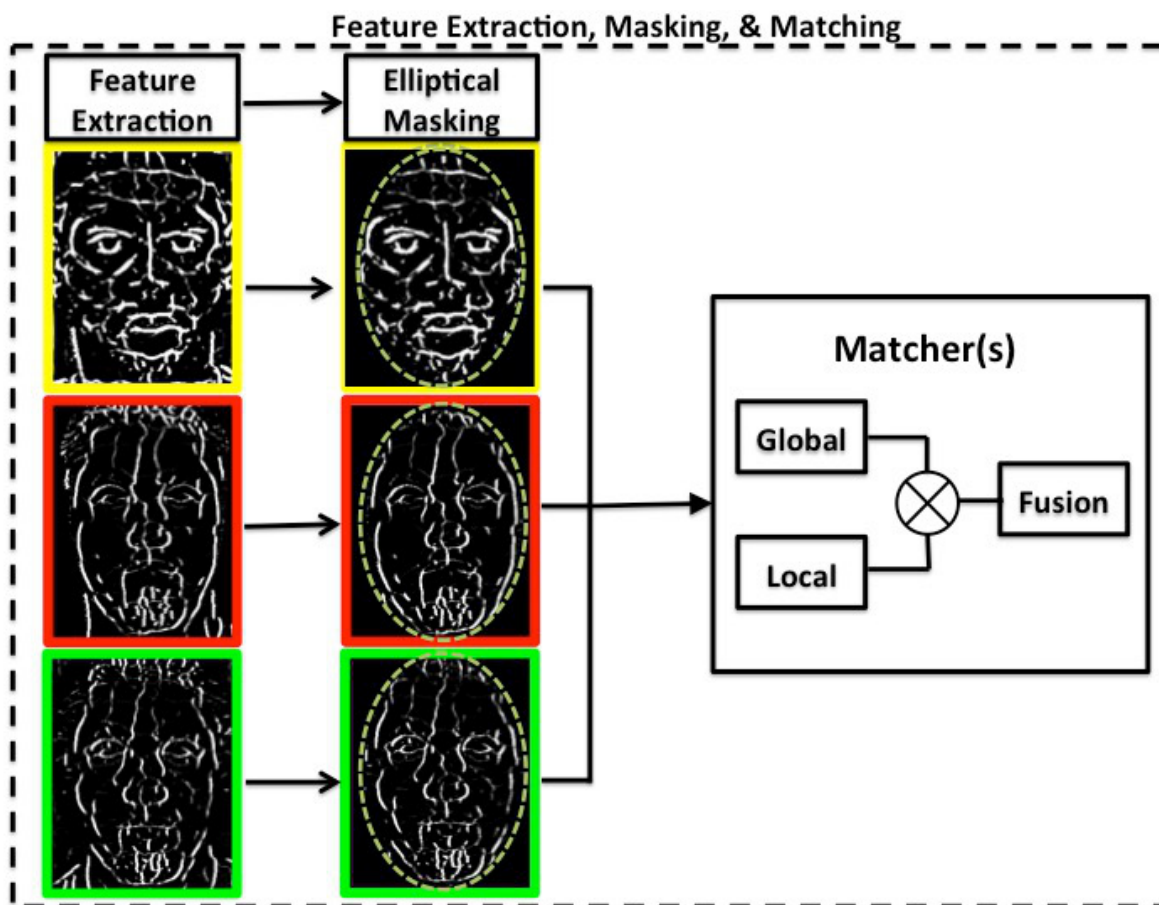
Figure 3.4: Methodology used to perform feature extraction over visible and passive infrared bands for visible spectrum (yellow), MWIR spectrum (red), and LWIR spectrum (green) images.

at the intersecting boundary of multiple objects with dissimilar temperatures in contact. In order to segment these features, morphological top-hat filtering is employed.

$$I_{open} = (I \ominus S) \oplus S, \tag{3.2}$$

$$I_{top} = I - I_{open} \tag{3.3}$$

I, I_{open}, and I_{top} are the original, opened, and white top hat segmented images, respectively. S is the structuring element, and $\ominus$ and $\oplus$ are the morphological operations for erosion and dilation respectively. Parameters for Perona-Malik anisotropic diffusion were empirically enhanced so that pixel intensities were smoothened without blurring edges in the face image (see Fig. 3.5). Parameters for top hat segmentation were chosen so that features could be segmented with the least amount of noise present (see Fig. 3.6). Empirical enhancement of our feature extraction process ensure the resultant images (see Fig. 3.7) do not contain noise, i.e. outlier edges that do not represent clear face-based physiological and geometrical features. Pixel normalization was the only pre-processing done on the images during this experiment, unlike the original work, which dilated, skeletonized, and bridged MWIR images [78].

3.4.2   Elliptical Masking

It is imperative that the elliptical mask is applied after the feature extraction stage to ensure that spurious, artificial feature points are not created by the mask's presence during top hat segmentation. In practice, image masking ensures no clothing is present during feature detection. When the elliptical mask segments the face, the part of the original image that is masked is set to a black background. Note also that the dimensions of the ellipse used for
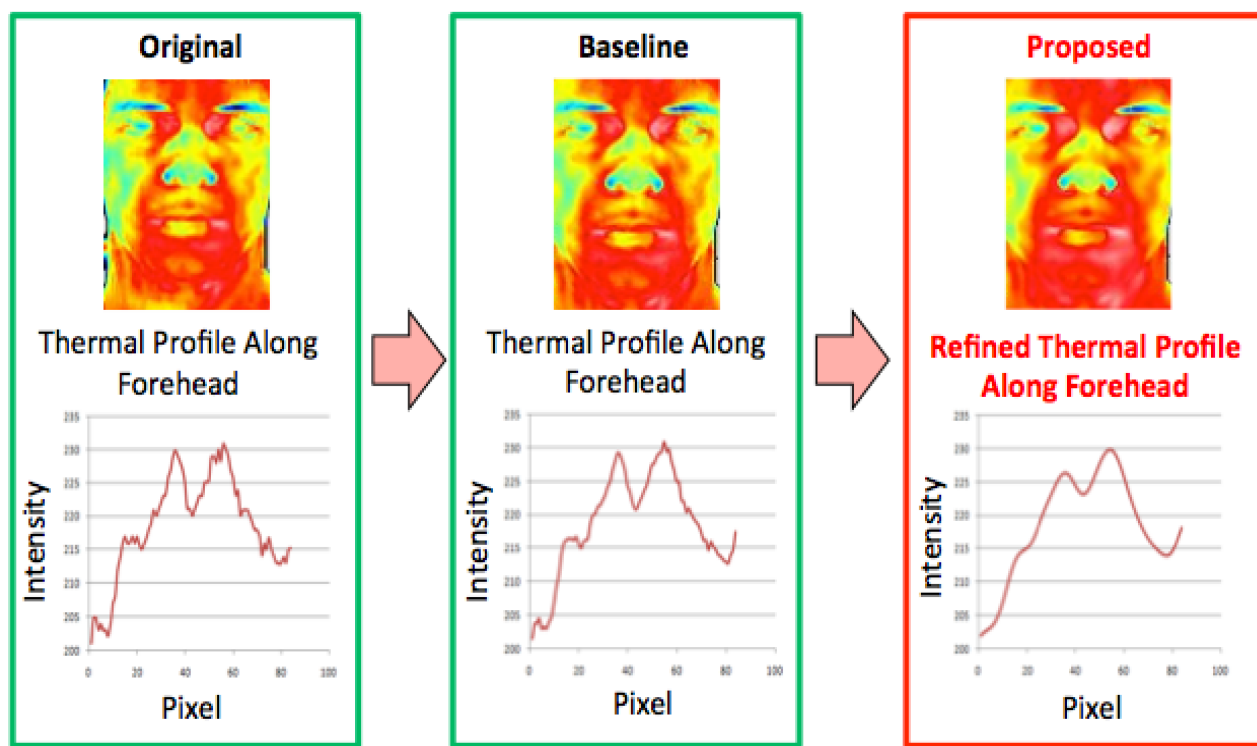
Figure 3.5: Thermal profile across subject forehead for original, baseline, and proposed methodology.
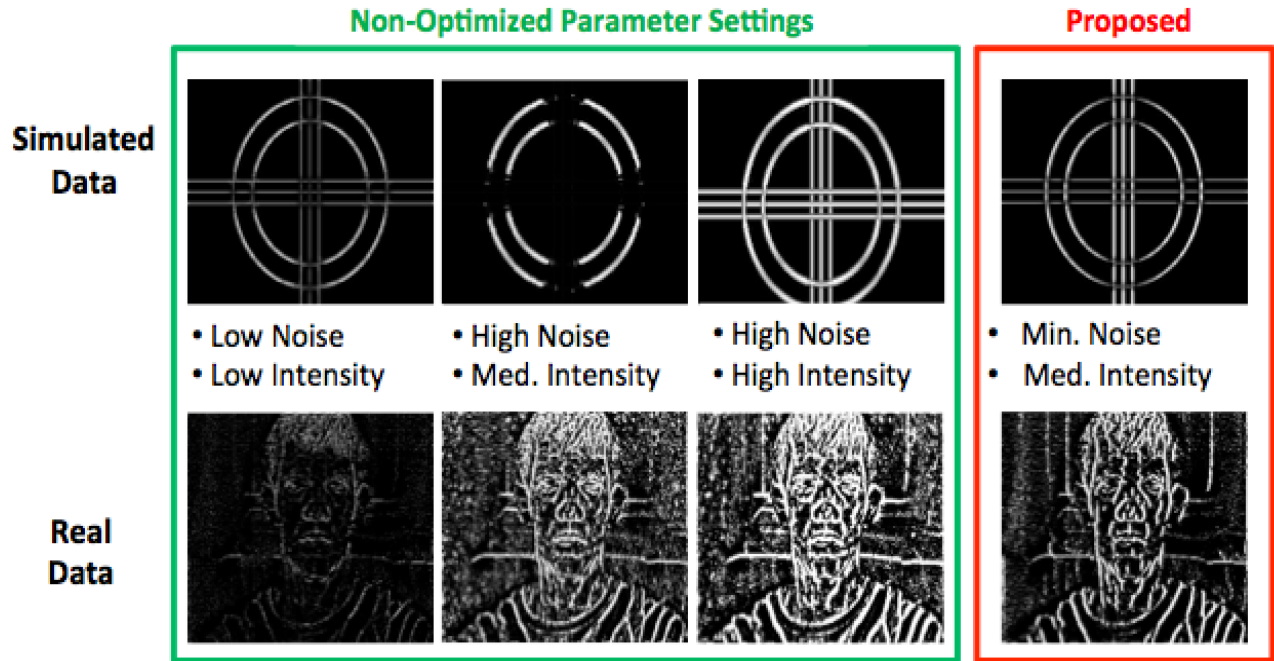
Figure 3.6: Enhancement of segmented data utilizing feature extraction. Simulated and real data was utilized to enhance parameter settings.

all images were fixed.

3.4.3 Matching Schemes

Many FR matching methods can be categorized as being global or local, depending on whether features are extracted from the entire face or from a set of local regions. Global features are usually represented by vectors fixed in length, which are compared during the matching process in a time efficient manner. On the contrary, local feature-based approaches first detect a set of fiducial points, using the surrounding pixel values. The number of matched fiducial points between an input and gallery template is used to calculate the match score. Since the number of fiducial points may vary depending on the subject, two sets of points from two different subjects cannot be compared directly. As a result, the matching scheme has to compare each fiducial point from one template against all the other

fiducial points in another template, increasing time for matching. Global features are more susceptible to variations in the face when all pixel values are encoded into a feature vector, especially with respect to geometric transformations.

Global Matcher: Segmentation Matching

It is usually a difficult task to extract on face information using only simple techniques. Our feature segmentation step extracts lines, edges, and wrinkles, which are unique to each subject, such as in the case of contours. In two face images of the same subject, similar features may still be found in contours. Especially in the case of subject's which may be identical twins. However, there may be remarkable differences in not only the shape of these contours but in the size as well, across subjects. Hence, identification may be carried out using matching of the segmented features. The segmented face features are compared using template matching.

   To find the maximum similarity between a gallery and probe image, the two feature segmented images are convoluted. The probe image is slid pixel by pixel across the gallery image in a top to bottom and left to right fashion. If $f(i,j)$ and $g(k,l)$ are the pixels at position (i,j) and (k,l) of probe and gallery feature segments respectively, then α, β measure the horizontal and vertical displacement between $f(i,j)$ and $g(k,l)$ during the traversal process. To account for the small differences that exist in the segmented features from different samples of the same subject, a 5×5 window around a white pixel is used for template matching. If for a white pixel in the probe segmented image, there is another white pixel in the 5×5 neighborhood of the corresponding position on the gallery segmented image, then the pixels are said to be matched. If $\hat{\alpha}, \hat{\beta}$ are the horizontal and vertical displacement respectively, which give the best matching result, then the maximum similarity $h_{i,j}(\hat{\alpha}, \hat{\beta})$

defined by (3.4) between corresponding feature segments can be obtained [79].

$$h_{i,j}(\hat{\alpha}, \hat{\beta}) \stackrel{\triangle}{=} \max_{\alpha,\beta} \sum_{i,j} h_{i,j}(\alpha, \beta) \tag{3.4}$$

$$h_{i,j}(\alpha, \beta) = \phi[\sum_{x=-2}^{2} \sum_{y=-2}^{2} (f_{i,j} \cdot g_{k+x,l+y})] \tag{3.5}$$

$k = i \pm \alpha, l = j \pm \beta,$

$\alpha = 0, 1, 2, ...320, \beta = 0, 1, 2, ...256$

$$\phi[x] = \begin{cases} 1, & \text{for } x \geq 1 \\ 0, & \text{for } x = 0 \end{cases} \tag{3.6}$$

For images of the same subject, after template matching, long overlapping segments (in the sense of (3.5)) are obtained. On the other hand, for different subjects, even if a high match score is obtained from (3.4), the existence of abundant segments overlapped by chance is expected. Therefore, if these short segments can be eliminated effectively, a stable performance of the discrimination method can be achieved. The final matching score $H(f, g)$ can be calculated using (3.7).

$$H_{f,g} = \frac{2}{F + G} \cdot \sum_{i,j} h_{i,j}(\hat{\alpha}, \hat{\beta}) \tag{3.7}$$

F and G denote the number of pixels in the feature segmented lines of the probe and gallery images respectively. The fragment removal threshold θ_i was optimized experimentally, and held constant for each image. Prior to matching, all images are filtered by removing pixel values below θ_i.
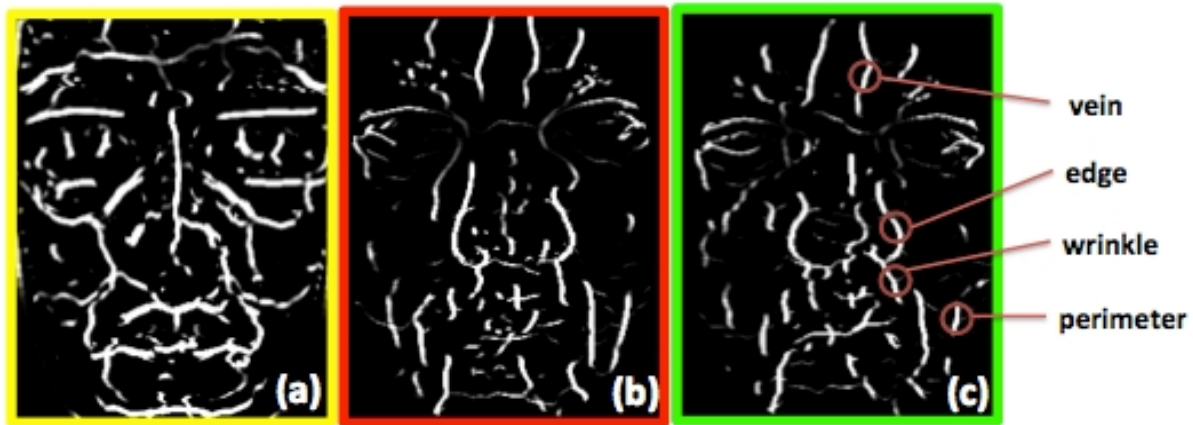
Figure 3.7: Sample normalized and binarized segmented feature consisting of veins, edges, wrinkles and parts of the face perimeter on: (a)Visible face images (yellow), (b) MWIR face (red), and (c) LWIR face images (green).

Local Matcher: Fiducial Point Matching

Fiduciary points are used as points of measure in many face recognition applications and play an important role in our localized matching algorithm. The use of fiducial points is appropriate due to the permanent and unique nature of the extracted features related to face-based subcutaneous information. Note also that although only face images with neutral expressions and poses were used in our experiments, fiducial points can be extracted and used for matching regardless of facial pose or expression.

- *Fingerprint-based Minutiae Detector:* The proposed method, a fingerprint based minutia point recognition system is used to detect features in the face [83]. Beforehand, the pixel intensities of the images are normalized so that they are binary. For our normalization, the minimum pixel intensity of the image is set as the threshold for visible images and the mean pixel intensity of the image is set as the threshold for passive infrared images. While traversing the image, if the current pixel intensity is smaller than the threshold, it receives a binary value of zero. If the current pixel intensity is larger
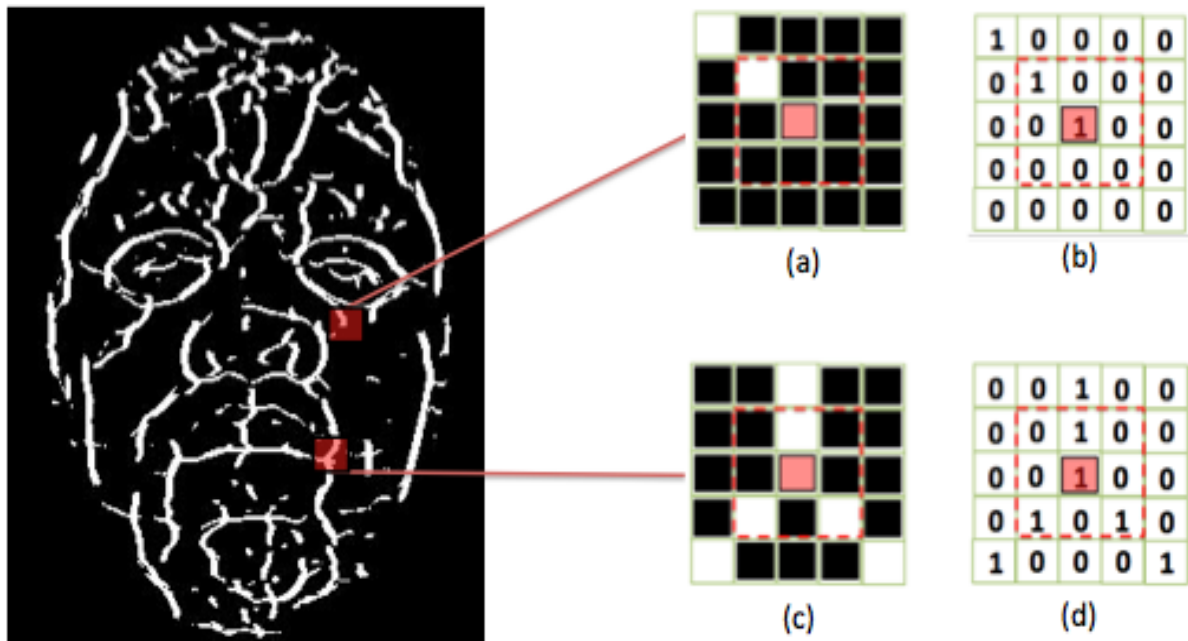
Figure 3.8: (i) Sample feature segmented subject face with (a) block representation of end point pixels, (b) binary representation of endpoint, (c) block representation of branch minutia point pixels and (d) binary representation of branch minutia point. Red box denotes marked minutia point.

than the threshold, it receives a binary value of one. During the traversal process, if the 3×3 neighborhood around a center white pixel had exactly three white pixels, it was labeled a branch point. If the center white pixel with a 3×3 neighborhood around it contained only 1 white pixel neighbor, then it was labeled an end point (see Fig. 3.8). Depending on the normalization technique, large clusters of both branch points and end points may be found.

Occasionally, artifacts may still be introduced into the segmented image, which later lead to spurious minutia. These false minutiae will significantly impact the performance of our fiducial point matching algorithm if they are classified as genuine minutia. Therefore, some mechanisms of removing false minutia are important in efficiency and effectiveness of extracted points. This is consistent mostly of the removal of either

branch points or end points that are close to each other, using a threshold distance. The threshold distance is computed as the average distance between two parallel neighboring segments in an image.

- *SIFT:* The detector extracts a number of frames from an image in a manner that is consistent with certain variations that may affect image quality such as external illumination. The SIFT feature detector was required in our work because we are only interested in x-coordinate and y-coordinate points (distinct feature key points). Such points are defined as resultant maxima and minima of Difference of Gaussians (DoG) applied to smoothed and re-sampled images [80]. In implementation, our SIFT detector is controlled primarily by two parameters, the peak threshold and the non-edge threshold. The peak threshold filters peaks of the scale-space that are too little in absolute value. As the peak threshold increases, fewer features are obtained. The edge threshold removes peaks of the DoG scale-space whose curvature is miniscule. As the edge threshold increases, more features are obtained. After applying this detector, for each key feature location, the x-axis and y-axis coordinates are extracted.

- *SURF:* The SURF method is very similar to SIFT. However, SURF is only a scale and rotation-invariant interest point detector and descriptor, offering a compromise between feature complexity and robustness for commonly occurring deformations [81]. The use of integral images allows for fast implementation of box type convolution filters. SURF aims to find salient regions in near constant time its use of integral images and box filters. When features are detected using the SURF algorithm, there is no need to iteratively apply the same filter to the output of a previously filtered level's images. After obtaining x-coordinate and y-coordinate, we move on to applying our local (e.g.
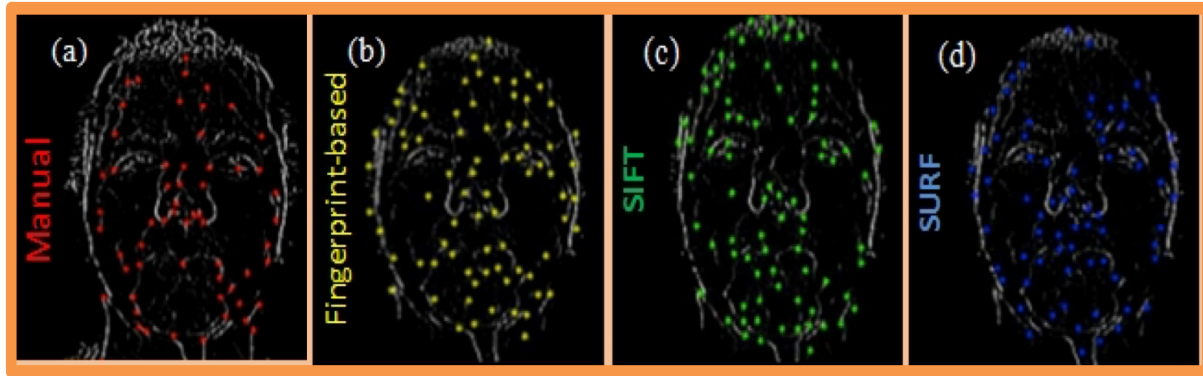
Figure 3.9: (a) points detected manually by forensic examiner, (b) points detected by fingerprint-based minutia detector, (c) points detected using SIFT algorithm (d) point detected using SURF algorithm.

fiducial point) matching algorithm using the detected points generated by SURF.

The fiducial points extracted using a forensic examiner (manual), fingerprint-based minutiae detector, SIFT, and SURF detectors represent points that overlay segmented features. A visual comparison of these points can be seen in Fig. 3.9.

Post-Processing: False Minutia Removal

For our fingerprint-based minutia detector, the preprocessing stage does not completely prepare the face image for matching. For example, false segment breaks due to insufficient amount of light or introduced artificats are not totally eliminated. These false minutia may affect the accuracy of the local matcher if they are simply regarded as genuine minutia. Therefore, some mechanisms for removing false minutia are essential to keep the local matching algorithm effective. My procedures in removing false minutia on the face image are:

1. If the distance between one bifurcation point and one endpoints is less than D and the two points are located in the same segment, remove both points. Where D is the average inter-segment width representing the average distance between two parallel

neighboring segments.

2. If the distance between two bifurcation points is less than D and they are in the same
   ridge, remove both bifurcation points.

3. If two endpoints are within a distance D and their directions are coincident with a
   small angle variation, and no other termination is located between the two endpoints;
   then the two endpoints are regarded as false minutia derived from a broken segment
   and are removed.

4. If two end points are located in a short ridge with length less than D, remove the two
   end points.

Fiducial Point Matching

A point alignment-based Random Sample Consensus (RANSAC) matching algorithm with
the ability of finding the correspondences between a stored set of gallery points and input set
of probe points is used. The set of input points are first aligned with the gallery points and
then a match score is computed, based on corresponding points. Let G and P be the gallery
and probe points we are trying to match. Because the correspondence points are computed
based on local information of each subject, G and P may not have the same number of
correspondence points: let m and n be the number of points for G (8) and P (9).

$$G = \{c_1, c_2, ...c_m\}, \text{ with } c_i = \{x_i, y_i\}, i = 1..m \tag{3.8}$$

$$P = \{c'_1, c'_2, ...c'_n\}, \text{ with } c'_j = \{x'_j, y'_j\}, j = 1..n \tag{3.9}$$

Two correspondence points c_i and c'_i are considered matching if close in position. This
can be written according to (3.10), using the spatial distance sd. Two minutia points from

each set are considered matching if their spatial distance (sd) is within a threshold (R_θ).

$$sd(c_i, c_j') = \sqrt{(x_i - x_j')^2 + (y_i - y_j')^2} < R_\theta \tag{3.10}$$

The distance between each possible point in the gallery set and point in the probe set is computed and sorted in increasing order. We then count the number of matches below a threshold, which is optimized for the fixed dimension of our face images. The highest counted number of correspondences; C_{best}, for each set of points is stored and used to compute the match score, M_S. We let:

$$M_S = \frac{(C_{best})^2}{tot\{G, P\}}. \tag{3.11}$$

3.4.4 Fusing Matchers at the Score Level

To enhance the recognition performance, confidence level fusion schemes can be invoked [84]. In this work, score level fusion is implemented to combine the match scores obtained from multiple matchers (global and local). The respective scores obtained individually for each matcher are fused together by the sum rule, after using min-max normalization. For a given vector S of similarity based scores of length Q, the min-max normalization of the input score vector may be computed using (3.12) for each score S_i, for all individual scores (i.e. $\sum_{i=1}^{Q}$). The operators $\min(S)$ and $\max(S)$ represent the minimum and maximum value of the score vector S respectively. Simple sum rule for both of our score vectors ($H_{f,g}$ and M_S) is carried out using (3.13), after min-max normalization (3.12) of each vector of matcher scores. For equation (3.13), m represents the total number of matchers.

$$S_{N_i} = \frac{S_i - \min(S)}{\max(S) - \min(S)}, \tag{3.12}$$

$$S_{F_i} = \frac{S_{N_i} + ... + S_{N_m}}{m}. \tag{3.13}$$

3.5 Face Recognition Systems

Both commercial and academic software were employed to perform the face recognition experiments: 1) Commercial software Identity Tools G8 provided by L1 Systems; 2) standard training-based face recognition methods provided by the CSU Face Identification Evaluation System [58], including *Principle Components Analysis* (PCA) ([49],[50], [85]), *a combined Principle Components Analysis and Linear Discriminant Analysis algorithm (PCA+LDA)* [48], the *Bayesian Intrapersonal/Extra-personal Classifier* (BIC) using either the Maximum likelihood (ML) or the Maximum *a posteriori* (MAP) hypothesis [86] and (3) *Local Binary Pattern* (LBP) method [87]. PCA rotates feature vectors from large, highly correlated sub- space to a small subspace which has no sample covariance between features. PCA is also used as part of an LDA algorithmic approach in order to reduce dimensionality of the feature vectors, and then the training data is further used to reduce dimensionality in a way that preserves distinguishing features. Both PCA and LDA algorithms may use distance metrics such as Euclidean Distance (EU), which result in the ordinary or standard distance between two feature vectors (PCA EU + LDA EU). BIC uses PCA during density estimation, in which parameters that define the Gaussian distributions, using the statistical methods properties of two subspaces: one for difference images that belong to the intrapersonal class and another for difference images that belong to the extrapersonal class. The two variants of the BIC algorithm include *maximum a posteriori* (MAP) and *maximum likelihood* (ML) classifier (BMAP + BML) [58]. The LBP operator is an efficient, nonparametric, and unifying approach to traditional divergent statistical and structural models of texture analysis. For each pixel in an image, LBP produces a binary code by thresholding its

value with the value of the center pixel. A histogram is created to count up occurrences of different binary patterns [87]. Cumulative match curves (CMC) curves comparing the proposed methodology with commercial and academic software for each respective spectrum can be seen in Figs. 3.11, 3.12 and 3.13.

3.6 Empirical Evaluation

The experimental scenarios investigated in this paper are the following: 1) validation of skin segmentation; 2) evaluation of global-based matching; 3) evaluation of local-based matching; 4) identification performance before and after fusion; 5) evaluation of matching using sub-facial face images; and 6) evaluation of system robustness when photometric normalization technique applied. Our entire multi-spectral database was used for the evaluation of eye detection, global-based matcher, local-based matcher, and fusion experiments. Due to the computational complexity of our system, 50 subjects are used for evaluation of matching using sub-facial face images, and 35 subjects are used to evaluate the CLAHE normalization of our proposed system. For evaluation of global-based matcher, local-based matcher, fusion, and matching of sub-facial face images, erroneously segmented skin is manually corrected, as well as eye locations that are incorrectly detected.

3.6.1 Skin Segmentation Validation

The skin segmentation algorithms presented in this work have yet to be validated in literature, therefore it is important the performance of the algorithms is validated. As with most segmentation methods that utilize thresholds, over segmentation is a major concern because the number of potential features that may be used for matching are reduced. Skin segmentation is more of a challenge in the visible spectrum due to variations in illumination. Fig. 3.10 gives examples of erroneous skin segmentation and correct skin segmentation results for
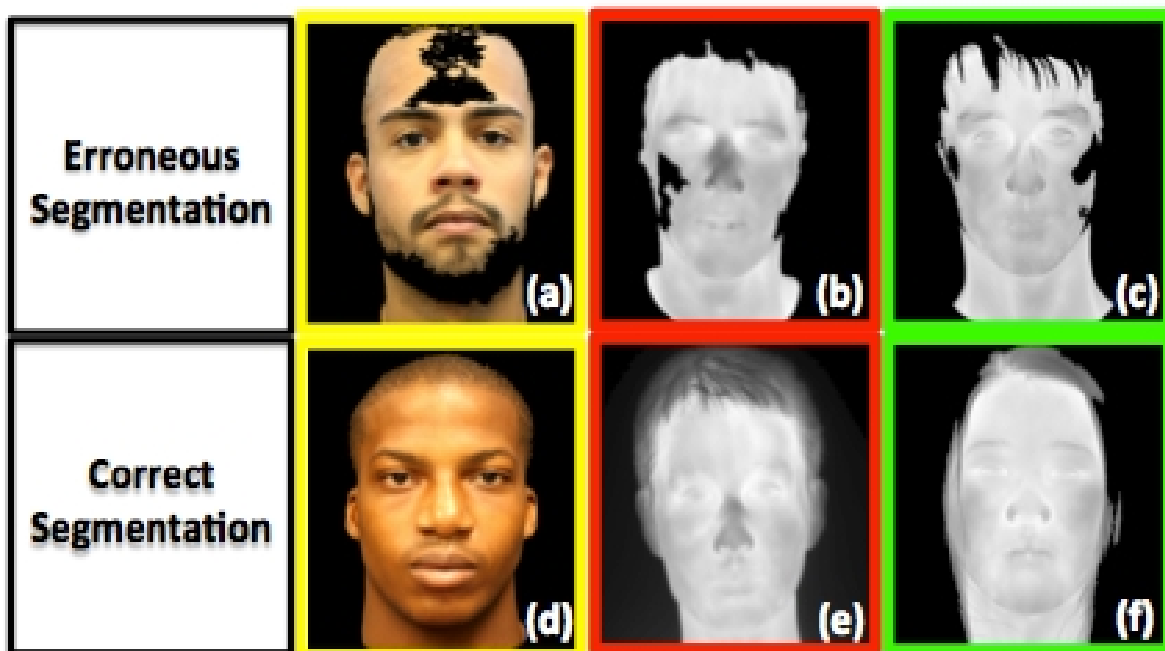
Figure 3.10: Validation of proposed skin segmentation algorithm across the visible and passive infrared spectrum: (a) Erroneously segmented visible face image (b) Erroneously segmented MWIR face image, and (c) Erroneously segmented LWIR face image, (d) Correctly segmented visible face image, (e) Correctly segmented MWIR face image, and (f) Correctly segmented LWIR face image.

all three spectrums.

3.6.2 FR Experiments using Global Matcher

In this experiment, the entire multi-spectral database is utilized to analyze the performance of the proposed feature segmentation algorithm in combination with our global-feature matcher (assuming no errors in pre-processing). We report the experimental results as CMC curves at the score level, for each spectrum separately. For 102 subjects in the visible spectrum, we achieved a rank-1 accuracy of 98.95%. For 102 subjects in the MWIR spectrum, we received a rank-1 accuracy of 94.21%. For 101 subjects in the LWIR spectrum, we achieved a rank-1 accuracy of 94.21%.

3.6.3 FR Experiments using Local Matcher

In this experiment, we utilize our entire multi-spectral database to investigate the performance of our feature segmentation algorithm in combination with our local-feature matcher (assuming no errors in pre-processing). We report the experimental results as CMC curves at the score level, for each spectrum individually. For 102 subjects in the visible spectrum, we achieved a rank-1 accuracy of 98.95%. For 102 subjects in the MWIR spectrum, we received a rank-1 accuracy of 99.47%. For 101 subjects in the LWIR spectrum, we achieved a rank-1 accuracy of 93.10%.

3.6.4 FR Experiments using Fused Matchers

For our entire visible, MWIR, and LWIR datasets, our two matchers achieve a rank-1 accuracy of 99.47%, 100%, and 99.43% respectively, when fused at the score level (assuming no errors in pre-processing). For our entire visible, MWIR and LWIR data, we examine verification results by plotting ROC curves after fusion of both matchers. For this work,
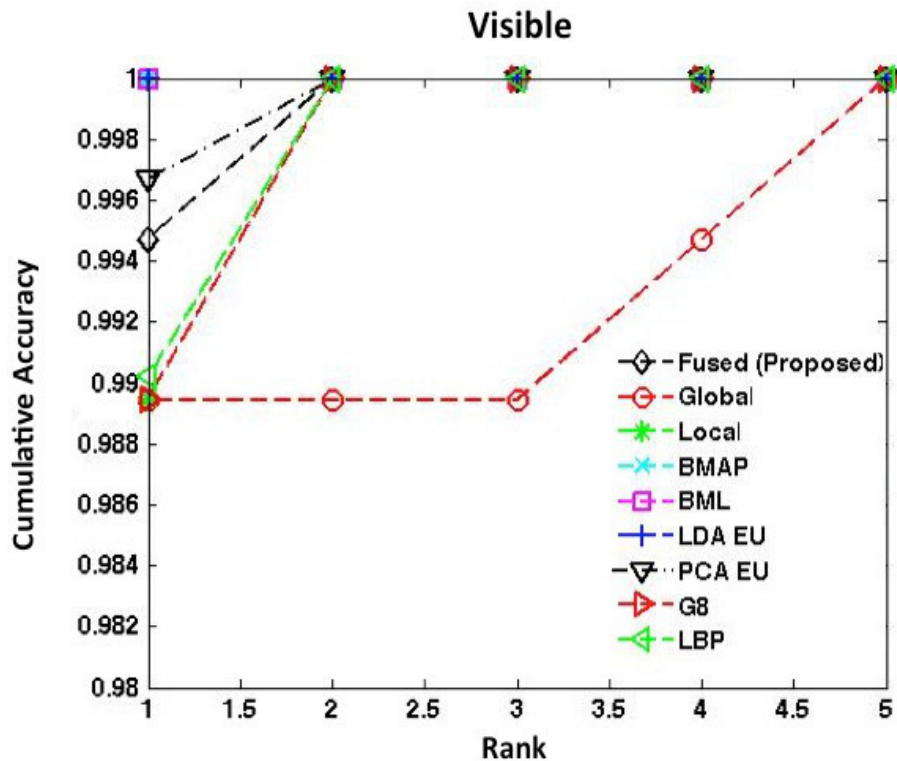
Figure 3.11: CMC curves comparing proposed methodology with commercial and academic software for entire database (assuming no errors in pre-processing) for visible spectrum.

we found that the visible, MWIR, and LWIR datasets, our fused matching approach produced an equal error rate (EER) of 1.6 %, 1%, and 2.7% respectively. It appears the MWIR spectrum has the lowest EER for verification experiments. ROC curves for the verification experiments can be seen in Fig. 3.14. It is evident from this experiment that fusion results in the best matching accuracy for our entire multi-spectral database. It is important to note that in the subsequent experiment, which uses a subset of our database (50 subjects per spectrum) in addition to SIFT and SURF, score-sum with min-max normalization does not always outperform each independent matcher.
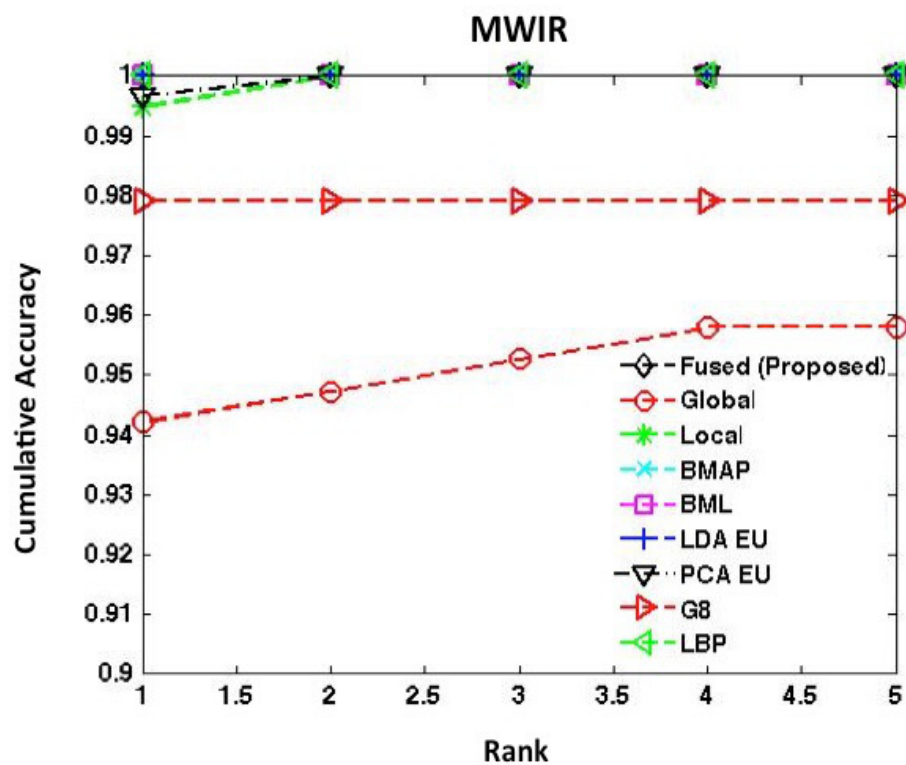
Figure 3.12: CMC curves comparing proposed methodology with commercial and academic software for entire database (assuming no errors in pre-processing) MWIR spectrum.
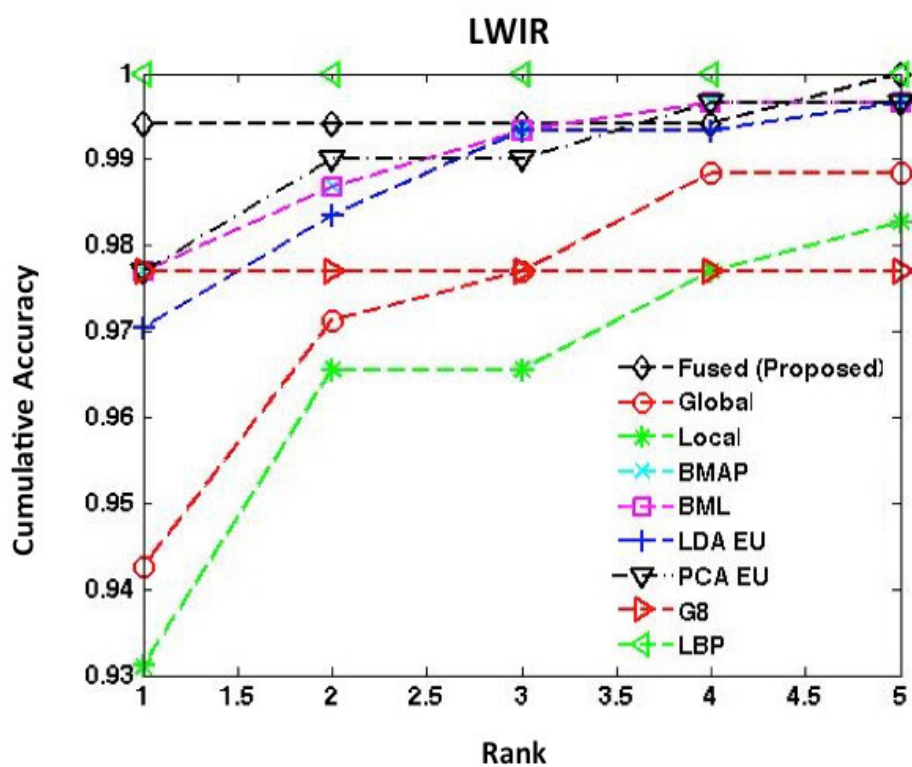
Figure 3.13: CMC curves comparing proposed methodology with commercial and academic software for entire database (assuming no errors in pre-processing) LWIR spectrum.
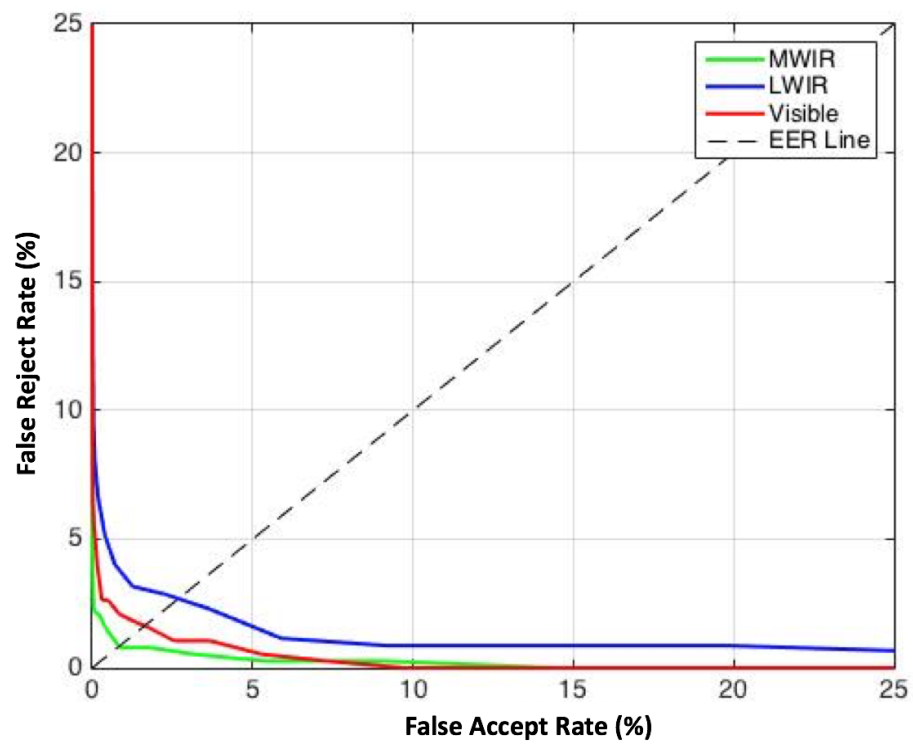
Figure 3.14: ROC curves comparing proposed methodology (after fusion) across MWIR, LWIR and Visible spectrums.

3.6.5   Matching using Sub-Facial Face Images

In this experiment, we use a subset of our entire multi-spectral face image database (50 subjects per spectrum). The automatically detected points extracted from holistic face images are partitioned depending on specific face regions, i.e. left eye, right eye, nose, chin, both eye regions, both eye and nose regions, both eye and nose regions along with the chin regions, and finally all disjoint points that are not in any aforementioned regions, but still part of the set (see Fig. 3.15). Each sub-region (e.g. left or right eye, nose and chin) is extracted using a 70×70 window around user-specific selected points centered on the left eye, right eye, nose, and chin respectively (see Fig. 3.15). Our automatic feature extraction method is evaluated against automatic point detection methods (SIFT and SURF), for all parts of the face. The results are summarized in Tables 3.1, 3.2 and 3.3 for each spectrum respectively. Fig. 3.16 shows each sub-region extracted for the global matcher and Fig. 3.17 shows minutia points detected per sub-facial region for the local matcher.

3.6.6   CLAHE Experiment

In this experiment, we use a subset of our entire multi-spectral face image database (35 subjects per spectrum). We evaluate the application of photometric normalization technique, contrast limited adaptive histogram equalization (CLAHE) [44]. The results are summarized in Table 3.4 for each spectrum.

Table 3.1: Identification results at rank-1 (%) for holistic and sub-facial features of visible images using our global, local, fused, and proposed (fused segmentation and minutiae) matchers for 50 subjects.

	Rank-1 Accuracy (%)						
	Matcher(s)						
	Global	Local			Fused		
Region(s)	Seg.	Minutiae	SIFT	SURF	**Proposed**	Seg. & SIFT	Seg. & SURF
Whole Face	98.00	98.00	55.00	99.00	98.00	99.00	100.00
Left Eye	83.00	14.00	10.00	45.00	48.00	46.00	81.00
Right Eye	82.00	17.00	06.00	43.00	42.00	34.00	75.00
Nose	78.00	14.00	07.00	24.00	50.00	40.00	69.00
Mouth	52.00	13.00	06.00	13.00	52.00	22.00	53.00
Eyes	89.00	32.00	13.00	72.00	69.00	44.00	94.00
Eyes and Nose	91.00	50.00	16.00	82.00	88.00	81.00	95.00
Eyes, Nose, Mouth	92.00	66.00	26.00	81.00	87.00	88.00	95.00
Disjoint	96.00	83.00	43.00	99.00	98.00	93.00	99.00

Table 3.2: Identification results at Rank-1 (%) for holistic and sub-facial features of MWIR images using our global, local, fused, and proposed (fused segmentation and minutiae) matchers for 50 subjects.

	Rank-1 Accuracy (%)						
	Matcher(s)						
	Global	Local			Fused		
Region(s)	Seg.	Minutiae	SIFT	SURF	**Proposed**	Seg. & SIFT	Seg. & SURF
Whole Face	97.36	100.00	97.36	94.73	100.00	100.00	100.00
Left Eye	61.00	52.00	44.73	31.57	74.00	71.05	57.89
Right Eye	68.42	62.00	43.42	22.37	81.00	75.32	77.63
Nose	65.78	73.00	43.42	38.15	81.00	69.73	73.68
Mouth	94.00	66.00	38.15	13.16	92.00	73.68	69.73
Eyes	76.31	82.00	76.31	52.63	92.00	85.52	89.47
Eyes and Nose	86.48	92.00	72.36	68.42	98.00	94.73	93.42
Eyes, Nose, Mouth	88.15	100.00	80.26	73.68	98.00	94.74	93.42
Disjoint	93.00	99.00	89.47	84.21	99.00	100.00	98.68

Table 3.3: Identification results at Rank-1 (%) for holistic and sub-facial features of LWIR images using our global, local, fused, and proposed (fused segmentation and minutiae) matchers for 50 subjects.

	Rank-1 Accuracy (%)						
	Matcher(s)						
	Global	Local			Fused		
Region(s)	Seg.	Minutiae	SIFT	SURF	**Proposed**	Seg. & SIFT	Seg.& SURF
Whole Face	97.00	81.00	99.00	94.59	97.00	94.00	100.00
Left Eye	57.00	10.00	11.00	31.00	40.00	41.00	55.40
Right Eye	64.00	08.00	21.00	26.00	44.00	43.00	61.00
Nose	54.00	11.00	15.00	36.00	39.00	40.00	64.00
Mouth	55.00	09.00	03.00	34.00	43.00	48.00	55.00
Eyes	73.00	17.00	21.00	51.00	62.00	62.00	82.00
Eyes and Nose	80.00	31.00	35.00	77.00	67.00	68.00	95.00
Eyes, Nose, Mouth	84.00	46.00	37.00	84.00	77.00	74.00	95.00
Disjoint	90.00	78.00	45.00	92.00	94.00	86.00	96.00

Table 3.4: Identification results at Rank-1 (%) for photometric normalization (CLAHE) using proposed global, local (minutiae) and fused matchers for 35 subjects.

	Rank-1 Accuracy (%)		
	CLAHE		
Spectrum	Global	Local(Minutiae)	**Proposed**
Visible	61.43	91.43	94.29
MWIR	94.29	97.14	95.71
LWIR	85.71	95.71	97.14
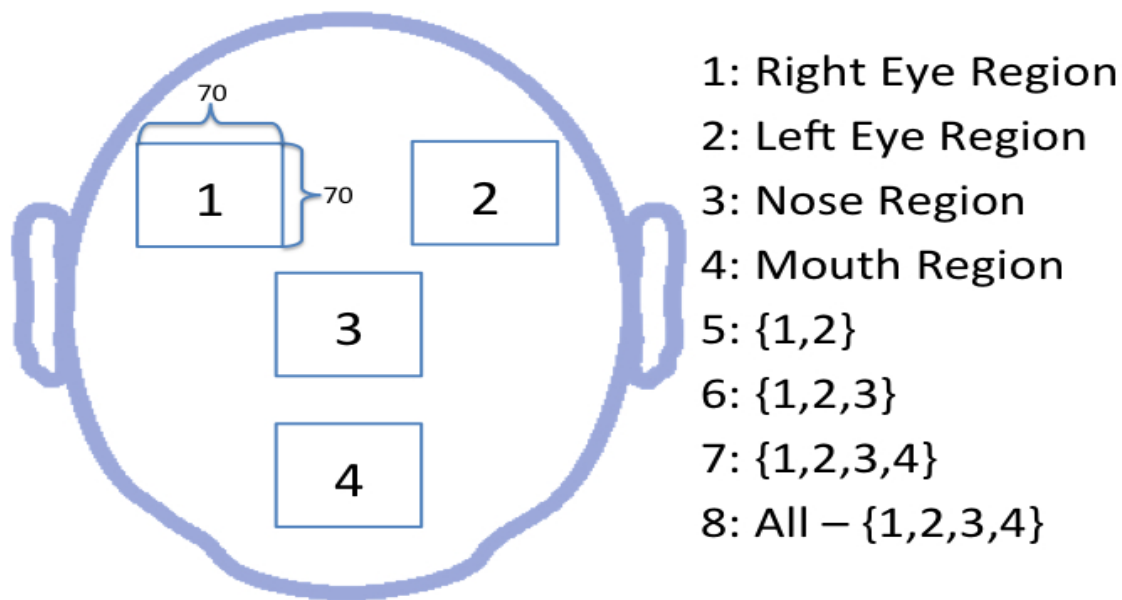
Figure 3.15: Face template for regions of face matched using 70x70 windows.

3.7   Chapter Summary

In this chapter, we investigated the use of extracted face based features composed of veins, edges, wrinkles, and perimeter lines, using two different matchers for face recognition. The experiments were carried out across the visible band and passive (MWIR and LWIR) bands for canonized faces.

Skin segmentation remains a difficult task, particularly in the visible spectrum due to variations in illumination, therefore requiring a number of different thresholds for the task. However, because the visible spectrum is a traditional research area in biometrics, pre-processing tasks such as eye detection hold an advantage in the visible spectrum. Eye detection in the passive infrared band is an area that still needs some attention. CLAHE appears to offer little to no advantage for our local matcher, however CLAHE may be advantageous
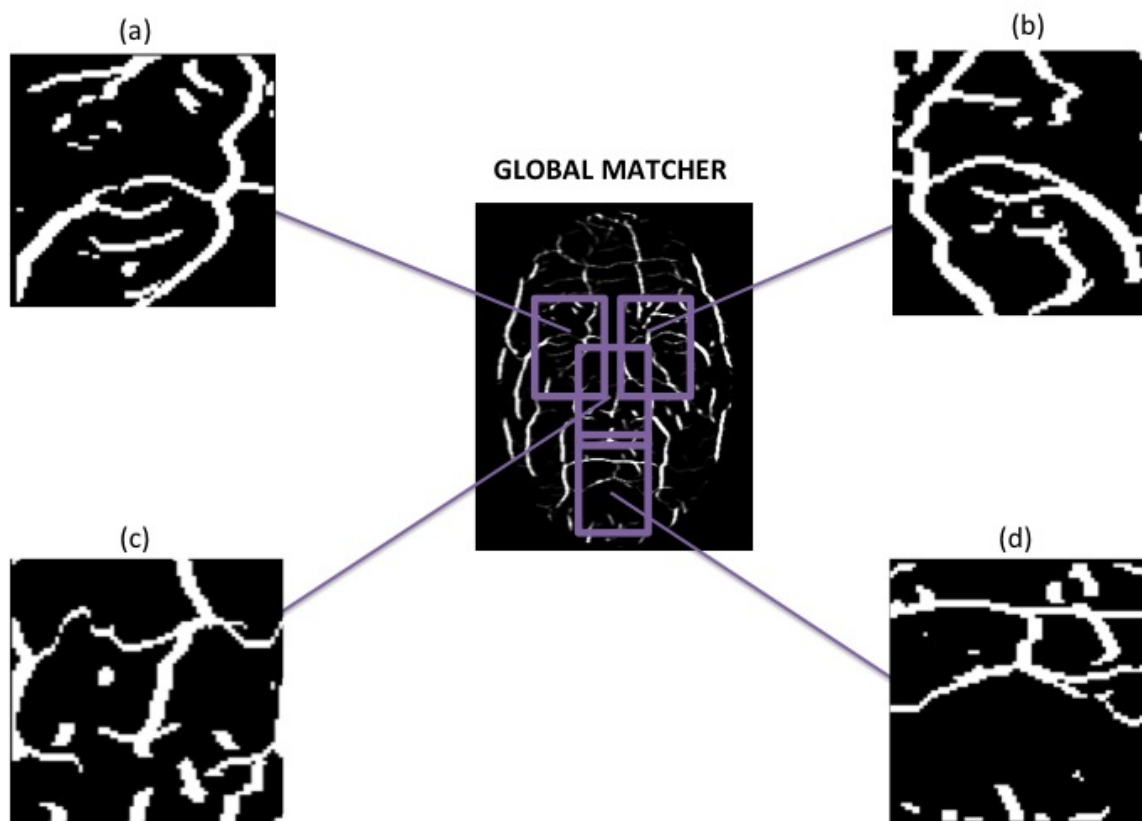
Figure 3.16:   Sample subject image of independent face region using global matcher and detected features: (a) Left Eye, (b) Right Eye, (c) Nose and (d) Mouth/Chin regions.
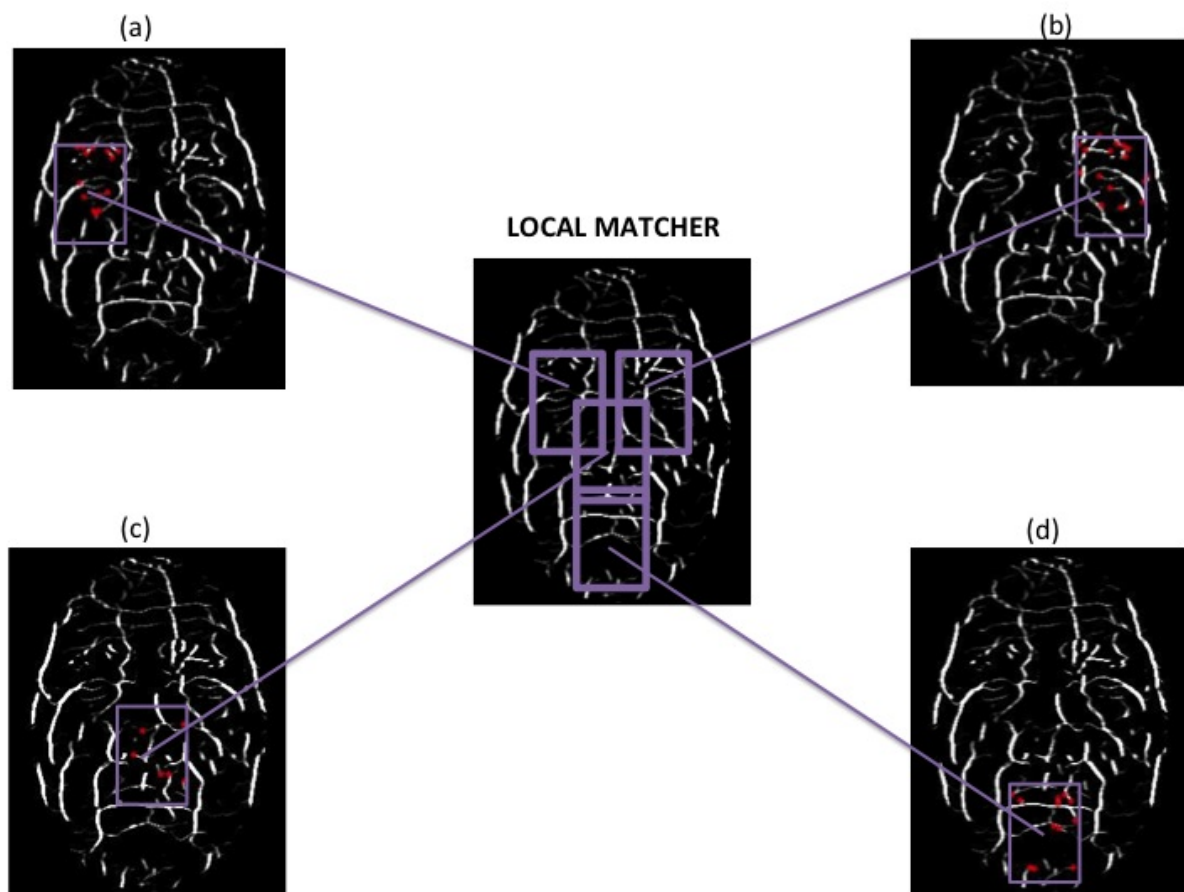
Figure 3.17: Sample subject image of independent face region using local matcher and detected features: (a) Left Eye, (b) Right Eye, (c) Nose and (d) Mouth/Chin regions.

for our global matcher. This is not a generalization of the effects of photometric normalization on our algorithm. It may be advantageous to explore other photometric normalization techniques in future work. Due to the use of physiological and geometric features for our matchers, eye detection accuracy is crucial to the sensitivity of our automated system. An advantage to our proposed approach is that it is semi-automated and does not require a design and the usage of a training set. This saves a lot of computational time and effort in processing. Another observation is that our pre-processing step reduces image storage size for a given subject from as large as 3 MB after video extraction, down to 1 KB, which is beneficial when storing a large number of subjects. Despite the fact that we only dealt with frontal face images, the design of our pre-processing step turned to be very important in achieving high recognition rates.

For the global matcher, the best performance was achieved in the MWIR spectrum with a 99.47% rank-1 accuracy. For the local matcher, the best performance was achieved in the visible spectrum with a 98.95% rank-1 accuracy. The number of subjects our system incorrectly identifies can be decreased, and overall performance increased with the addition of samples to the image gallery. Although the global matcher is computationally efficient, the local matcher outperforms the global matcher across the visible and MWIR spectrums. When matchers are fused, a rank-1 accuracy of at least 99.43% is achieved across all spectrums.

A subset of each dataset was used for experimenting on different methods for feature point detection, along with matching individual parts of the face, such as the eye, nose, and chin regions. A very interesting result from our study was in the case where our proposed methodology was applied to the disjointed face sub-region (i.e. the region of the face excluding the eyes, nose, and chin regions) in the MWIR spectrum. In this case we achieved at least a 93% rank-1 accuracy using the global matcher and 99% rank-1 accuracy using the local matcher. Another interesting result in the MWIR spectrum was that the mouth region could be matched using our global matcher with 94% confidence. These partial face match-

ing results suggest that the features extracted around the cheek, forehead and chin regions are unique per individual, particularly in the MWIR spectrum. It appears that our global matcher outperforms our local matcher in matching sub-facial regions for each spectrum except the MWIR spectrum. Fusing matchers for sub-facial regions does not always increase the recognition performance, but performance is always increased for holistic face matching when matchers are fused. For each subset of our databases, it took less time to match subjects using the global matcher in comparison to the local matcher. Through our experiments and baseline comparison, we can conclude that the FR in LWIR spectrum doesn't perform as well as FR in the visible spectrum. However, a reason for the decrease in performance may be due to the lower resolution images in our LWIR dataset. Furthermore, although users could be identified using parts of the face instead of the whole face, face sub-regions should be used in conjunction with other biometric approaches to boost performance.

Chapter 4

Heterogeneous Face Recognition

4.1 Bridging the Visible and Passive IR Bands: Cross-Spectral FR Evaluation

We conduct cross-spectral FR experiments to compare the matcher proposed in Chapter 4 with previous literature [2]. We can hypothesize that our proposed feature extraction and matching algorithm will not perform well in cross-spectral FR scenarios because it is highly dependent on the consistency and alignment of extracted features being matched. If non-identical contours are not extracted consistently across each respective spectrum, face recognition accuracy will decrease (see Fig. 4.1). A subset of data collected in the MWIR spectrum was used in experimentation. In the work done by [2], six different discriminants were used (LBP,LTP, PHOG, SIFT, TPLBP and FPLBP), two different illumination normalization approaches (SQI and DoG), and four similarity measures (L1, L2, Chi and DT). Fusion at the feature level was also employed both before and after applying illumination normalization when using the L1 and L2 measures, but omitted in Table 4.1. While all combinations of descriptors, illumination normalization and distance measures were tested in [2],
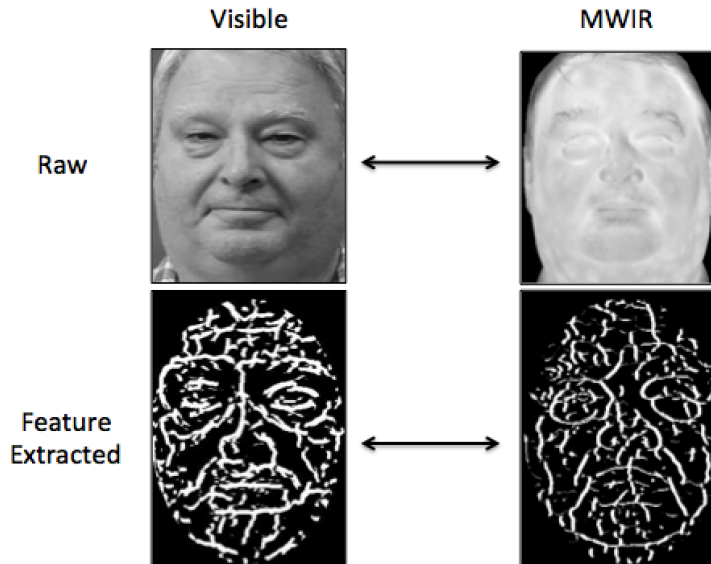
Figure 4.1: Comparison of raw (top row) and segmented (bottom row) features from visible and MWIR face images used in cross-spectral FR evaluation.

we compare their best results with our proposed matcher from Chapter 4 prior to fusion. The results of the experiments illustrate the limitations of our proposed matcher and the potential upside of localized texture based matcher(s) such as LBP/LTP when employing cross-spectral FR. Results can be seen in Table 4.1.

4.2   Introduction

There are various challenges in comparing passive IR (probes) against visible face images (gallery), as many largely deployed FR systems usually require the individual to enroll using their visible face images. Previous work showed that the use of geometric and photometric normalization techniques, applied in both gallery and probe sets, can improve same-band face matching performance ([2, 88]). This is not the case for cross-spectral matching. Therefore, it is vital that approaches for alleviating the matching problem between passive IR and visible

Table 4.1: Rank-1 FR results (%) for Cross-Spectral FR Evaluation of VIS-MWIR using Normalization (NORM) and Similarity Measurements (SM) [2]

| ALGORITHM (NORM) [SM] | Rank-1 (%) FR Accuracy | | | |
	MWIR Gallery vs. VIS Probe	VIS Gallery vs. MWIR Probe	Subjects	Samples
LBP (DoG) [Chi]	30.80	16.7	39	7
LBP (DoG) [DT]	19.70	18.00	39	7
LTP (DoG) [Chi]	32.50	21.80	39	7
LTP (DoG) [DT]	23.90	21.20	39	7
LBP+SIFT+PHOG (SQ1) [L1]	37.60	18.60	39	7
LBP+SIFT+PHOG (SQ1) [L2]	36.80	21.20	39	7
TPLBP (DoG) [Chi]	**53.90**	**42.31**	39	7
FPLBP (DoG) [Chi]	29.90	28.80	39	7
Proposed (CLAHE) [Local]	2.38	2.38	42	4
Proposed (CLAHE) [Global]	1.70	1.70	42	4

face images receive more attention. In Fig. 4.2 we can clearly see the benefits of geometric and photometric normalization, i.e. it results in improving similarity scores (cross-spectral matching when using Local Binary Patterns). In Fig. 4.2(b) Geometric Normalization is shown to improve the score by a factor of 7. In Fig. 4.2(c) Photometric Normalization further increases the geometrically normalized score by a factor of 1.5. However, we can also see that we are still far away in generating a similarity score as good as the one computed when comparing images from the same spectrum (e.g. visible against visible) as seen in Fig. 4.2(d). In order to further improve such a score, one of the solutions discussed in the literature is the usage of image synthesis (see Fig. 4.5). However, while there is existing work on image synthesis of human faces [89–93] the approaches are different, and in these studies the authors did not focus on addressing the cross-spectral face matching problem for 2D face images in the context of (VIS-IR) FR systems. More recent work [94] focuses on thermal to visible face recognition through the use of deep neural networks to capture and learn the largely non-linear association between the two spectrums, while preserving identity information. We use an alternate approach to determine whether the cross-spectral
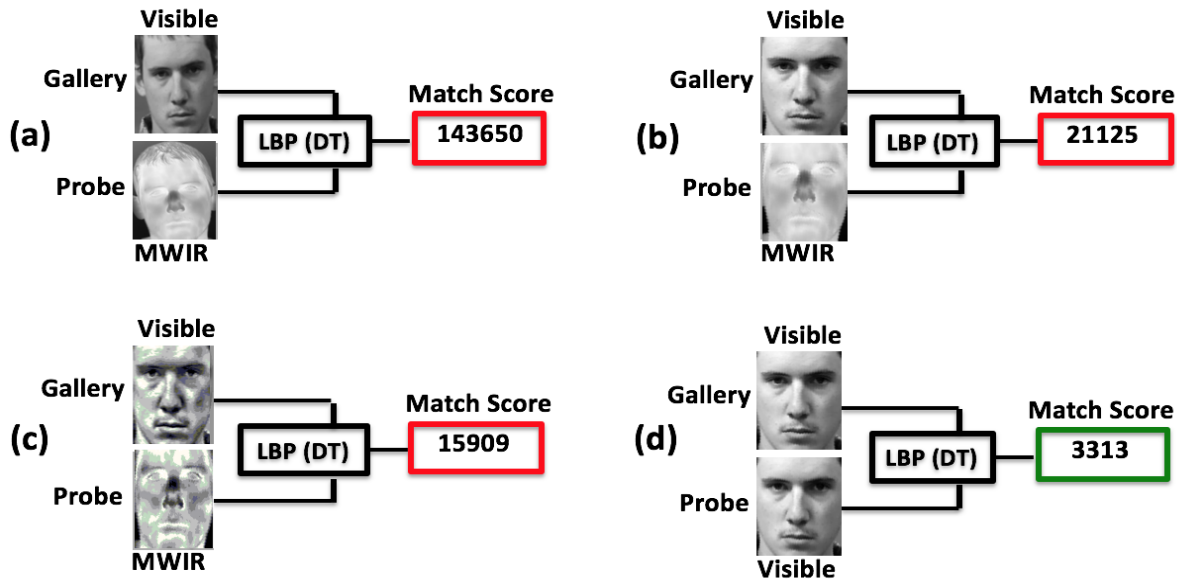
Figure 4.2: LBP (DT) is a distance based matcher (lower score indicates closer match, 0 is perfect match). Match scores for the same subject (a) across different spectrums using Face Detection, (b) across different spectrums with Geometric Normalization, (c) across different spectrums with Geometric and Photometric Normalization (CLAHE), (d) across the visible spectrum using Geometric Normalization

gap (in terms of face recognition accuracy) can be further improved, across all subjects, after image synthesis and image denoising is applied. The question is, whether the cross-spectral gap (in terms of face recognition accuracy) can be further improved, across all subjects, after image synthesis, image denoising, and cohort filtering is applied. Fig. 4.3 illustrates main components typically found in a FR system and the contributions to those respective components made in this work.

4.3   Heterogenous FR Analysis and Methods

Performing a direct match between images from different spectrums (i.e., passive infrared and visible) is challenging because each spectrum contains different information pertaining to the subject. In the visible spectrum, we are able to determine pixel intensity across the face
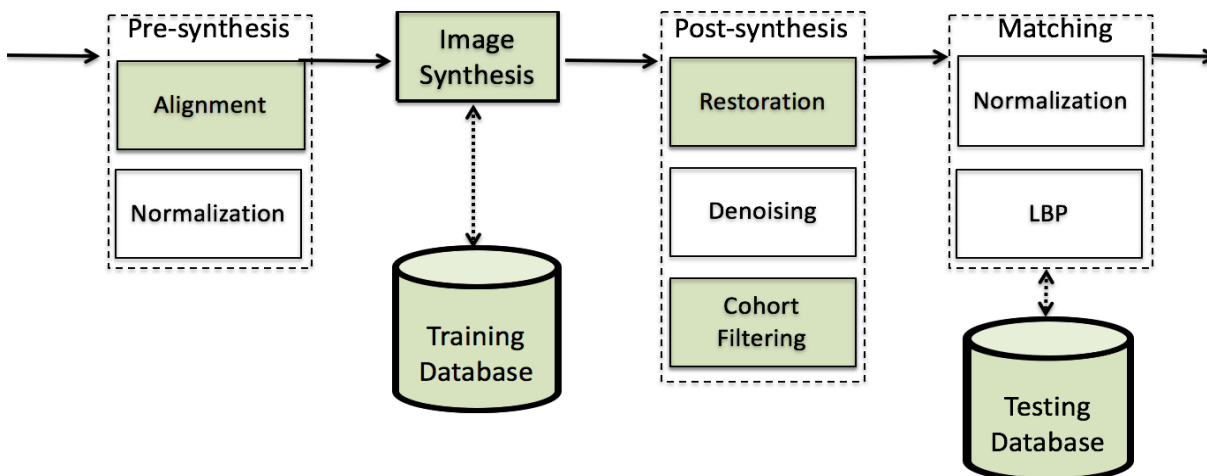
Figure 4.3: Main components typically found in a FR system. The shaded green objects illustrate areas in which contributions are made. There aren't any novel contributions that were made to the area of photometric normalization, denoising, or matching (LBP) in this work.

of the subject. On the other hand, the passive infrared spectrum reveals thermal properties across the face of the subject. There has been a concerted effort in research to provide alternate approaches to various heterogeneous face recognition problems, despite spectrum of operation. One of the earliest contributions to heterogeneous FR research was the use of viewed sketches for sketch recognition, and eventually into other modalities such as NIR.

Tang et al. pioneered work in heterogeneous FR with a number of approaches to synthesize a visible image using a sketch (or vice-versa) [95–98]. In [95] Tang and Wang studied an effective photo-to-sketch transformation method using eigen-faces. In [96], pseudo-sketches of photos are automatically synthesized using LLE inspired local geometry preserving algorithm and training photo sketch pair samples. Kernel based nonlinear discriminant analysis (KNDA), which combines the nonlinear kernel trick with linear discriminant analysis (LDA), is adopted to match the probe sketch with the pseudo-sketches. Later, Liu et al. [97] performed the transformation using a Bayesian MAP framework that integrates tensor modeling and statistical optimization to ensure inference performance of the difficult vision problem.

Wang and Tang also proposed a multiscale Markov Random Fields (MRF) model, which learns the structure of the face across different sized ranges, for converting a sketch into a photograph [98]. In the NIR domain, Lei at al. [99] proposed a NIR imaging and a statistical learning approach of tensor modeling with CCA based mapping for facial shape recovery from a single image. The heterogeneous FR problem is a major challenge in FR systems, but becomes even more complex when other challenges such as pose and illumination invariation are introduced. Recently, Lei and Li [100] suggested solving the same problem via a low dimensional representation for each face, using a discriminative graph embedding method.

Several works [101–103] research lighting and pose variations in heterogeneous FR across the Sketch-VIS and NIR-VIS scenarios. Zhang et al. [101] introduce shape priors, robust patch matching, and new compatibility terms to synthesize face sketches from photos with different lighting and poses (Sketch-VIS). Their method utilized by Zhang et al. [101] is formulated using the multiscale MRF. Chen et al. [102] propose a patch based transformation method where a virtual sample is synthesized from an input sample. Through their methodology, they are able to reduce the intrapersonal difference caused by completely different lightings (NIR-VIS). Li et al. [103] overcome the problem of illumination variation by extracting LBP features from NIR images. In [103] the AdaBoost approach is utilized to learn a powerful FR engine based on the constant representation.

Aside from the generative transformation-based approaches, recent research in heterogeneous FR utilize approaches that are discriminative feature-based [34, 104–108], which have demonstrated promising matching results in both the sketch and NIR domains. Through the initial use of local feature discriminants such as local binary patterns (LBP) and SIFT descriptors, these approaches represent face images. In [34] Klare and Jain incorporate SIFT feature discriminants and an RS-LDA based scheme to match NIR to visible light images. In [106] a novel method of heterogeneous FR that uses a common feature-based representation for both NIR images as well as VIS images is proposed. Linear discriminant analysis

is performed on a collection of random subspaces to learn discriminative projections. NIR and VIS images are matched directly using the random subspace projections, and sparse representation classification. Sarfraz et al. [94] use deep learning methods to benchmark the Carl thermal-visible dataset (NVESD) where there are changing activity levels and variations in subject-to-camera distance, and illumination. Other implementations, on top of using non-linear dimensionality reduction and manifold learning, also use photometric normalization for optimal feature discrimination based on the spectrum of operation, and image reconstruction using the training data during the testing phase, in place of inferred features. An example of how image synthesis works is provided in Fig. 4.4. Please note that unlike other heterogeneous thermal-visible matching approaches, this work only uses the facial information (after face detection and normalization) for synthesis, restoration and matching. We do not use the entire thermal head signature that includes more features that may result in enhanced accuracy as for example in [109]. Chen et al. [110] addressed the heterogeneous FR problem by generating multiple subspaces through a cascaded subspace learning scheme based on whitening transformation, factor analysis and discriminant analysis models.

4.4   Image Synthesis Analysis and Methods

The upside of synthesis-based methods is that once conversion has been completed, existing FR algorithms can be used for matching. We review three types of approaches for image synthesis: (i) face synthesis analysis; (ii) subspace methods; (iii) 3D-based approaches.

4.4.1   Face synthesis analysis

Li et al. [91] propose a stereoscopic synthesis method that produces frontal face images based on two different poses of face images that are co-captured. In [111] face images are transformed from one type to another using face analogy software and then subsequently

synthesized query images are matched against gallery images. Zhang et al. [112] developed a face synthesis approach where corresponding sparse coefficients of visible and NIR images are assumed to be alike through learning pairs of an over-complete dictionary.

4.4.2 Subspace Methods

In [92] the authors augment a challenging database consistent of just one sample per subject by synthesizing new face samples of various degrees using edge-based information. Yi et al. [90] and Dou et al. [113] utilized canonical correlation analysis (CCA) to learn the relationship between face pairs using 9 out of 10 samples from each subject for the training algorithm, and the remaining sample for conversion.

4.4.3 3D-based Methods

Video can be used to extract 3D features instead of utilizing a 2D face image. Ansari et al. [93] created a database of 3D textured face models composed of 114 subjects using stereo images and a generic face mesh model for 3D FR application. In [89] a 3D generic face model is aligned with each frontal face image.

4.5 Heterogeneous FR Framework

The formulation of the cross-spectral matching problem that we want to solve is as follows. Given a IR image I_p as input, we estimate the target VIS image V_p with the help of a training set of IR images I_q and the corresponding VIS images V_q. We represent each IR or VIS image as a set of small image patches that overlap. I_p and V_p have the same number of patches, and each IR image in I_q and the corresponding VIS image in V_q also have the same number of patches.

Utilizing the IR patches, linear models are used to predict their corresponding VIS patches. An input IR patch is represented as $y \in \Re^D$, where D is its pixel number. Then from each training IR-VIS image pair, we extract patches of the same size at the same position as y. These IR and VIS patches are denoted as Y and X respectively. To predict y's VIS counterpart x, relations learned from X and Y are used. Following the assumption of image manifolds ([114–116]), we take Y and X as samples drawn from two manifolds. Therefore, the neighborhood of y in Y can be seen as lying on a linear subspace. A search is conducted for the K nearest neighbors in Y for y, Y_N. Their corresponding VIS patches are denoted as X_N. Canonical Correlation Analysis (CCA) is used to model the linear relations between the linear subspaces spanned by Y_N and X_N. CCA discovers one set of axes for each dataset, along which these two sets of data co-vary most. From the viewpoint of learning, CCA finds the most linear predictable components for the two sets.

In the proposed approach, patch size and degree of overlap between adjacent patches are taken into consideration. Ideally, each patch generated for the VIS image V_p should be related appropriately to the corresponding patch in the IR image I_p and also preserve some inter-patch relationships with adjacent patches in V_p. In Figure 4.4, we draw a flowchart for the proposed solution to our cross-spectral matching problem.

4.6   Feature Analysis

The correspondence between VIS and IR images is a many-to-many mapping, therefore it is not possible to learn a simple linear relationship between the two image spaces. Instead, a solution can be extracting features and learning the relationship between the feature spaces. We aim to extract features from the original image with the desired properties: (1) The relationship between two feature spaces is unvarying, i.e. there is a one-to-one mapping between them that can easily be learned from the training set, and performs well when gen-
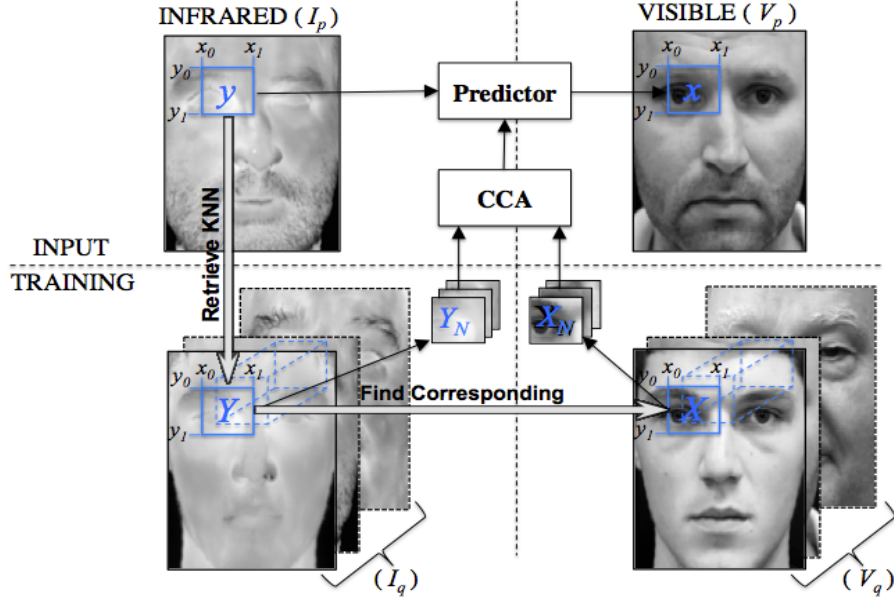
Figure 4.4: Flow chart of proposed image synthesis.

eralized to the test set; (2) The features should contain enough information to approximately recover the images in two distinct feature spaces. Unfortunately, the two properties conflict with each other in our problem. Principal Component Analysis (PCA), which is known as the EigenFace method [117] in FR, is a popular method to extract features. For our problem it well satisfies the second condition above, but two sets of principal components, extracted from a VIS image and the corresponding IR image, have weak correlations. CCA finds pairs of directions that yield the maximal correlations between two data sets or two random vectors, i.e. the correlations between the projections (features) of the original data projected onto these directions are maximized. CCA has the desired characteristics as given in the above property (1). But unlike PCA, several CCA projections are not sufficient to recover the original data, because the found directions may not be able to cover the principal variance of the data set. However, it has been found that regularized CCA is a satisfying trade-off between the two desired properties [118].

4.6.1 Manifold Learning Dimensionality Reduction

Our method is based on the assumption that small patches in VIS and IR images form manifolds with similar local geometry in two distinct spaces. This assumption is valid because the resulting representation is stable, and hence independent of the spectrum, as long as the embedding is isometric. Each patch, represented as a feature vector, corresponds to a point in one of the two feature spaces. Recently, some new manifold learning (or nonlinear dimensionality reduction) methods have been proposed to automatically discover low-dimensional nonlinear manifolds in high-dimensional data spaces and embed them onto low-dimensional embedding spaces, using tractable linear algebraic techniques that are not prone to local minima. These include isometric feature mapping (Isomap) [119, 120], locally linear embedding (LLE) [121, 122], and Laplacian eigenmap [123, 124]. Our image synthesis method to be described below has been inspired by these manifold learning methods, particularly LLE.

4.7 Cross-Spectral VIS-IR Database

A total of two different datasets were utilized for our experiments. One of the datasets was collected and assembled for our experiments in our lab. Each dataset consists of only full frontal face images with a neutral facial expression for every subject. Three different methods are performed for pre-processing all subjects prior to synthesis. A total of about 78 subjects were used in the construction of our database. Each subject had 4 samples that were used for our matching and synthesis experiments (see Fig. 4.6).

- **WVU** The (1) VIS-MWIR subset consists of 308 bitmap images (154 for probe and 154 for gallery) with four sequential images in time per subject (77 subjects). Visible images for this database were extracted from videos captured in our laboratory,
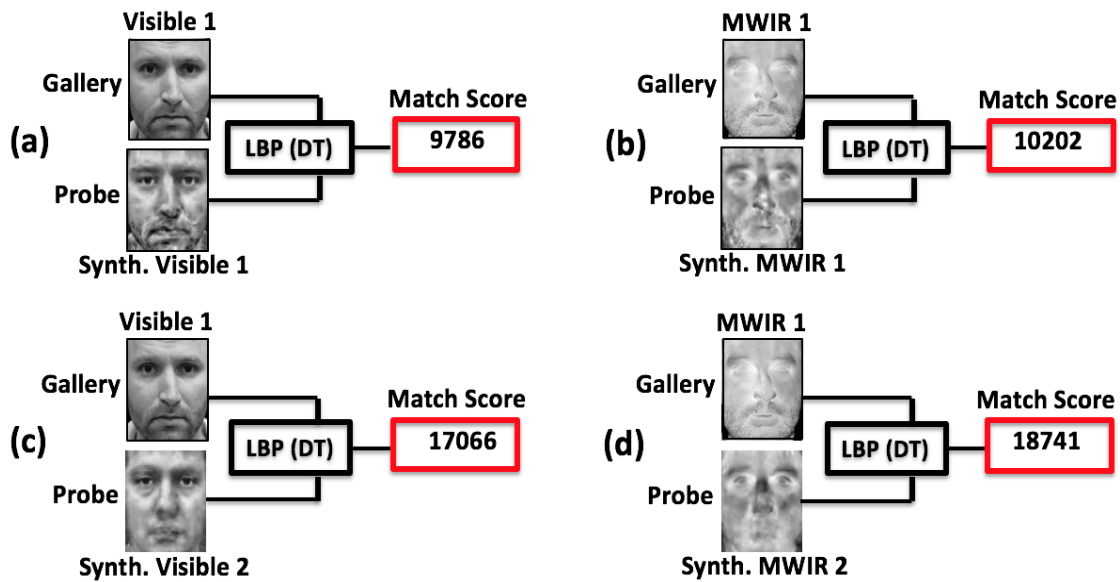
Figure 4.5: LBP (DT) is a distance based matcher (lower score indicates closer match, 0 is perfect match). Match scores for the same subject after geometric normalization and synthesis (a) visible subject 1 to synthesized visible subject 1, (b) MWIR subject 1 to synthesized MWIR subject 1, (c) visible subject 1 to synthesized visible subject 2, and (d) MWIR subject 1 to synthesized MWIR subject 2. Scores are compelling when compared to Fig. 4.2.

using a Canon EOS 5D Mark II camera, where a full image contained the subject's complete head and shoulders. This digital SLR camera produces ultra-high resolution RGB color images or videos, with a resolution of 1920×1080 pixels. The face images are obtained by obtained from the movie files in JPEG format. The MWIR images for this database were extracted from videos captured in our laboratory, using a FLIR SC8000 MWIR camera, where a full image contained the subject's complete head and shoulders. The infrared camera produces high definition thermal videos, with a resolution of 1024×1024. The (2) VIS-LWIR subset consists of 312 bitmap images (156 for probe and 156 for gallery) with four sequential images in time per subject (78 subjects). Visible images for this subset were extracted from videos captured using the aforementioned Canon EOS 5D Mark II camera. The LWIR images for this subset were extracted from videos captured in our laboratory, using a FLIR SC600 LWIR camera, where a full image contained the subject's complete head and shoulders. The science-grade infrared camera produces high-resolution LWIR images or videos, with a resolution of 640×480 pixels. The first 2 samples were utilized as gallery images, while the remaining 2 samples were the probe images. It is noteworthy that images between sensor pairs were not captured simultaneously or co-registered (e.g. captured in multiple bands at the same time), making our database more challenging given our patch-based approach.

- **NVESD:** The NVESD dataset [125] was acquired as a joint effort between the Night Vision Electronic Sensors Directorate of the U.S. Army Communications-Electronics Research, Development and Engineering Center (CERDEC), and the U.S. Army Research Laboratory (ARL). The portion of NVESD dataset examined two experimental conditions: vigorous exercise in the form of a fast paced walk and subject-to-camera

range (1 m, 2 m, and 4 m). A group of 25 subjects were imaged before and after exercise at each of the three ranges. Another group of 25 subjects were at rest and imaged at each of the three ranges. All 50 subjects were used to create the dataset, however only a subject-to-camera range of 1m and 2m was used for our dataset. For the (1) VIS-MWIR subset, visible images were captured using the Basler Scout GigE Vision Sensor sc640-74gm equipped with Sigma 24 mm f/1.8 EX DG Aspherical Macro Large Aperture Wide Angle Lenses. The visible sensor was used to acquire 8-bit grayscale facial images and was connected to GigE via a Netgear router connected to the collection PC. The sensor has pixel pitches of 10 μm and spectral responses of 400-1000 nm, with a peak at 500 nm. The images were acquired for 15 seconds per capture at a resolution of 640x480 at 30 Hz. Software known as JAI Camera Control Tool was used to obtain the images and store them in raw, uncompressed AVI and TIFF formats. The MWIR face images were acquired using a DRS sensor. The MWIR DRS sensor was used to acquire 16-bit (12-bit) grayscale facial images in each band. The MWIR sensor has a pixel pitch of 12 μm and a spectral response of 3-5 μm. The same aformentioned Basler Scout sc6470 camera was used to acquire visible images for the (2) VIS-LWIR subset. The LWIR face images were acquired using a DRS sensor. The LWIR DRS sensor was used to acquire 16-bit (12-bit) grayscale facial images in each band. The LWIR sensor has a pixel pitch of 15 μm and a spectral response of 8-12 μm. Images were acquired for 15 seconds per capture at a resolution of 640x480 at 30 Hz. AutoIt software was used to acquire the images and store them in the .raw format. Each acquisition lasted 15 seconds at 30 frames per second for each camera, with all the sensors started and stopped almost simultaneously (subject to slight offsets because of human reaction time). To form a set of gallery and probe images for face recognition, a frame was extracted at the 1 second and 14 second marks for each video. The first 2 samples of the gallery and probe sets respectively were constructed using

DATABASE			
	DATASETS	**DESCRIPTION**	**Number of (Subjects / Samples per Subject)**
WVU	1.VISIBLE & MWIR	Controlled Face (Canon Mark II and FLIR SC800)	77 / 4
	2.VISIBLE & LWIR	Controlled Face (Canon Mark II and FLIR SC600)	78 / 4
NVESD	1.VISIBLE & MWIR	Controlled Face (Basler Scout sc640 and DRS MWIR)	50/ 4
	2.VISIBLE & LWIR	Controlled Face (Basler Scout sc640 and DRS LWIR)	50/ 4

Figure 4.6: Brief description of the databases used in this work.

the still frames of a $1m$ and $2m$ subject-to-camera range. Images between sensor pairs were captured simultaneously.

4.7.1 Co-Registered Images

Co-registration does not allude to the alignment of faces, but rather the use of multiple sensors, simultaneously during the time of acquisition, such as in the NVESD database. The location where images are acquired, and also timing (e.g. multiple samples and sessions) are an important dependence of the suggested heterogeneous framework. This requires careful consideration in experimental capture scenarios where it can be inferred that sensitivity in one location may behave similar to another. Although mitigating location dependency is a challenging task, there are major implications of this location dependency in the design and evaluation of algorithms, and this is supported through the use of co-registered images.

4.8 Image Synthesis Approach

In Fig. 4.5(a) we see a sample MWIR image synthesized to its visible counterpart (synthesized visible). In Fig. 4.5(b) a sample visible image is synthesized to its MWIR counterpart (synthesized MWIR). We have also included sample results for the difference of images from different synthesized subjects in Fig. 4.5(c-d) for completeness in analyzing the discrimina-

(a)   **(b)**   **(c)**   **(d)**
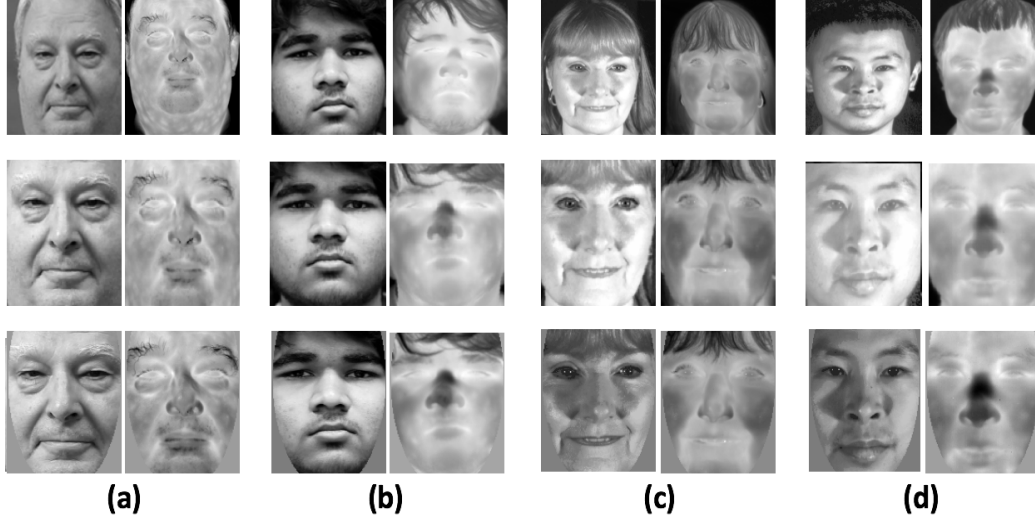
Figure 4.7:   Example images from two separate subjects from each of the two heterogeneous face recognition datasets tested in our study: (a) WVU VIS-MWIR, (b) WVU VIS-LWIR, (c) NVESD WVU-MWIR and (d) NVESD WVU-LWIR. The top row contains VIS-IR pairs of subjects aligned using face detection. The middle row contains images normalized using proposed geometric normalization. The bottom row corresponds to images normalized using CSU geometric normalization.

tive ability of face representation with respect to our algorithm. Image synthesis appears to be more favorable than cross-spectral matching using heterogeneous face images in combination with photometric normalization. Although, not as ideal as comparing images from the same spectrum (e.g. visible against visible) as seen in Fig. 4.2(d), image synthesis helps bridge the gap. Prior to feature extraction through CCA, a photometric normalization technique (e.g. CLAHE) is applied, in order to strengthen the relationship between patches of differing spectra.

4.8.1   Canonical Correlation Analysis

Through the use of two random variables with zero-mean $\mathbf{x}$, a $p \times l$ vector, and $\mathbf{y}$, a $q \times l$ vector, CCA finds the 1st pair of directions $\mathbf{w}_1$ and $\mathbf{v}_1$ with maximum correlation between the projections $x = \mathbf{w}_1^{\mathrm{T}}\mathbf{x}$ and $y = \mathbf{v}_1^{\mathrm{T}}\mathbf{y}, \max \rho(\mathbf{w}_1^{\mathrm{T}}\mathbf{x}, \ \mathbf{v}_1^{\mathrm{T}}\mathbf{y}) \ , s.t. \ Var((\mathbf{w}_1^{\mathrm{T}}\mathbf{x} =$

1) and $Var(\mathbf{v_1}^T\mathbf{y} = 1)$, where the correlation coefficient is ρ, the variables x and y are known as the first canonical variates, and the $\mathbf{w_1}$and$\mathbf{v_1}$ represents the initial correlation direction vector. CCA finds kth pair of directions $\mathbf{w_k}$ and $\mathbf{v_k}$ which satisfies:(1) $\mathbf{w_k}^T\mathbf{x}$ and $\mathbf{v_k}^T\mathbf{y}$ are not correlated to the previous *k-1* canonical variates; (2) the correlation between $\mathbf{w_k}^T\mathbf{x}$ and $\mathbf{v_k}^T\mathbf{y}$ is optimized under the constraints $Var((\mathbf{w_1}^T\mathbf{x} = 1)$ and $Var(\mathbf{v_1}^T\mathbf{y} = 1)$. Then $\mathbf{w_k}^T\mathbf{x}$ and $\mathbf{v_k}^T\mathbf{y}$ are called the kth canonical variates, and $\mathbf{w_k}$ and $\mathbf{v_k}$ are the kth correlation direction vector, k $\leq$ min$(p,\ q)$. The solution for the correlation of coefficients and directions is not different from the generalized eigenvalue problem seen here,

$$(\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}^{T} - \rho^2\Sigma_{xx})\mathbf{w} = 0 \ , \tag{4.1}$$

$$(\Sigma_{xy}^{T}\Sigma_{xx}^{-1}\Sigma_{xy} - \rho^2\Sigma_{yy})\mathbf{v} = 0 \ , \tag{4.2}$$

where Σ_{xx} and Σ_{yy} are the self-correlation while the Σ_{xy} and Σ_{yx} are the co-correlation matrices respectively. Through CCA, the correlation of the two data sets are prioritized, unlike PCA, which is designed to minimize the reconstruction error. Generally speaking, a few projections (canonical variates) are not adequate to recover the original data well enough, so there is no guarantee that the directions discovered through CCA cover the main variance of the paired data. In addition to the recovery problem, the overfitting problem should be accounted and taken care of as well. If a small amount of noise is present in the data, CCA is so sensitive it might produce a good result to maximize the correlations between the extracted features, but the features may likely model the noise rather than the relevant information in the input data. In this work we use a method called regularized CCA [118]. This approach has proven to overcome the overfitting problem by adding a multiple of the identity matrix $\lambda\mathbf{I}$ to the co-variance matrix Σ_{xx} and Σ_{yy}.

4.8.2   Feature Extraction using CCA

Local features are extracted, instead of features that are holistic, because the latter features seem to fail capturing localized characteristics and facial traits. The datasets used in training CCA consists of paired VIS and IR images. The images are divided into patches that overlap by the same amount at each position, where there exists a set of patch pairs for CCA learning. CCA locates directional pairs $\mathbf{W}^{(i)} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k]$ and $\mathbf{V}^{(i)} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]$ for VIS and IR patches respectively, where the superscript (i) represents the index of the patch (or the location of the patch within the face image). Each column of $\mathbf{W}$ or $\mathbf{V}$ is a directionary vector, which is unitary, but between different columns it is not orthogonal. For example, if we take a VIS patch $\mathbf{p}$ (which can be vectorized as a column) at position i, we are able to extract the CCA feature of the patch $\mathbf{p}$, using $\mathbf{f} = \mathbf{W}^{(i)T}\mathbf{p}$, where $\mathbf{f}$ is the feature vector belonging to the patch. For each patch and each position at each patch, we are able to acquire CCA projections using our preprocessed training database face images. Projection onto the proper directions is used to extract features, then at each patch location i we get the VIS $\mathbf{O}_v{}^i = \{\mathbf{f}_{v,j}{}^i\}$ and IR training sets $\mathbf{O}_{ir}{}^i = \{\mathbf{f}_{ir,j}{}^i\}$ respectively.

4.8.3   Reconstruction using Training Patches

In our reconstruction phase that occurs during testing, we use explicitly learned LLE weights in conjunction with our training data to reconstruct the patch and preserve the global manifold structure. Reconstructing the original patch $\mathbf{p}$ through the vectorized feature $\mathbf{f}$ is an arduous task. We are unable to recover the patch by $\mathbf{p} = \mathbf{W}\mathbf{f}$ as we do in PCA because $\mathbf{W}$ is not orthogonal. However, the original patch can be obtained by solving the least squares problem below,

$$\mathbf{p} = \arg_p\min||\mathbf{W}^T\mathbf{p} - \mathbf{f}||_2{}^2, \tag{4.3}$$

or to add an energy constraint,

$$\mathbf{p} = \mathbf{arg}_p \mathbf{min} ||\mathbf{W}^T\mathbf{p} - \mathbf{f}||_2{}^2 + ||\mathbf{p}||_2{}^2. \tag{4.4}$$

The least squares problem can be solved effectively using the scaled conjugate gradient method. In order for the above reconstruction method to be feasible, the feature vector $\mathbf{f}$ has to contain enough information about the original patch. The original patch can be recovered using LLE [121] when fewer features, represented as canonical variates, can be extracted. The assumption that localized geometries pertaining to the manifold of the feature space and that of the patch space are similar, is taken into consideration (see [114]). The patch from the image to be converted and its corresponding features have similar reconstruction coefficients. If $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_k$ are the patches whose features $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_k$ are $\mathbf{f}$'s k nearest neighbors, and $\mathbf{f}$ is able to be recovered using neighboring features with $\mathbf{f} = \mathbf{Fw}$, where $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_k]$, $\mathbf{w} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k]^\mathbf{T}$, we can reconstruct the original patch using $\mathbf{p} = \mathbf{Pw}$, where $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_k]$. Using a probe IR image, we divide it into smaller patches, and obtain the feature vector $\mathbf{f}_{ir}$ of every patch. When we infer the corresponding VIS feature vector $\mathbf{f}_v$, the VIS patch can be obtained using $\mathbf{p} = \mathbf{Pw}$ for reconstruction and then the patches will be combined into a VIS facial image. A sample illustration of the reconstruction process can be seen in Fig. 4.8 for K=5 nearest neighbors.

4.9   Face Image Denoising

Unwanted noise is introduced into the image through the image synthesis process (see Fig. 4.10). Therefore, image denoising [126] is considered as a worthy post-synthesis step that could help improve FR accuracy. Simple image filtering is not ideal for recovering useful image content because it can remove important frequency components in the pipeline. To
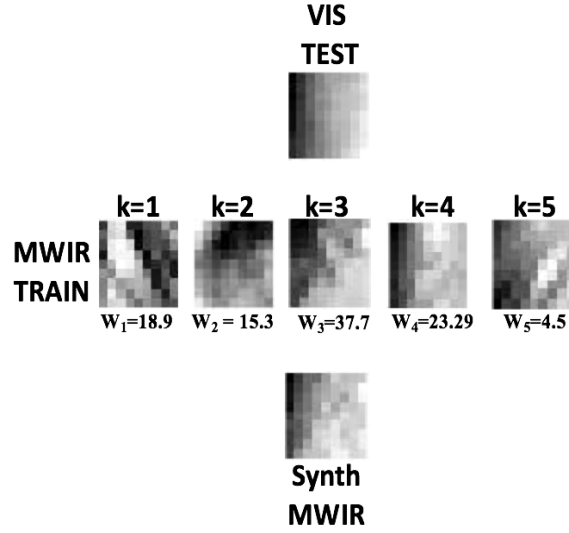
Figure 4.8: Sample illustration of input VIS patch, and corresponding training MWIR patches, for k=5 nearest neighbor. The reconstructed and synthesized MWIR image using the training patches and locally linear embedded weights.

help alleviate the challenge of effective removal of noise and subsequent image restoration, linear denoising (e.g. filtering), and nonlinear denosising (e.g. thresholding) is explored as an alternative to alleviating the noise introduced during image synthesis.

4.9.1   Denoising

1) Image Filtering-Based Linear Denoising: Linear methods can be used for image denoising so that the noise that perturbs an image is suppressed as much as possible. The filtering strength can be controlled by the filter width γ: higher values of γ increase the blurring effect. When 2-D FIR filters are designed and used with the windowing method technique, γ represents the window size of the digital filter in terms of pixels. In this paper, we found the boxcar filter with a fixed window size to be optimal for experimentation. Linear methods can cause image blurring. Therefore, this filter is efficient in denoising smooth images but not images with several discontinuities.
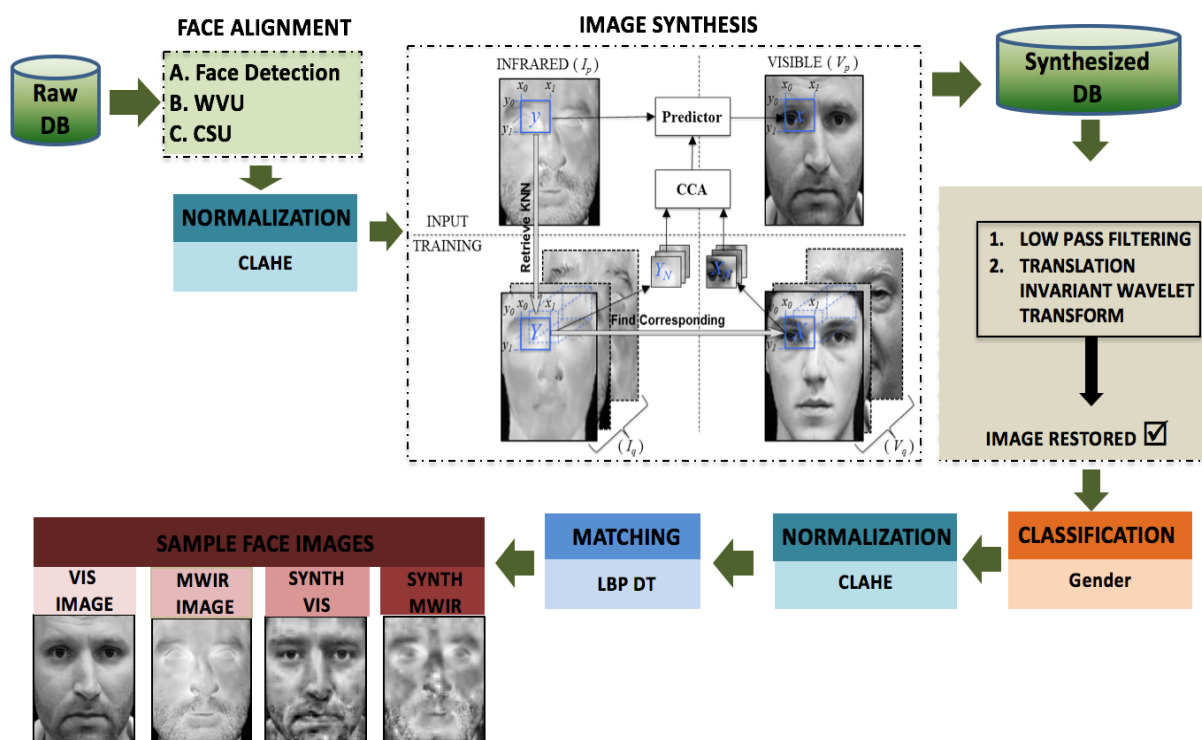
Figure 4.9:   Schematic of the heterogeneous overview methodology which consists of normalization, synthesis, restoration, denoising, and matching.

2) Thresholding-Based Nonlinear Denoising: When wavelets are used to deal with the problem of image denoising [126], the necessary steps involved are the following: 1) Apply discrete wavelet transform (DWT) to the noisy image by using a wavelet function (e.g., Daubuchies, Symlet, etc.). 2) Apply a thresholding estimator to the resulting coefficients thereby suppressing those coefficients smaller than a certain amplitude. 3) Reconstruct the denoised image from the estimated wavelet coefficients by applying the inverse discrete wavelet transform (IDWT). The idea of using a thresholding estimator for denoising was systematically explored for the first time in [127]. An important consideration here is the choice of the thresholding estimator and threshold value used since they impact the effectiveness of denoising. Different estimators exist that are based on different threshold value quantization methods, viz., hard, soft, or semisoft thresholding. Each estimator removes redundant coefficients using a nonlinear thresholding based on (4), where h is the noisy observation, ψ_m is the mother wavelet function, $m = (i, j)$ (2^i is the scale and j is the position of the wavelet basis), Ω is the thresholding estimator, q is the thresholding type, and T is the threshold used.

If x is an input signal, then the estimators used in this paper are defined based on (4.6) - (4.8), where mu is a parameter greater than 1, and the superscripts H, S, and SS denote hard, soft, and semisoft thresholding, respectively. In nonlinear thresholding-based denoising methods [see (4.5)], translation invariance means that the basis is translation invariant $\forall m, \forall r \in \Phi$, where Φ is a lattice of $\Re^d$ and $d = 2$ for an image signal.While the Fourier basis is translation invariant, the orthogonal wavelet basis ψ_m is not (in either the continuous or discrete settings)

$$\hat{h} = \sum_{|(h, \phi_m)| > T} (h, \psi_m)\psi_m = \sum_m \Omega_T^q((h, \psi_m))\psi_m \tag{4.5}$$

$$\Omega_T^H(x) = \begin{cases} x, & \text{if } |x| > T \\ 0, & \text{if } |x| \leq T \end{cases} \tag{4.6}$$

$$\Omega_T^S(x) = \begin{cases} \text{sgn}(x) \cdot (|x - T|), & \text{if } |x| > T \\ 0, & \text{if } |x| \leq T \end{cases} \tag{4.7}$$

$$\Omega_T^{SS}(x) = \begin{cases} 0, & \text{if } |x| \leq T \\ x, & \text{if } |x| > \mu T \\ \text{svn}(x) \cdot \frac{|x-T|}{\mu-1}, & \text{if } |x| > T, \text{otherwise.} \end{cases} \tag{4.8}$$

Image denoising using the traditional orthogonal wavelet transforms may result in visual artifacts. Some of these can be attributed to the lack of translation invariance of the wavelet basis. One method to suppress such artifacts is to average out the translation dependence, i.e., through cycle spinning as proposed by Coifman [128].

$$\Theta_{\text{TI}}(h) = \frac{1}{\Phi} \cdot \sum_{\tau \in \Phi} \Theta(h_\tau)_{-\tau} \tag{4.9}$$

where $\forall \tau \in \Phi, \Theta_{\text{TI}}(h) = \Theta_{\text{TI}}(h_r)_{-\tau}$, . This is called cycle spinning denoising. If we have an N-sample data, then pixel precision translation invariance is achieved by having N wavelet translation transforms (vectors) or $|\Phi| = N$. Similar to cycle spinning denoising, thresholding-based translation invariant denoising can be defined as

$$\Theta_{\text{TI}}(h) = \frac{1}{\Phi} \cdot \sum_{m,\tau \in \Phi} \omega_T^q((h, (\Psi_m)_\tau))(\Psi_m)_\tau. \tag{4.10}$$

4.9.2   Methodological Steps

The entire overview of this framework is illustrated in Fig. 4.9. The pertinent stages of the methodology proposed in this work are described below:

1. ***Pre-Processing:*** Our proposed approach is patch-based, therefore it is important that the correct corresponding patches overlap as precisely as possible in both spectra. We experiment with three different face image pre-processing techniques, all discussed in detail below. The metric we use for performance evaluation is rank-1 identification accuracy (CMC). The left and right eye coordinates are manually annotated on the raw images prior to pre-processing. Samples of the face images after pre-processing can be seen in Fig. 4.10.

   - **Face Detection:** For the visible spectrum of our database, Viola & Jones face detection algorithm [41] is used to determine the rectangular overlay or boundary around the face. This algorithm has been regarded to perform efficiently on facial images captured in the visible spectrum, but additional training is necessary for the passive IR band. However, there were still several limitations when Viola & Jones is applied to the passive IR band of our database, due to the lack of training data (not many available and the operational cost to collect more with both our cameras was prohibited). To compensate, blob detection based approach is applied in our passive infrared band images, resulting in 85% better detection accurary than Viola & Jones (whose haar cascades are trained specifically for visible data).

   - **CSU Normalization:** Colorado State University's (CSU) Face Identification Evaluation System [58] FR software is first utilized for pre-processing. The normalization is a spatial transformation, which utilizes the left and right eyes as
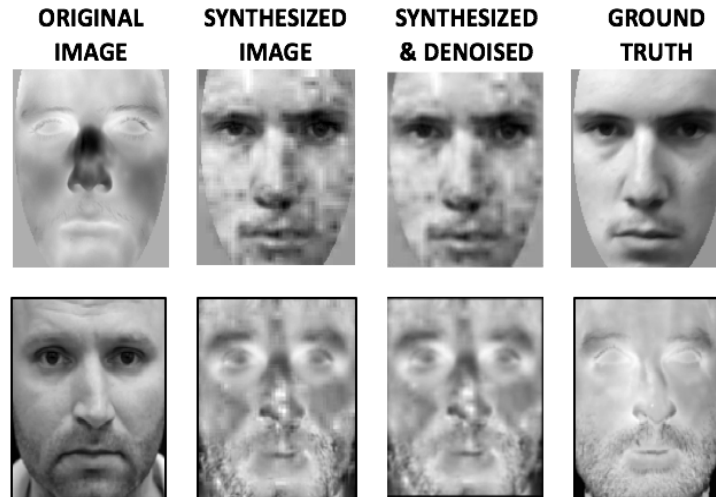
Figure 4.10: Example original, synthesized, synthesized and denoised and ground truth images from two separate subjects. The subject on top row (MWIR to VIS) was normalized using CSU normalization, while the subject on the bottom row (VIS to MWIR) was normalized using our proposed normalization technique.

control points. Shapes in the original image are unchanged, but the image is distorted by a combination of translation, rotation, and scaling. After geometric normalization, the image is cropped using an elliptical mask so that only the face from the forehead to the chin and cheek to cheek can be seen.

- **Normalization (Proposed):** A standard interocular distance is set and the eye locations are centered and aligned onto a single horizontal plane and resized to fit the desired distance. Each face image was geometrically normalized based on the manually found locations to have an interocular distance of 60 pixels with a resolution of 111×121 pixels. There is no elliptical mask applied in our approach, in contrast to the CSU normalization software.

2. *Image Synthesis & Denoising:* The formulated image synthesis methodology is

combination of manifold learning and non-linear dimensionality reduction. We utilize the leave one out method during synthesis, where the sample left out of the training set is used for conversion from one spectrum to another. Through the image synthesis algorithm, we are able to convert the datasets described and create their synthesized versions. After the synthesized data is created, it is later used for identity authentication. We restore the synthesized images from the previous step using a combination of linear denoising and thresholding. Noniterative denoising methods are a possible solution for the noise problem through numerical calculations that are explicitly solved. Noniterative methods are usually easier to implement and are not computationally complex.

3. **Face Recognition Systems:** We utilize the *Local Binary Patterns* (LBP) method [63] for FR due to its previous use and success with the cross-spectral face recognition problem [129]. The LBP operator is an efficient, nonparametric, and unifying approach to traditional divergent models for analyzing texture that are statistical and structural based. Occurrences of different binary patterns are then counted up using a histogram. Initially, the LBP method was introduced as a texture descriptor. By thresholding a 3×3 window based on the value of the center pixel, patterns in an image can be computed. Next, the result which is a binary pattern, is converted into a decimal value. The set of points that have been sampled and spaced evenly on a circle represent the local neighbor hood.

The LBP operator used in this work can be described as $LBP_{P,R}^{u^2}$ where P refers to the total sample points placed on a circle with radius R. The symbol u^2 takes into consideration the patterns in our experiments that are most frequently occurring, and represents the uniform pattern. The pattern is capable of characterizing local regions that contain corners as well as edges, which is what makes it vital. The binary pattern

for pixels, lying in a circle $f_p, p = 0, 1, ..., P - 1$ with f_c as the center pixel can be computed mathematically as follows:

$$S(f_p - f_c) = \begin{cases} 1, & \text{if } f_p - f_c \geq 0 \\ 0, & \text{if } f_p - f_c \leq 0 \end{cases} \qquad (4.11)$$

Following this a binomial weight 2^P is assigned to each sign $S(f_p - f_c)$ to compute the LBP code,

$$LBP_{P,R} = \sum_{p=0}^{P-1} S(f_p - f_c) 2^P \qquad (4.12)$$

LBP is not variant to monotonic gray-level transformations. It is worth mentioning that one downside to LBP is that it is usually sensitive to noise in image regions that are homogeneous since the center of the pixel region is thresholded to compute the binary code. Subsequently, *Local Tertiary Patterns* (LTP) method [63] has been introduced to overcome such a challenge, where the quantization can be performed as:

$$S(f_p - f_c) = \begin{cases} 1, & \text{if } f_p - f_c \geq t \\ 0, & \text{if } |f_p - f_c| \leq t \\ -1, & \text{if } f_p - f_c \leq -t \end{cases} \qquad (4.13)$$

A pattern that is 3-valued (tertiary), as opposed to a pattern that is binary, is the result of the LTP operator. In addition, different patterns can be produced through the optimization of the threshold t. The use of this threshold also makes the LTP code a bit more tolerable to noise. Two measures were used, i.e., chi-squared distance and distance transform [63], to measure the similarity between histogram features,. The

chi-squared distance can be mathematically defined as:

$$\chi^2(n, m) = \frac{1}{2} \sum_1^l \frac{h_n(k) - h_m(k)}{h_n(k) + h_m(k)}, \tag{4.14}$$

where h_n and h_m are the two histogram feature vectors, l represents the length of the vectorized feature, and n and m are sample vectors extracted from an image of the gallery and probe sets respectively. The distance transform (defined as the distance or similarity metric from image X to image Y) is equated as follows:

$$D(X, Y) = \sum_{Y(i,j)} w(d_X^{k_{Y(i,j)}}(i, j)), \tag{4.15}$$

where $k_{Y(i,j)}$ is the code value of pixel (i, j) of image Y, and w is a user-controlled penalty function [63]. The distance transform (DT) was established to be a method that is more consistent when aiming to achieve higher FR accuracy for the LBP/LTP. When comparing selected matchers (e.g. LBP vs LTP), LBP holds a slight edge over LTP in many scenarios, particularly in our proposed approach. The cumulative match characteristic (CMC) curve is used to measure the identification accuracy of the system. With this metric, the ranking potential of the system can be measured, showing the 1 : m identification performance.

4.9.3   Gender Classification

Gender-based cohort classification is achieved using an approach that detects HOG (Histogram of Oriented Gradient) features [130] and a SVM (Support Vector Machine) classifier [131].
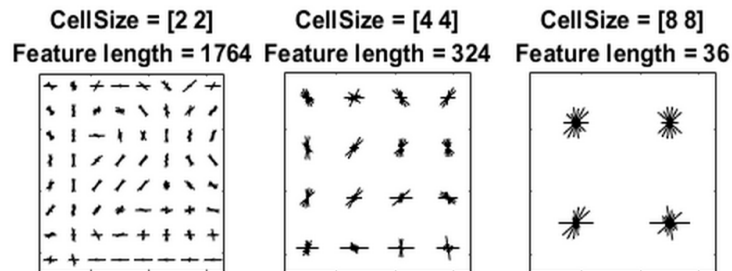
Figure 4.11: Trade off between cell size and feature length in encoded HOG feature vectors.

HOG Feature Extraction

In order to train the classifiers, it is important that vectoized HOG features are extracted from the images used for training. The extracted HOG feature that is vectorized should be capable of encoding an precise amount of information pertaining to the subject. However, there is a trade-off betwen cell size and the feature length in the vectorized HOG feature. Pertaining to shape, Fig. 4.11 shows that a cell size of 8×8 is larger, but a lot of inormation is lost, when compared to a cell size of 2×2. A safe choice for many applications is a cell size of 4×4. Ideally, the HOG parameters can be varied while iterating classifier training and testing for the best settings.

SVM Classification

SVM is a kernel based method and has mostly been used for two class classification [131]. Through the use of non-linear mapping, kernel algorithms are capable of mapping data from a original space into a higher dimensional feature space. The downside of this is that in high dimensional spaces, the curse of dimensionality is evident, although there exists a workaround for finding the scalar products in the feature space. When considering two feature space vectors, calculation of the scalar product can be done explicitly with the help of kernel functions. Originally, the SVM method was well known as a linear classifier [132]. As time went on, there was modification in order to allow data to be non-linearly mapped
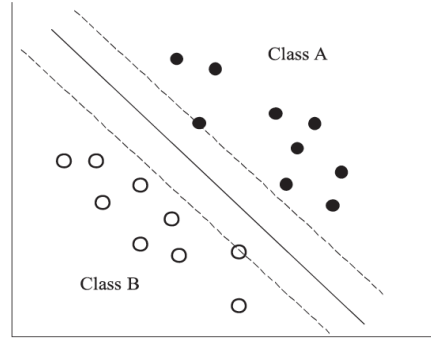
Figure 4.12: Example of separation that occurs between two different classes using SVM.

into the feature space, with use of a kernel-based method. In Fig. 4.12, an example of the separation that occurs between data with SVM can be seen.

4.10   Empirical Evaluation

The experimental scenarios we evaluate in this work are as follows: 1) baseline experiments; 2) optimization of image synthesis; 3) post synthesis image denoising w.r.t. FR accuracy; 4) automatic classification experiments; and 5) identification performance after gender-based filtering. After optimizing our selected matcher for the given problem (e.g. LBP/LTP), the distance transform (DT) appears to be a more consistent method in achieving higher FR accuracy. When comparing selected matchers (e.g. LBP vs LTP), LBP holds a slight edge over LTP in many scenarios. For our selected texture based matcher (e.g., LBP DT), we evaluate the challenge of image alignment using varied pre-processing within our proposed synthesis approach during experimentation. We trained our synthesis and classification algorithms using a Leave-One-Out approach, i.e. take one image sample out of the training dataset and use that sample as test image for synthesis (the IR image as the input and the VIS image as the ground truth, and vice-versa); the remaining samples of the subject are used for training within our system.

4.10.1   Baseline FR Experiments

We employ a set of baseline experiments (cross-spectral face matching) by using commercial and academic based software: 1) Commercial-of-the-shelf (COTS) identity software tools (G8) provided by MorphoTrust (formerly L1) ; 2) Face Identification Evaluation System which contains standard training-based face recognition methods developed by the CSU [58], including PCA ([49],[50], [85]), PCA+LDA [48], the *Bayesian Intrapersonal/Extra-personal Classifier* (BIC) using either the Maximum likelihood (ML) or the Maximum *a posteriori* (MAP) hypothesis [86]. Distance metrics such as Euclidean Distance (EU) are used by both PCA and LDA, which result in the ordinary or standard distance between two feature vectors (PCA EU + LDA EU); 3) *Local Binary Pattern* (LBP) method [87], as aforementioned in the methodological steps for our face recognition pipeline.

Utilizing a commercial matcher (G8), the rank-1 identification rate achieved is 40.26% for the WVU VIS-MWIR dataset and 62.82% for the WVU VIS-LWIR dataset. For CSU academic matcher, the maximum rank-1 identification rate recorded is 19.23% for WVU VIS-LWIR and 8.97% for WVU VIS-MWIR using the LDA algorithm. For our NVESD dataset, the Bayesian algorithm had the best results with a rank-1 identification rate of 38.00% for NVESD VIS-LWIR and 20.00% for NVESD VIS-MWIR combination. With regard to the LBP DT matcher, the rank-1 identification rate achieved is 5.29% for the WVU VIS-MWIR spectrum. For the WVU VIS-LWIR spectrum, the rank-1 identification rate achieved for the LBP DT matcher is 5.19%. When evaluating the LBP DT matcher on the NVESD VIS-MWIR dataset, a rank-1 of 20.00% was achieved. For the LBP DT matcher on the NVESD VIS-LWIR dataset, a rank-1 of 14.00% was recorded. The baseline results using COTS G8 for the WVU dataset is shown in Table 4.2. We are unable to provide results for COTS G8 algorithm with the NVESD dataset due to an expiring license. For training-based academic matcher CSU, the baseline results for both the WVU and NVESD datasets can be seen in

Table 4.2: Baseline Rank-1 FR results (%) for VIS-MWIR and VIS-LWIR Face Matching Experiments.

WVU Rank-1 Raw Baseline (CSU) FR Accuracy (%)				
Methodology	VIS-LWIR	LWIR-VIS	VIS-MWIR	MWIR-VIS
L1 Systems (G8)	**62.82**	61.54	40.26	37.01
Bayesian MAP	**11.54**	**11.54**	7.69	5.13
Bayesian ML	**11.54**	10.26	7.69	5.13
LDA Euclidean	11.54	**19.23**	7.69	8.97
LDA IdaSoft	14.10	**19.23**	8.97	6.41
PCA Euclidean	**8.97**	5.13	7.69	6.41
PCA MahCosine	**15.38**	**15.38**	5.13	7.69

Table 4.3: NVESD Dataset Baseline Rank-1 FR results (%) for VIS-MWIR and VIS-LWIR CSU Face Matching Experiments.

NVESD Rank-1 Raw Baseline (CSU) FR Accuracy (%)				
Methodology	VIS-LWIR	LWIR-VIS	VIS-MWIR	MWIR-VIS
Bayesian MAP	**38.00**	10.00	20.00	10.00
Bayesian ML	**32.00**	12.00	18.00	12.00
LDA Euclidean	**32.00**	30.00	14.00	10.00
LDA IdaSoft	**32.00**	30.00	16.00	16.00
PCA Euclidean	**20.00**	14.00	4.00	4.00
PCA MahCosine	**32.00**	28.00	16.00	16.00

Tables 4.2 and 4.3 repsectively. The baseline results using texture based LBP DT matcher can be seen in Tables 4.4, and 4.5 for both datasets.

4.10.2   FR Experiments after Image Synthesis

There are several parameters to be chosen in our proposed synthesis algorithm, such as the size of patches, the number of canonical variates k (the dimensionality of feature vector)

Table 4.4: WVU Dataset Baseline Rank-1 FR results (%) for Pre-Processed VIS-MWIR and VIS-LWIR Face Matching Experiments (LBP DT).

		Baseline Rank-1 FR Accuracy (%)		
Gallery	Probe	Face Detect	WVU	CSU
VIS	MWIR	4.55	3.90	**5.19**
MWIR	VIS	**3.90**	**3.90**	1.95
VIS	LWIR	2.56	**4.49**	**4.49**
LWIR	VIS	3.85	**5.13**	1.92

Table 4.5: NVESD Dataset Baseline Rank-1 FR results (%) for Pre-Processed VIS-MWIR and VIS-LWIR Face Matching Experiments (LBP DT).

		Baseline Rank-1 FR Accuracy (%)		
Gallery	Probe	Face Detect	WVU	CSU
VIS	MWIR	6.00	**12.00**	6.00
MWIR	VIS	16.00	**20.00**	**20.00**
VIS	LWIR	4.00	**8.00**	6.00
LWIR	VIS	14.00	14.00	**16.00**

we take for every patch, and the number of the neighbors we use to train the canonical directions. Generally speaking, the correlation between pairs of IR and VIS patches of a smaller size is weaker, so the inference is less reasonable. While the larger the patch size, makes the correlation stronger, more canonical variates are needed to represent the patch, which makes training samples much sparser in the feature space. The size of all the images in our database despite pre-processing methodology is, 320×256 during the synthesis step, and we choose a patch size of 9×9 with 3-px overlapping. Since the projections (features) onto the former pairs of direction have stronger correlations, choosing fewer features makes the inference more robust, while choosing more features gives a more precise adaptation of the original patch. Similarly, when we choose a larger number of neighbors, K, there are more samples, which makes the algorithm more robust but computationally expensive. We choose 5 features and 100 neighbors for LLE. Once we have converted a spectrum from VIS to MWIR, MWIR to VIS, VIS to LWIR, or LWIR to VIS, our heterogeneous cross-spectral matching problem can now be considered to be an homogenous intra-spectral matching

problem again. CLAHE normalization is applied to both sets of gallery and probe images after synthesis and prior to matching. Although not practical, our matching experiments after synthesis are tested using the synthesized images as both gallery and probe set, for each spectrum.

With respect to the WVU dataset, we achieve a maximum rank-1 identification rate of 85.06% when using LBP DT matcher after synthesis for WVU VIS-MWIR and a maximum rank-1 identification rate of 79.49% for the WVU VIS-LWIR spectrums, using WVU normalization. For the NVESD dataset, we achieve a maximum rank-1 identification rate of 98.00% when using LBP DT matcher after synthesis for NVESD VIS-MWIR and a rank-1 identification rate of 100.00% for the NVESD VIS-LWIR spectrums, using CSU normalization. The synthesis results can be seen in Tables 4.6 and 4.7 for the synthesized WVU and NVESD datasets respectively.

We perform verification experiments on both datasets for practical scenarios (e.g. where the gallery is not synthesized). For the WVU dataset, we find that VIS-Synth. VIS (LWIR) EER is 36%, the MWIR-Synth. MWIR (VIS) EER is 40%, the LWIR-Synth. LWIR (VIS) EER is 41%, and the VIS-Synth. VIS (MWIR) EER is 42%. For the NVESD dataset, we find that VIS-Synth. VIS (LWIR) EER is 20%, the MWIR-Synth. MWIR (VIS) EER is 29%, the LWIR-Synth. LWIR (VIS) EER is 26%, and the VIS-Synth. VIS (MWIR) EER is 20%. Much like the identification experiments we performed, we find that NVESD dataset performs better than the WVU dataset given the suggested experimental protocol. ROC curves for the verification experiments can be seen in Fig. 4.13. Experiments suggest that, in some scenarios, a matcher resulting in good rank-1 identification accuracy (closed-set) may not perform well in a verification scenario [133].

Table 4.6: Rank-1 FR results (%) for Synthesized WVU VIS-MWIR and VIS-LWIR datasets using selected matcher (LBP DT).

WVU Synthesized Rank-1 FR Accuracy (%) LBP DT				
Gallery	Probe	Face Detect	WVU	CSU
VIS	Synth. VIS (MWIR)	19.48	**31.17**	27.27
Synth. VIS (MWIR)	VIS	68.18	**76.62**	72.73
MWIR	Synth. MWIR	24.68	**35.71**	26.62
Synth. MWIR	MWIR	58.44	**85.06**	76.62
VIS	Synth. VIS (LWIR)	8.97	**40.38**	27.56
Synth. VIS (LWIR)	VIS	53.85	**85.26**	66.67
LWIR	Synth. LWIR	26.28	38.46	**45.51**
Synth. LWIR	LWIR	56.41	**79.49**	77.56

Table 4.7: Rank-1 FR results (%) for Synthesized NVESD VIS-MWIR and VIS-LWIR datasets using selected matcher (LBP DT).

NVESD Synthesized Rank-1 FR Accuracy (%) LBP DT				
Gallery	Probe	Face Detect	WVU	CSU
VIS	Synth. VIS (MWIR)	50.00	68.00	**74.00**
Synth. VIS (MWIR)	VIS	**86.00**	**86.00**	84.00
MWIR	Synth. MWIR	62.00	56.00	**88.00**
Synth. MWIR	MWIR	90.00	94.00	**98.00**
VIS	Synth. VIS (LWIR)	54.00	68.00	**76.00**
Synth. VIS (LWIR)	VIS	**86.00**	84.00	78.00
LWIR	Synth. LWIR	**80.00**	60.00	39.00
Synth. LWIR	LWIR	90.00	96.00	**100.00**

(a)   **(b)**
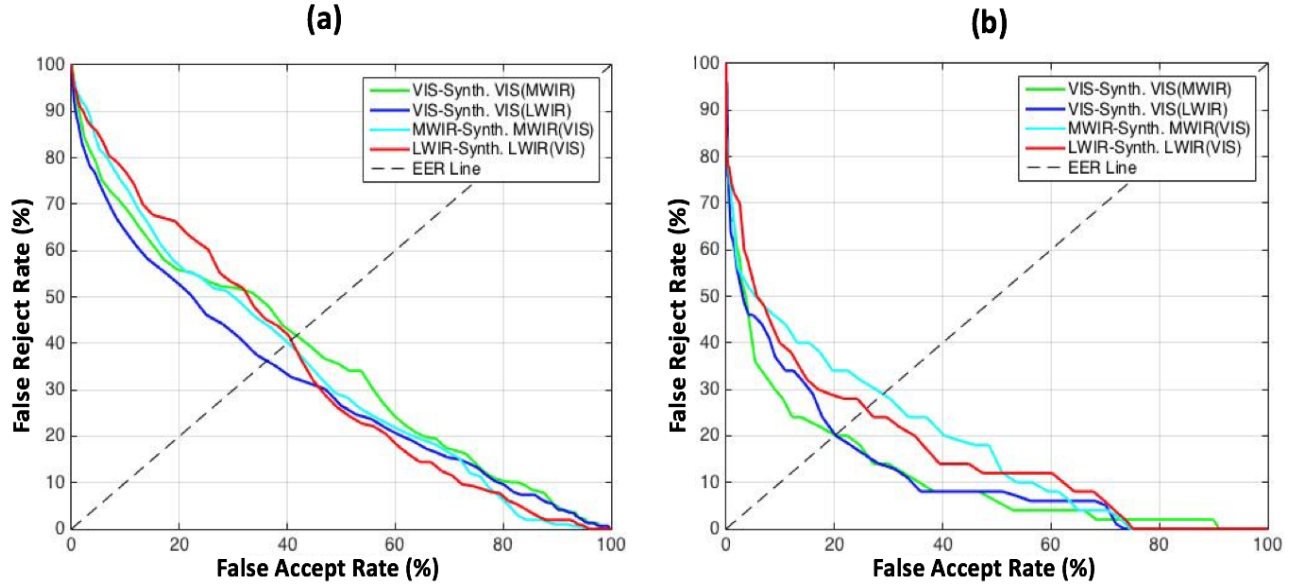
Figure 4.13: ROC curves comparing proposed methodology (after synthesis) for practical scenarios, (e.g. where the gallery is not synthesized) for: a) WVU dataset; (b) NVESD datset.

4.10.3   FR Experiments after Face Image Denoising

In this evaluation, we determine the effects of applying a combination of filtering and TI-denoising on synthesized images in order to improve FR accuracy of our datasets under practical scenarios (e.g. gallery images are not synthesized). Both synthesized and ground truth (gallery and/or probe) sets were LP filtered and subsequently denoised. We optimize our proposed image denoising parameters, LP filter type and sigma value threshold used for TI-denosing, using CMC rank-1 accuracy as a metric. First, we apply an LP Filter to minimize distortion due to the subsampling. The type of LP filter used is a boxcar filter with a fixed window size. Through previous experimentation [2], we found a window size of 3 to be optimal. After we are able to LP filter the image, denoising is carried out using the TI-denosing scheme. The sigma value chosen for TI-denoising appears to be optimal depending on whether we are denoising synthesized images or ground truth images. Synthesized images received a sigma value of 3, while ground truth images were only slightly denoised with a

Table 4.8: Rank-1 FR results (%) for Denoised Synthesized WVU VIS-MWIR and VIS-LWIR datasets using selected matcher (LBP DT).

WVU Denoised Synthesized Rank-1 FR Accuracy (%) LBP DT				
Gallery	Probe	Face Detect	WVU	CSU
VIS	Synth. VIS (MWIR)	21.43	30.52	**31.17**
Synth. VIS (MWIR)	VIS	68.83	75.32	**77.27**
MWIR	Synth. MWIR	27.27	42.86	**51.30**
Synth. MWIR	MWIR	59.74	**81.17**	76.62
VIS	Synth. VIS (LWIR)	5.13	**37.82**	23.72
Synth. VIS (LWIR)	VIS	53.85	**81.41**	79.49
LWIR	Synth. LWIR	25.00	43.59	**57.69**
Synth. LWIR	LWIR	60.90	78.85	**80.77**

sigma value of .01.

With respect to the WVU dataset, we achieve a maximum rank-1 identification rate of 81.17% when using LBP DT matcher after synthesis for WVU VIS-MWIR and a rank-1 identification rate of 80.77% for the WVU VIS-LWIR spectrums, using WVU and CSU normalization respectively. For the NVESD dataset, we achieve a maximum rank-1 identification rate of 98.00% when using LBP DT matcher for NVESD VIS-MWIR and a rank-1 identification rate of 96.00% for the NVESD VIS-LWIR spectrums, using CSU normalization.

The results for FR of synthesized images after image denoising can be seen in Tables 4.8 and 4.9 for the restored WVU and NVESD datasets respectively. Identification rates (Rank-1 to Rank-5) can be seen for our collected data and proposed methodology (after denoising) when compared to classic academic matchers, in Figs. 4.15 , 4.16, 4.18 and 4.17 respectively.

Table 4.9: Rank-1 FR results (%) for Denoised Synthesized NVESD VIS-MWIR and VIS-LWIR datasets using selected matcher (LBP DT).

NVESD Denoised Synthesized Rank-1 FR Accuracy (%) LBP DT				
Gallery	Probe	Face Detect	WVU	CSU
VIS	Synth. VIS (MWIR)	66.00	70.00	**78.00**
Synth. VIS (MWIR)	VIS	86.00	86.00	**92.00**
MWIR	Synth. MWIR	74.00	67.00	**88.00**
Synth. MWIR	MWIR	90.00	92.00	**98.00**
VIS	Synth. VIS (LWIR)	**88.00**	71.00	76.00
Synth. VIS (LWIR)	VIS	88.00	82.00	**90.00**
LWIR	Synth. LWIR	82.00	66.00	**88.00**
Synth. LWIR	LWIR	90.00	92.00	**96.00**

Table 4.10: SVM gender classification accuracy for NVESD dataset preprocessed using face detection and trained using leave-one-sample out method.

NVESD (Face Detection) Gender Classification Accuracy (%) using SVM				
	Train: WVU MWIR		Train: WVU LWIR	
	Male	Female	Male	Female
Male	**100.00**	0.00	**100.00**	0.00
Female	0.00	**100.00**	0.00	**100.00**

4.10.4   FR Experiments after Gender Classification

We achieve gender-based cohort classification scheme using HOG (Histogram of Oriented Gradient) features [130] and a SVM (Support Vector Machine) classifier [131]. The cell size that we use to encode the vectorized HOG feature is 4×4. This offers a compromise between a cell size of 8×8 and cell size of 2×2. We test both datasets using the training scheme detailed in Empirical Evaluation section. When training using the leave-one-out scheme, we achieve perfect performance across all pre-processing methods and tested datasets. Tables showing classification accuracy per gender, per dataset and training scenario can be seen in Tables 4.10, 4.11, 4.12, 4.13, 4.14, and 4.15.

Table 4.11: SVM gender classification accuracy for WVU dataset preprocessed using face detection and trained using leave-one-sample out method.

WVU (Face Detection) Gender Classification Accuracy (%) using SVM				
	Train: NVESD MWIR		Train: NVESD LWIR	
	Male	Female	Male	Female
Male	**100.00**	0.00	**100.00**	0.00
Female	0.00	**100.00**	0.00	**100.00**

Table 4.12: SVM gender classification accuracy for NVESD dataset preprocessed using WVU face normalization and trained using leave-one-sample out method.

NVESD (WVU) Gender Classification Accuracy (%) using SVM				
	Train: WVU MWIR		Train: WVU LWIR	
	Male	Female	Male	Female
Male	**100.00**	0.00	**100.00**	0.00
Female	0.00	**100.00**	0.00	**100.00**

4.10.5   FR Experiments after Demographic Filtering

The face datasets are filtered by gender and the resulting system performance is evaluated. The objective is to determine whether the hypothesis that face matching performance improves with a filtering hold, such as gender-based classification. Two different gender subsets,

Table 4.13: SVM gender classification accuracy for WVU dataset preprocessed using WVU face normalization and trained using leave-one-sample out method.

WVU (WVU) Gender Classification Accuracy (%) using SVM				
	Train: NVESD MWIR		Train: NVESD LWIR	
	Male	Female	Male	Female
Male	**100.00**	0.00	**100.00**	0.00
Female	0.00	**100.00**	0.00	**100.00**

Table 4.14: SVM gender classification accuracy for NVESD dataset preprocessed using CSU face normalization and trained using leave-one-sample out method.

NVESD (CSU) Gender Classification Accuracy (%) using SVM				
	Train: WVU MWIR		Train: WVU LWIR	
	Male	Female	Male	Female
Male	**100.00**	0.00	**100.00**	0.00
Female	0.00	**100.00**	0.00	**100.00**

Table 4.15: SVM gender classification accuracy for WVU dataset preprocessed using CSU face normalization and trained using leave-one-sample out method.

WVU (CSU) Gender Classification Accuracy (%) using SVM				
	Train: NVESD MWIR		Train: NVESD LWIR	
	Male	Female	Male	Female
Male	**100.00**	0.00	**100.00**	0.00
Female	0.00	**100.00**	0.00	**100.00**

one for each dataset, were used for this experiment. For the gender-based subsets we split the data into (1) male and (2) female gallery and probe. With respect to the WVU VIS-MWIR gender-based subsets, 52 subjects were male, while 25 subjects were female. For the WVU VIS-LWIR gender-based subsets, 57 subjects were male, while 21 subjects were female. For both the NVESD VIS-MWIR and VIS-LWIR gender-based subsets, 35 subjects were male, while 15 subjects were female. A pie chart representing gender-based distribution for both datasets can be seen in Fig. 4.14.

With respect to the WVU dataset, we achieve a maximum rank-1 identification rate of 88.96% when using LBP DT matcher after synthesis for WVU VIS-MWIR and a rank-1 identification rate of 90.38% for the WVU VIS-LWIR spectrum, using WVU normalization. For the NVESD dataset, we achieve a maximum rank-1 identification rate of 98.00% when using LBP DT matcher and WVU pre-processing, after synthesis for NVESD VIS-MWIR and a rank-1 identification rate of 96.00% for the NVESD VIS-LWIR spectrum, using WVU and CSU normalization respectively.

The results for FR of the denoised synthesized images after image denoising and demo-
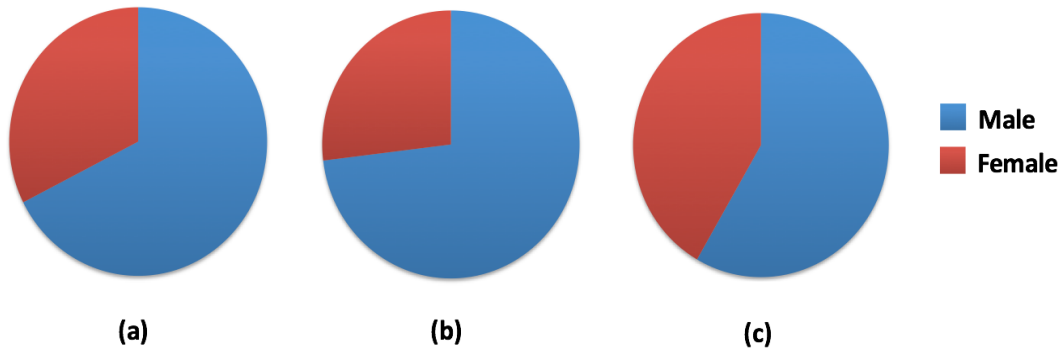
Figure 4.14: Gender-based filtering of dataset for (a) WVU VIS-MWIR, (b) WVU VIS-LWIR and (c) NVESD VIS-IR. NVESD VIS-MWIR and NVESD VIS-LWIR both have the same amount of subjects (50) and male-female proportion.

Table 4.16: Rank-1 FR results (%) for Cohort Filtered and Denoised Synthesized WVU VIS-MWIR and WVU VIS-LWIR datasets using selected matcher (LBP DT).

WVU Cohort Filtered & Denoised Synthesized Rank-1 FR Accuracy (%) LBP DT				
Gallery	Probe	Face Detect	WVU	CSU
VIS	Synth. VIS (MWIR)	42.51	51.30	**54.54**
Synth. VIS (MWIR)	VIS	84.41	83.77	**88.96**
MWIR	Synth. MWIR	41.56	68.18	**68.83**
Synth. MWIR	MWIR	76.62	**88.96**	85.71
VIS	Synth. VIS (LWIR)	22.44	**49.36**	41.67
Synth. VIS (LWIR)	VIS	70.51	**90.38**	88.46
LWIR	Synth. LWIR	43.59	**76.28**	73.72
Synth. LWIR	LWIR	78.85	**90.38**	89.10

graphic filtering can be seen in Tables 4.16 and 4.17 for the synthesized WVU and NVESD datasets respectively.

Table 4.17: Rank-1 FR results (%) for Cohort Filtered and Denoised Synthesized NVESD VIS-MWIR and NVESD VIS-LWIR datasets using selected matcher (LBP DT).

NVESD Cohort Filtered & Denoised Synthesized Rank-1 FR Accuracy (%) LBP DT				
Gallery	Probe	Face Detect	WVU	CSU
VIS	Synth. VIS (MWIR)	66.00	75.00	**78.00**
Synth. VIS (MWIR)	VIS	86.00	86.00	**92.00**
MWIR	Synth. MWIR	73.00	71.00	**96.00**
Synth. MWIR	MWIR	90.00	94.00	**98.00**
VIS	Synth. VIS (LWIR)	63.00	73.00	**76.00**
Synth. VIS (LWIR)	VIS	88.00	84.00	**94.00**
LWIR	Synth. LWIR	82.00	73.00	**90.00**
Synth. LWIR	LWIR	90.00	**98.00**	96.00
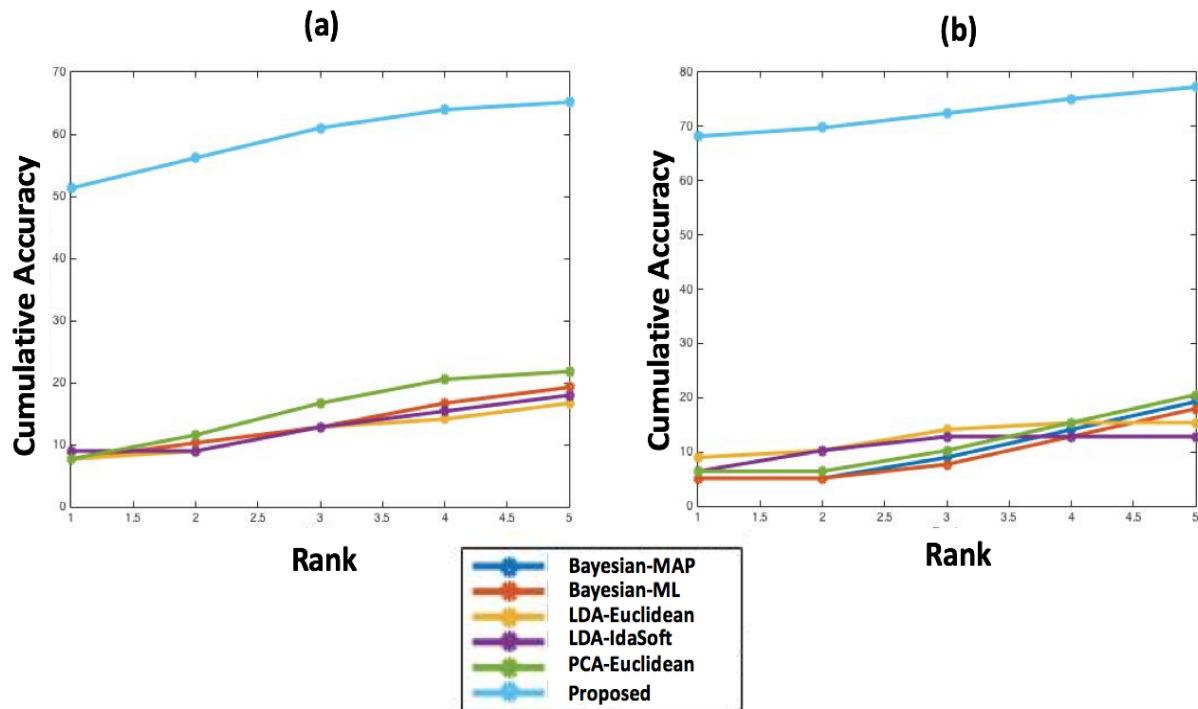
Figure 4.15: (a) Identification rates (Rank-1 to Rank-5) for WVU VIS gallery and MWIR probe. (b) Identification rates (Rank-1 to Rank-5) for WVU MWIR gallery and VIS probe.
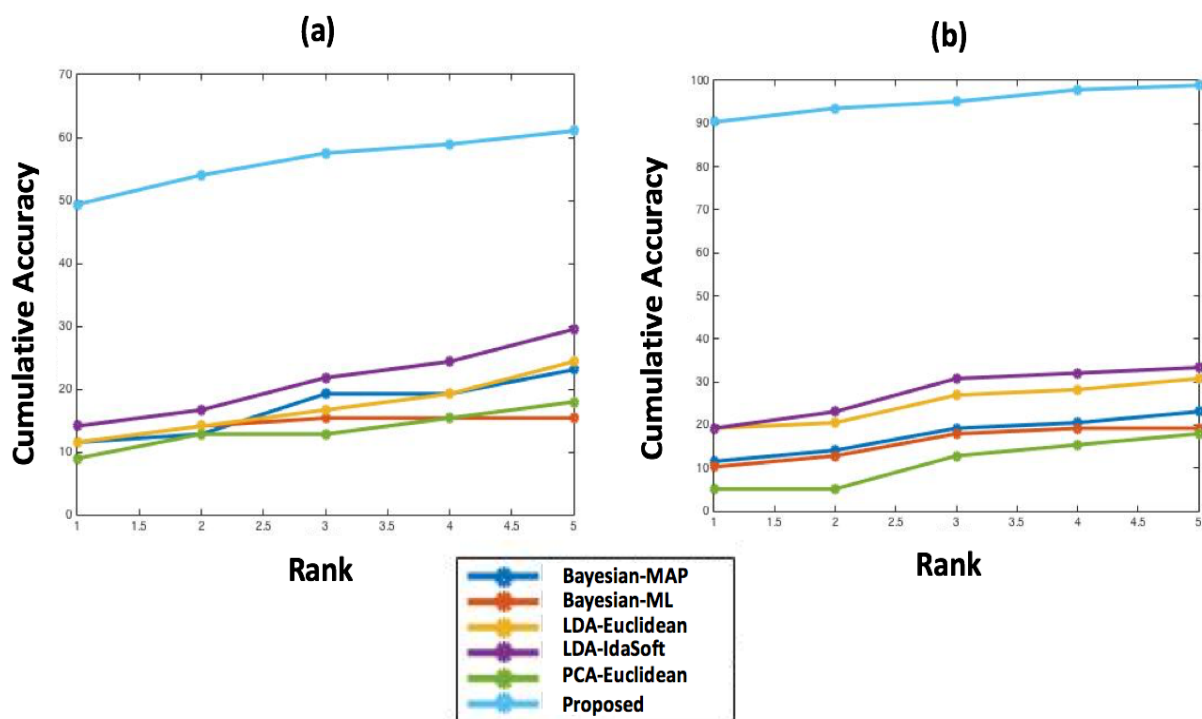
Figure 4.16: (a) Identification rates (Rank-1 to Rank-5) for WVU VIS gallery and LWIR probe. (b) Identification rates (Rank-1 to Rank-5) for WVU LWIR gallery and VIS probe.
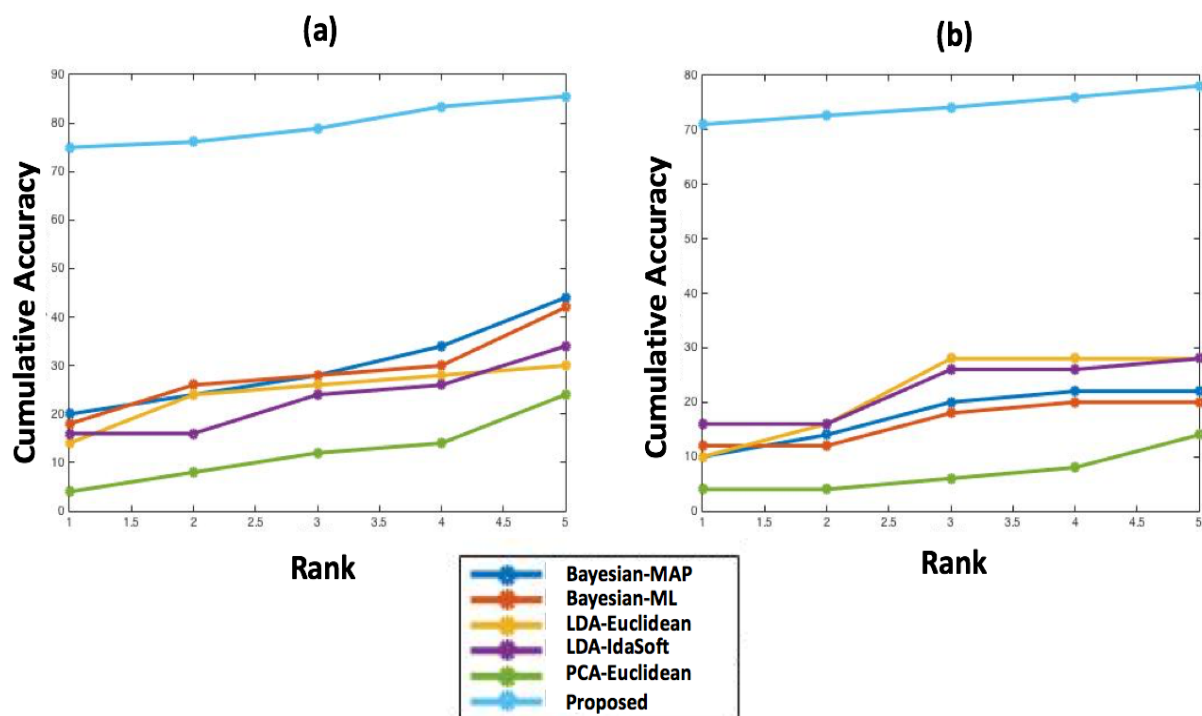
Figure 4.17: (a) Identification rates (Rank-1 to Rank-5) for NVESD VIS gallery and MWIR probe. (b) Identification rates (Rank-1 to Rank-5) for NVESD MWIR gallery and VIS probe.
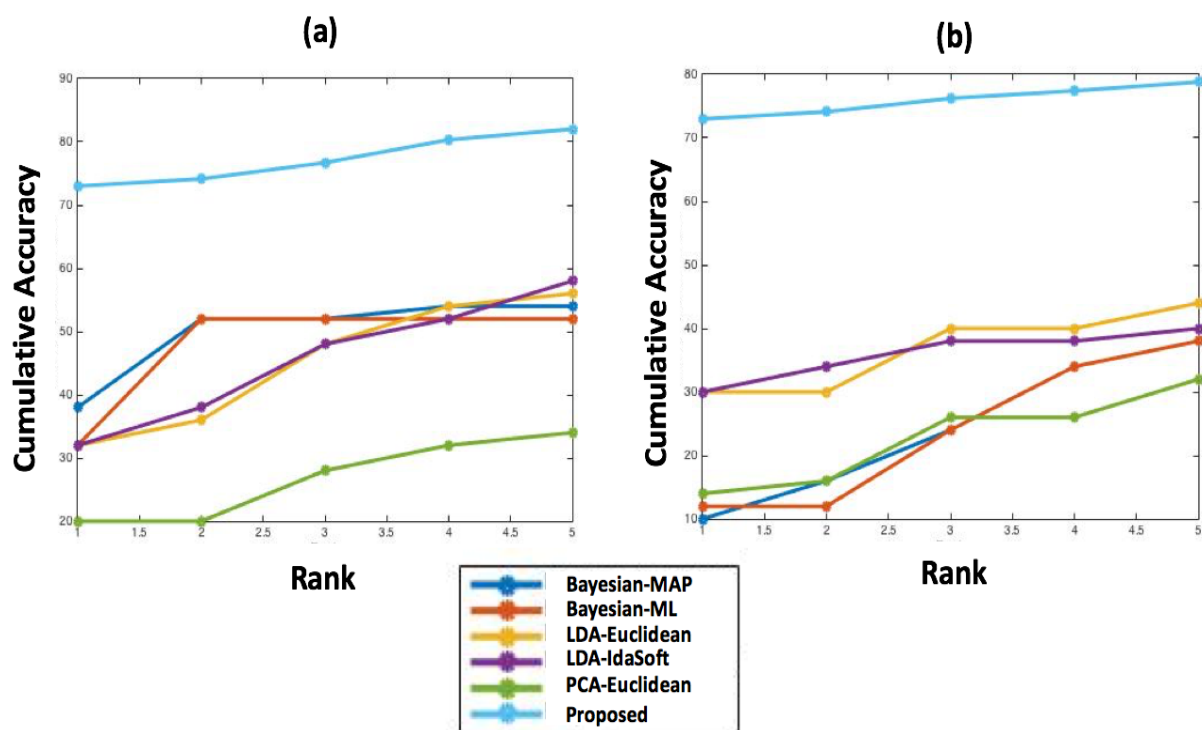
Figure 4.18: (a) Identification rates (Rank-1 to Rank-5) for NVESD VIS gallery and LWIR probe. (b) Identification rates (Rank-1 to Rank-5) for NVESD LWIR gallery and VIS probe.

4.11   Chapter Summary

We study the problem of image synthesis as a means to bridge the informational gap between face images pertaining to two different spectral bands. Our study shows that image alignment is important in achieving higher FR accuracy for the proposed approach. Experimental results show that recognition accuracy is much higher when the synthesized face image is used as a gallery set, as opposed to the probe set. We believe there is so much difference in rank-1 scores when the synthetic dataset is the gallery vs. the probe because there is more data present in the raw image. When a synthesized face image is used as the gallery, all information in the synthesized face image should be present in the raw face image of the same subject. However, when the raw face image is the gallery, the synthesized face probe image is likely missing some information that is present in the raw face gallery image. In practical applications, the use of raw face images as the gallery set would be the more realistic scenario. The image denoising step increases the score where we use the slightly denoised raw images as the gallery set, irrespective of the spectral band. However, rank-1 accuracy is decreased when a denoised synthesized image is used as the gallery image. The image denoising step was particularly valuable on the datasets pre-processed using face detection and CSU normalization, excluding matching using VIS gallery and synthesized VIS probe. The image denoising step decreases face recognition accuracy for our proposed geometric normalization pre-processing step. FR software which may be COTS, is difficult to evaluate because it may contain proprietary geometric and photometric normalization, in conjunction with restoration which are unable to be accounted for by the user. This is challenging because when using COTS packages, in addition to ours, a performance drop can be expected when compared to using our proposed approach.

Utilizing our image synthesis approach, we achieve a maximum rank-1 identification rate of 85.06% when using LBP DT matcher after synthesis for WVU VIS-MWIR and a max-

imum rank-1 identification rate of 79.49% for the WVU VIS-LWIR spectrum, using WVU normalization. For the NVESD dataset, we achieve a maximum rank-1 identification rate of 98.00% when using LBP DT matcher after synthesis for NVESD VIS-MWIR and a rank-1 identification rate of 100.00% for the NVESD VIS-LWIR spectrum, using CSU normalization. After denoising of our data, we achieve a maximum rank-1 identification rate of 81.17% when using LBP DT matcher after synthesis for WVU VIS-MWIR and a maximum rank-1 identification rate of 80.77% for the WVU VIS-LWIR spectrum, using WVU and CSU normalization respectively. For the NVESD dataset, we achieve a maximum rank-1 identification rate of 98.00% when using LBP DT for NVESD VIS-MWIR and a rank-1 identification rate of 96.00% for the NVESD VIS-LWIR spectrum, using CSU normalization. I think it is an important to note that although overall identification rate deceases in scenarios where there is a synthesized gallery set, identification accuracy improves in modes of operation that mimic reality (e.g. when the gallery set is not synthesized). After gender filtering of our synthesized and denoised data, we achieve a maximum rank-1 identification rate of 88.96% when using LBP DT matcher after synthesis for WVU VIS-MWIR and a rank-1 identification rate of 90.38% for the WVU VIS-LWIR spectrum, using WVU normalization. For the NVESD dataset, we achieve a maximum rank-1 identification rate of 98.00% when using LBP DT matcher and WVU pre-processing, after synthesis for NVESD VIS-MWIR and a rank-1 identification rate of 96.00% for the NVESD VIS-LWIR spectrum, using WVU and CSU normalization respectively. Overall, the assembled WVU dataset appears to be more challenging, perhaps because the images were not co-registered while being captured.

Chapter 5

Conclusions

In this dissertation we proposed a framework, for both homogeneous and heterogeneous FR systems using multi-spectral sensors. For homogeneous FR systems, we formulate and develop an efficient, semi-automated, direct matching-based FR framework, that is designed to operate efficiently when face data is captured using either visible or passive IR sensors. Thus, it can be applied in both daytime and nighttime environments. First, input face images are geometrically normalized using our pre-processing pipeline prior to feature-extraction. Then, face-based features including wrinkles, veins, as well as edges of facial characteristics, are detected and extracted for each operational band (visible, MWIR, and LWIR). Finally, global and local face-based matching is applied, before fusion is performed at the score level. Although this proposed matcher performs well when same-spectrum FR is performed, irregardless of spectrum, a challenge exists when cross-spectral FR matching is performed. The second framework aims to solve the heterogeneous, cross-spectral FR problem, and deals with the issue of bridging the gap across the visible and passive infrared (MWIR and LWIR) spectrums. Specifically, we investigate the benefits and limitations of using synthesized visible face images from thermal and vice versa, in cross-spectral face recognition systems when utilizing canonical correlation analysis (CCA) and manifold learning dimensionality

reduction (LLE). Finally, by conducting an extensive experimental study we establish that the combination of the proposed synthesis and demographic filtering scheme increases system performance in terms of rank-1 identification rate.

To reiterate, this dissertation provides the following contributions:

5.1   Contributions

This dissertation details the following set of novel contributions to the science:

1. The proposal of an efficient and automated homogeneous FR framework which is applied on full-frontal face images acquired using visible, MWIR, and LWIR sensors. The pre-processing pipeline and feature extraction methodology ensure that no artificial information is introduced into the face images. The pre-processing pipeline is instrumental in standardizing the face images across our dataset, which is necessary prior to feature extraction and subsequent experimental evaluation.

2. Designed experiments to quantitatively evaluate the benefits of global vs. local based matchers, specifically the performance of such matchers individually, and after fusion at the score level. For the global matcher, the best performance was achieved in the MWIR spectrum with a 99.47% rank-1 accuracy. For the local matcher, the best performance was achieved in the visible spectrum with a 98.95% rank-1 accuracy. When matchers are fused, a rank-1 accuracy of at least 99.43% is achieved across all spectrums. Experiments were also designed to evaluate the robustness of the proposed approach with the challenging problem of facial occlusion. Partial face matching is simulated through the partitioning of the whole face into regions. A very interesting result from this work was in the case where our proposed methodology was applied to the disjointed face sub-region (i.e. the region of the face excluding the eyes, nose, and

chin regions) in the MWIR spectrum. This suggested disjoint face features are just as important as those located near or around eyes, nose, mouth, and can be matched with high accuracy as well.

3. Formulation of a framework for the heterogeneous FR challenge, when matching visible face images to thermal face images, and vice versa. In constructing our cross-spectral face recognition system, the benefits and limitations of using canonical correlation analysis (CCA) and manifold learning dimensionality reduction (LLE) are investigated. Experimental results show that recognition accuracy is much higher when the synthesized face image is used as a gallery set, as opposed to the probe set.

4. Demonstration of challenges associated with face matching in cross-spectral condition, including utilization of three different methods for pre-processing face images before feature extraction methods are applied. Cohort-specific matching (per gender) instead of blind-based matching (when all images in the gallery are matched against all in the probe set) is explored, in order to increase face recognition accuracy. An extensive experimental study is conducted to establish that the combination of the proposed synthesis, restoration, and demographic filtering scheme increases system performance in terms of rank-1 identification rate. After gender filtering of our synthesized and denoised data, we achieve a maximum rank-1 identification rate of 88.96% when using LBP DT matcher after synthesis for WVU VIS-MWIR and a rank-1 identification rate of 90.38% for the WVU VIS-LWIR spectrum, using WVU normalization. For the NVESD dataset, we achieve a maximum rank-1 identification rate of 98.00% when using LBP DT matcher and WVU pre-processing, after synthesis for NVESD VIS-MWIR and a rank-1 identification rate of 96.00% for the NVESD VIS-LWIR spectrum, using WVU and CSU normalization respectively.

5.2   Future Work

Although the foundation of this work provides several contributions that assist in both homogeneous and heterogeneous multi-spectral FR scenarios, there are additional aspects of this work that merit additional investigation moving forward.

With respect to homogeneous FR scenario, there has been a vast amount of research towards the collection and research of full-frontal profile face images. However, our results show the feasibility of using parts of the face for identification, in comparison to using the holistic face image. Furthermore, although users could be identified using parts of the face instead of the whole face, face sub-regions should be used in conjunction with other biometric approaches to boost performance. This needs further investigation, and if such an observation holds in large datasets, it can assist current FR systems in cases of facial occlusion or disguise. Following that, the application of our proposed methodology to datasets consisting of subjects with facial occlusion or disguise intending to counterfeit a biometric system would provide further insights in this research field. Also, to further validate the efficiency of our proposed FR approach and matchers, other datasets such as that of the University of Notre Dame [134], [30] should also be considered. Results suggest that the proposed algorithm is robust to minor pose and expression variations, assuming that the face is normalized accurately and similar poses are matched to one another. In the context of a biometric system, we need to be able to estimate a pose angle for the subject's face in order to constrain matches to similar poses. A quick solution that alleviates this challenge is the use of different poses as samples for each respective subject in both gallery and probe. Nevertheless, this would result in an increase of the computational complexity of our proposed system that could be avoided with prior knowledge of the pose angle. The problem of automated pose estimation and side-profile FR across the visible, MWIR and LWIR spectrums should be reserved for future work as well. Finally, another area that

merits further investigation in our homogeneous FR framework is the use of object-oriented code and parallel computing to speed up local fiducial point extraction and matching.

With respect to the proposed heterogeneous FR framework, the image synthesis approach can potentially discard important textural information from the face image. Therefore, this problem should be addressed in future work. One possible improvement could be the use of super-resolution or inpainting algorithms after image synthesis to help improve noise introduced during image synthesis of frontal face images. Another future direction is to extend the proposed approach to different spectral bands, variation of poses, and inclusion of image fusion as a cross function for better FR accuracy.

Through the use of two databases collected under different registration techniques (e.g. captured in two different spectrum at the same time), we are able to quantitatively examine the importance of co-registration. In instances where an algorithm may be alignment based, or there is learning based on the overlay of patches, there is a need for reasonably geometrized faces. Therefore, co-registration is a necessary step, and further future study is necessary where location dependency is mitigated so that design and evaluation are more robust.

Finally, another area that merits further investigation is automated filtering (classification) of face images by gender and race. Such an effort is expected to improve the integrity of the overall synthesis approach. The relationship between the IR spectrum and visible spectrum is highly non-linear and difficult to model. The heterogeneous FR framework utilized in this work uses statistical learning methods to try and estimate a manifold where two spectrums could lie close together. Through our results, we gather that manifold learning is highly data dependent and requires a good approximation of the underlying distribution of data. Given data restraints, particularly in IR to visible FR where datasets are very limited. The collection and organization of such data, particularly data that has been co-registered, should be considered in the future. Also, the use of techniques that may be able to learn mappings by adjusting projection coefficients over the training set, such as neural networks

should be considered as well.

Appendices

Appendix A

Sensor Experiments

A.1 Blackbody Experiments

An Santa Barbara Infrared (SBIR) Inc 920S differential black body and controller is used to carry out blackbody experiments. The blackbody allows the user to increase the temperature from 0 to 60 °C (0 to 333 °K). In the sensor experiments the temperature is incremented by a value of 5 °C in order to see the effects in the MWIR and LWIR bands over different temperature ranges. These figures can be seen in Figs. A.1 & A.2.

The results of this experiment conclude that thermal energy is initially detected by the blackbody around 25 °C for both the MWIR and LWIR sensor. However, for the LWIR sensors it takes about another 15 °C for the sensor to pick up peak temperature in the image, while it takes MWIR a 5 °C range. The LWIR sensor is a lot more sensitive in the 20 - 40 °C range.
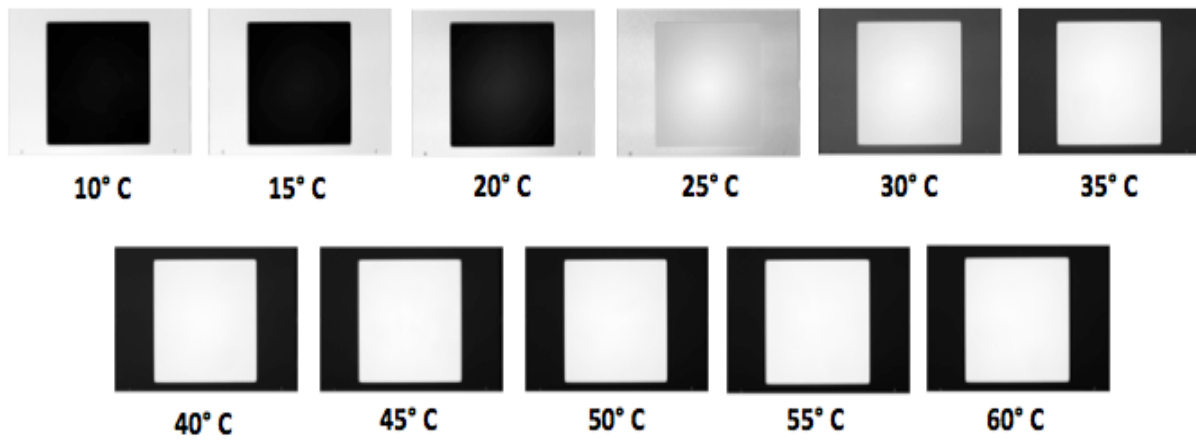
Figure A.1: SBIR Blackbody at various temperatures (C °) using MWIR sensor.
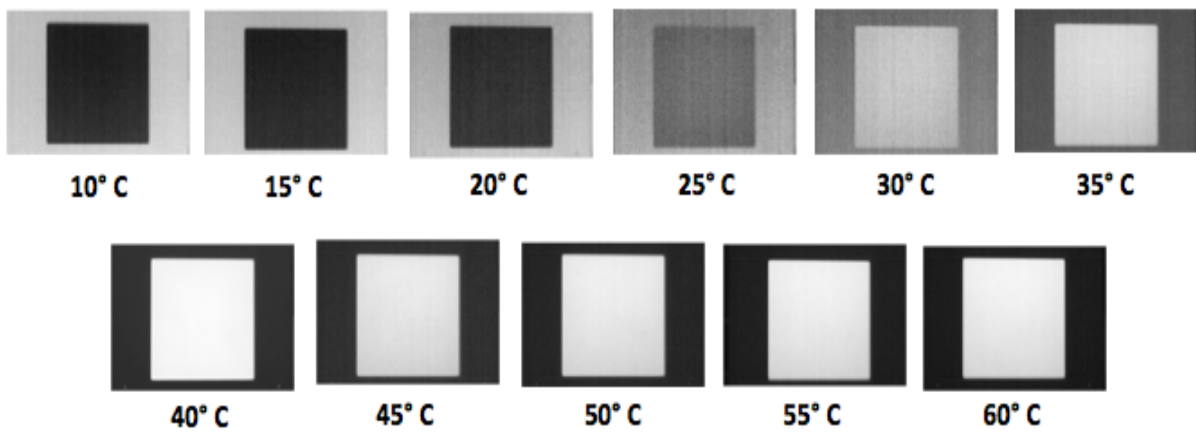
Figure A.2: SBIR Blackbody at various temperatures (C °) using LWIR sensor.

A.2   Actively and Passively Illuminated Face Recognition Experiment

There are two primary types of sensing used in detecting energy, known as passive and active sensing, or illumination. For example, the sun as a source of energy or radiation, may be either reflected (e.g. visible wavelengths) or absorbed and then re-emitted (e.g. thermal infrared wavelengths). Passive sensors detect and measure energy that is naturally available from the observed scene or object. Active sensors provide an external energy source for illumination which is directed towards the target being investigated.

Active sensing is usually used to radiate significantly within IR bands where the environment surrounding the object being imaged is at a much lower temperature. Active and passive sensing has been modeled using indoor and outdoor light sources. It has been qualitatively observed that IR imagery of human faces is invariant to changes in indoor and outdoor illumination, although there hasn't been much quantitative analysis to confirm this in open literature. In the SWIR and NWIR bands, some artificial illumination sources can be used to increase contrast when imaging by using reflected energy from objects being imaged. However, an advantage of the passive IR band is that most of the objects imaged within the band are good sources of IR energy, therefore no active illumination is necessary. If this holds true, there should not be an affect on FR between active illumination and passive illumination on the object being imaged. Measurements made in LWIR are inherently less affected by active illumination from ambient sources such as the sun and light bulbs. [135]

In an attempt to determine the effects of active and passive sensing on match scores (and inherently FR) in the IR band, we collect data samples both indoors and outdoors in the MWIR and LWIR spectrums respectively. After the face images are captured, they are normalized using WVU geometric normalization techniques. We look at the LBP distance transform (DT) between two samples of the same subject for four different scenarios, as well
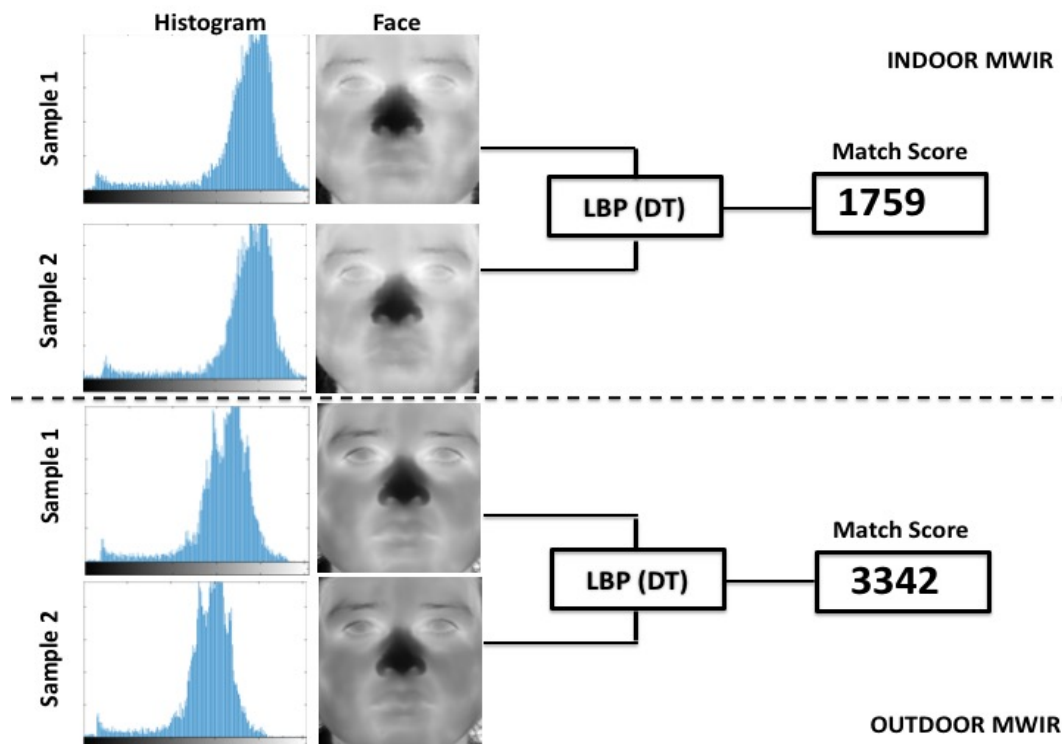
Figure A.3: Histogram and normalized face images of different samples for the same subject. LBP or difference scores between the two normalized face samples using (Top) Active/Indoor and (Bottom) Passive/Outdoor illumination in the MWIR band.
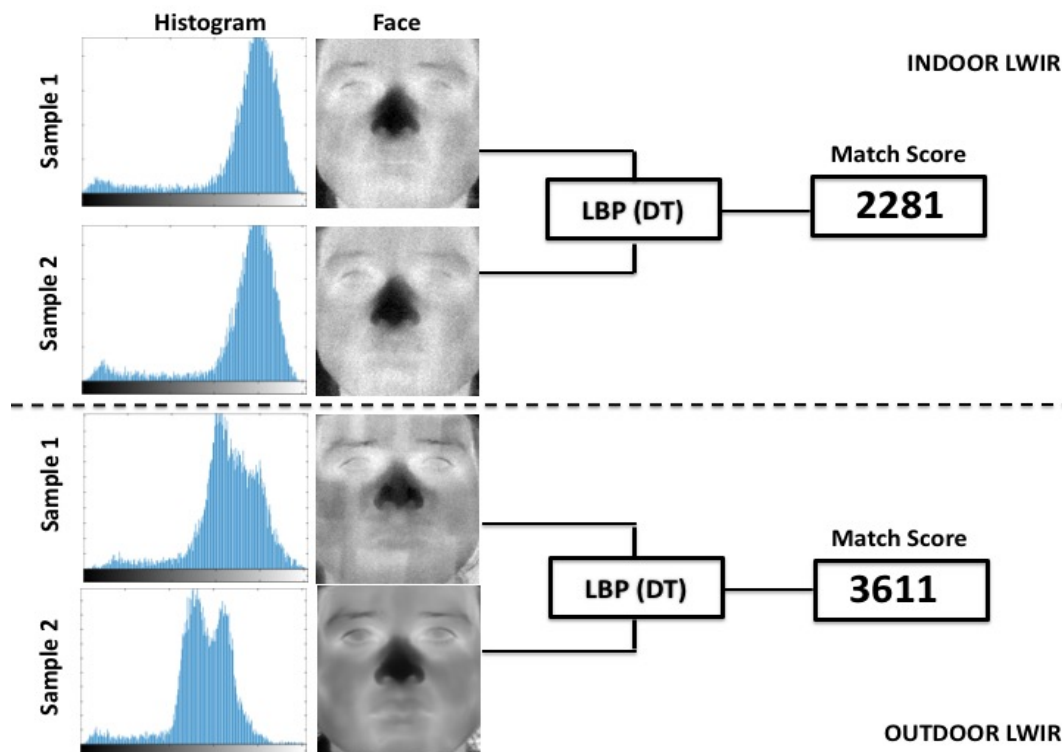
Figure A.4: Histogram and normalized face images of different samples for the same subject. LBP or difference scores between the two normalized face samples using (Top) Active/Indoor and (Bottom) Passive/Outdoor illumination in the LWIR band.

as the histogram for each subject sample as a quantitative measurement. LBP is a distance based metric, therefore a lower score indicates a stronger match, where a match score of 0 would be considered a perfect match.

Through our results, it appears that face appeared warmer indoor (e.g. modeled active illumination) according to the histogram, but there was greater variation in temperature when the subject was outdoors (e.g. modeled passive illumination). However, our results indicate that the temperature difference caused by external illumination source does not have a significant impact on FR for a single subject. When simulating active illumination (indoors), we found a match between two subjects to be stronger (smaller LBP score or distance), when compared to passive illumination (outdoors) irrespective of spectrum operation. Active illumination in the passive IR band would be useful in finding a hot object against a cold background, especially with obscuration like fog or smoke. Following this observation, active illumination in the passive IR band may be used to increase the signal to noise ratio in an image as well.

Appendix B

Cross-Validation of Gender Classifier

In this appendix, the gender classification approach introduced in 4.9.3 using HoG features and SVM classifier is further evaluated. Through cross-validation, we are able to estimate how our results would extend to a data set that is independent. The primary objective of cross-validation is to try and accurately predict how an approach will perform under practical constraints. When modeling prediction, the training dataset is created using a given dataset, that is tested against a training set that is unknown, with independent samples [136]. The goal of cross-validation is to try and limit overfitting by defining a test dataset for a approach in the training phase (e.g. validation dataset). This will result in an insightful look into how the approach will generalize to an unknown dataset, for instance from a practical problem. Cross-validation is also useful with respect to this work, because we are limited by our data, where further samples of IR images are difficult to obtain and collect.

Using our available WVU and NVESD databases, we partition the training and testing for each database into 4 different sets: (1) 20% training and 80% testing; (2) 40% training and 60% testing; (1) 60% training and 40% testing; (1) 80% training and 20% testing. The samples that are chosen for both training and testing sets are independent samples chosen at random with no bias. The sets are mutually exclusive. The results can be seen in Tables

Table B.1: Gender classification results (%) for Face Detected WVU & NVESD Passive IR Bands using 20% Training and 80% Testing.

	Gender Classification Accuracy (%) for Face Detected Images using HoG-SVM							
	WVU MWIR: Train 20% Test 80%		NVESD MWIR: Train 20% Test 80%		WVU LWIR: Train 20% Test 80%		NVESD LWIR: Train 20% Test 80%	
	Male	Female	Male	Female	Male	Female	Male	Female
Male	99.00	1.00	86.00	14.00	98.00	2.00	100.00	0.00
Female	18.00	82.00	12.00	88.00	37.00	63.00	17.00	83.00

Table B.2: Gender classification results (%) for Face Detected WVU & NVESD Passive IR Bands using 40% Training and 60% Testing.

	Gender Classification Accuracy (%) for Face Detected Images using HoG-SVM							
	WVU MWIR: Train 40% Test 60%		NVESD MWIR: Train 40% Test 60%		WVU LWIR: Train 40% Test 60%		NVESD LWIR: Train 40% Test 60%	
	Male	Female	Male	Female	Male	Female	Male	Female
Male	98.00	2.00	98.00	2.00	97.00	3.00	100.00	0.00
Female	15.00	85.00	12.00	88.00	29.00	71.00	11.00	89.00

B.1 - B.12 for the WVU and NVESD databases respectively.

Table B.3: Gender classification results (%) for Face Detected WVU & NVESD Passive IR Bands using 60% Training and 40% Testing.

	Gender Classification Accuracy (%) for Face Detected Images using HoG-SVM							
	WVU MWIR: Train 60% Test 40%		NVESD MWIR: Train 60% Test 40%		WVU LWIR: Train 60% Test 40%		NVESD LWIR: Train 60% Test 40%	
	Male	Female	Male	Female	Male	Female	Male	Female
Male	100.00	0.00	100.00	0.00	96.00	4.00	100.00	0.00
Female	7.00	93.00	0.00	100.00	0.00	100.00	0.00	100.00

Table B.4: Gender classification results (%) for Face Detected WVU & NVESD Passive IR Bands using 80% Training and 20% Testing.

	Gender Classification Accuracy (%) for Face Detected Images using HoG-SVM							
	WVU MWIR: Train 80% Test 20%		NVESD MWIR: Train 80% Test 20%		WVU LWIR: Train 80% Test 20%		NVESD LWIR: Train 80% Test 20%	
	Male	Female	Male	Female	Male	Female	Male	Female
Male	100.00	0.00	100.00	0.00	96.00	4.00	100.00	0.00
Female	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00

Table B.5: Gender classification results (%) for WVU Geometric Normalized WVU & NVESD Passive IR Bands using 20% Training and 80% Testing.

	Gender Classification Accuracy (%) for WVU Normalized Images using HoG-SVM							
	WVU MWIR: Train 20% Test 80%		NVESD MWIR: Train 20% Test 80%		WVU LWIR: Train 20% Test 80%		NVESD LWIR: Train 20% Test 80%	
	Male	Female	Male	Female	Male	Female	Male	Female
Male	90.00	10.00	96.00	4.00	99.00	1.00	100.00	0.00
Female	44.00	55.00	13.00	87.00	28.00	72.00	21.00	79.00

Table B.6: Gender classification results (%) for WVU Geometric Normalized WVU & NVESD Passive IR Bands using 40% Training and 60% Testing.

	Gender Classification Accuracy (%) for WVU Normalized Images using HoG-SVM							
	WVU MWIR: Train 40% Test 60%		NVESD MWIR: Train 40% Test 60%		WVU LWIR: Train 40% Test 60%		NVESD LWIR: Train 40% Test 60%	
	Male	Female	Male	Female	Male	Female	Male	Female
Male	96.00	4.00	91.00	9.00	97.00	3.00	100.00	0.00
Female	4.00	96.00	42.00	58.00	9.00	91.00	0.00	100.00

Table B.7: Gender classification results (%) for WVU Geometric Normalized WVU & NVESD Passive IR Bands using 60% Training and 40% Testing.

	Gender Classification Accuracy (%) for WVU Normalized Images using HoG-SVM							
	WVU MWIR: Train 60% Test 40%		NVESD MWIR: Train 60% Test 40%		WVU LWIR: Train 60% Test 40%		NVESD LWIR: Train 60% Test 40%	
	Male	Female	Male	Female	Male	Female	Male	Female
Male	100.00	0.00	100.00	0.00	97.00	3.00	100.00	0.00
Female	0.00	100.00	17.00	83.00	3.00	97.00	0.00	100.00

Table B.8: Gender classification results (%) for WVU Geometric Normalized WVU & NVESD Passive IR Bands using 80% Training and 20% Testing.

	Gender Classification Accuracy (%) for WVU Normalized Images using HoG-SVM							
	WVU MWIR: Train 80% Test 20%		NVESD MWIR: Train 80% Test 20%		WVU LWIR: Train 80% Test 20%		NVESD LWIR: Train 80% Test 20%	
	Male	Female	Male	Female	Male	Female	Male	Female
Male	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00
Female	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00

Table B.9: Gender classification results (%) for CSU Geometric Normalized WVU & NVESD Passive IR Bands using 20% Training and 80% Testing.

	Gender Classification Accuracy (%) for CSU Normalized Images using HoG-SVM							
	WVU MWIR: Train 20% Test 80%		NVESD MWIR: Train 20% Test 80%		WVU LWIR: Train 20% Test 80%		NVESD LWIR: Train 20% Test 80%	
	Male	Female	Male	Female	Male	Female	Male	Female
Male	91.00	9.00	98.00	2.00	97.00	3.00	97.00	3.00
Female	40.00	60.00	25.00	75.00	34.00	66.00	25.00	75.00

Table B.10: Gender classification results (%) for CSU Geometric Normalized WVU & NVESD Passive IR Bands using 40% Training and 60% Testing.

	Gender Classification Accuracy (%) for CSU Normalized Images using HoG-SVM							
	WVU MWIR: Train 40% Test 60%		NVESD MWIR: Train 40% Test 60%		WVU LWIR: Train 40% Test 60%		NVESD LWIR: Train 40% Test 60%	
	Male	Female	Male	Female	Male	Female	Male	Female
Male	95.00	5.00	98.00	2.00	95.00	5.00	100.00	0.00
Female	24.00	76.00	11.00	89.00	8.00	92.00	42.00	58.00

Table B.11: Gender classification results (%) for CSU Geometric Normalized WVU & NVESD Passive IR Bands using 60% Training and 40% Testing.

	Gender Classification Accuracy (%) for CSU Normalized Images using HoG-SVM							
	WVU MWIR: Train 60% Test 40%		NVESD MWIR: Train 60% Test 40%		WVU LWIR: Train 60% Test 40%		NVESD LWIR: Train 60% Test 40%	
	Male	Female	Male	Female	Male	Female	Male	Female
Male	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00
Female	10.00	90.00	0.00	100.00	3.00	97.00	25.00	75.00

Table B.12: Gender classification results (%) for CSU Geometric Normalized WVU & NVESD Passive IR Bands using 80% Training and 20% Testing.

	Gender Classification Accuracy (%) for CSU Normalized Images using HoG-SVM							
	WVU MWIR: Train 80% Test 20%		NVESD MWIR: Train 80% Test 20%		WVU LWIR: Train 80% Test 20%		NVESD LWIR: Train 80% Test 20%	
	Male	Female	Male	Female	Male	Female	Male	Female
Male	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00
Female	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00

Appendix C

Heterogeneous Framework Using Cross-Databases

In this appendix, some components of the image synthesis approach introduced in Chapter 4 for the heterogeneous framework is further analyzed. Since some of the methodology used in this work requires training, namely SVM, and CCA. it is important that experiments using cross-databases instead of two separate single-database are included. A cross-database can help further evaluate if the introduced approach is able to be generalized. In the results detailed in this appendix, the same experimental setup is utilized, with the only difference being that subjects from different databases were used for testing and training. The purpose of these experiments are to test if the components of the heterogeneous framework which require training generalize well to an unseen data. In other words, all components which require training will be done exclusively with the WVU database, and tested with the NVESD database, and vice-versa.

Table C.1: Rank-1 FR results (%) for Synthesized WVU VIS-MWIR and VIS-LWIR datasets using disjoint NVESD training and selected matcher (LBP DT).

WVU Synthesized (NVESD Trained) Rank-1 FR Accuracy (%) LBP DT				
Gallery	Probe	Face Detect	WVU	CSU
VIS	Synth. VIS (MWIR)	1.30	**4.55**	2.60
Synth. VIS (MWIR)	VIS	3.25	**5.19**	3.90
MWIR	Synth. MWIR	0.00	**3.25**	1.30
Synth. MWIR	MWIR	0.65	**8.44**	4.55
VIS	Synth. VIS (LWIR)	0.64	**7.05**	6.41
Synth. VIS (LWIR)	VIS	0.00	6.41	**8.33**
LWIR	Synth. LWIR	1.28	**2.56**	**2.56**
Synth. LWIR	LWIR	1.28	**7.05**	5.77

C.1   FR Experiments after Image Synthesis

With respect to testing the WVU database using NVESD database for training, we achieve a maximum rank-1 identification rate of 8.44% when pre-processing using WVU normalization and LBP DT matcher after synthesis for WVU VIS-MWIR. For the WVU VIS-LWIR spectrum, we achieve a maximum rank-1 identification rate of 8.33% for the WVU VIS-LWIR spectrums, using CSU normalization. When testing the NVESD database using WVU database for training, we achieve a maximum rank-1 identification rate of 10.00% when using LBP DT matcher after synthesis for NVESD VIS-MWIR and a rank-1 identification rate of 14.00% for the NVESD VIS-LWIR spectrums, using WVU normalization. The synthesis results can be seen in Tables C.1 and C.2 for the synthesized WVU and NVESD datasets respectively.

Table C.2: Rank-1 FR results (%) for Synthesized NVESD VIS-MWIR and VIS-LWIR datasets using WVU training and selected matcher (LBP DT).

NVESD Synthesized (WVU Trained) Rank-1 FR Accuracy (%) LBP DT				
Gallery	Probe	Face Detect	WVU	CSU
VIS	Synth. VIS (MWIR)	4.00	**4.00**	4.00
Synth. VIS (MWIR)	VIS	6.00	**8.00**	4.00
MWIR	Synth. MWIR	0.00	4.00	**10.00**
Synth. MWIR	MWIR	2.00	**10.00**	**10.00**
VIS	Synth. VIS (LWIR)	0.00	**10.00**	2.00
Synth. VIS (LWIR)	VIS	2.00	0.00	**6.00**
LWIR	Synth. LWIR	0.00	4.00	**6.00**
Synth. LWIR	LWIR	2.00	**14.00**	10.00

C.2 Cross-Database FR Experiments after Image De-noising

With respect to testing the WVU database using NVESD database for training, we achieve a maximum rank-1 identification rate of 5.84% when pre-processed using WVU normalization and LBP DT matcher after synthesis for WVU VIS-MWIR. A rank-1 identification rate of 9.62% using WVU normalization is achieved for the WVU VIS-LWIR spectrum. When testing the NVESD database using WVU database for training, we achieve a maximum rank-1 identification rate of 12.00% when using LBP DT matcher for NVESD VIS-MWIR and NVESD VIS-LWIR spectrums, using CSU normalization. The results for FR of synthesized images after image denoising can be seen in Tables C.3 and C.4 for the restored WVU and NVESD datasets respectively.

Table C.3: Rank-1 FR results (%) for Denoised Synthesized WVU VIS-MWIR and VIS-LWIR datasets using NVESD training and selected matcher (LBP DT).

WVU Denoised Synthesized (NVESD trained) Rank-1 FR Accuracy (%) LBP DT				
Gallery	Probe	Face Detect	WVU	CSU
VIS	Synth. VIS (MWIR)	0.65	2.60	**3.90**
Synth. VIS (MWIR)	VIS	1.95	5.19	**6.49**
MWIR	Synth. MWIR	3.90	**5.84**	2.60
Synth. MWIR	MWIR	0.00	**5.84**	4.55
VIS	Synth. VIS (LWIR)	1.92	**5.77**	3.21
Synth. VIS (LWIR)	VIS	0.00	2.56	**7.69**
LWIR	Synth. LWIR	1.28	3.21	**3.85**
Synth. LWIR	LWIR	1.28	**9.62**	7.05

Table C.4: Rank-1 FR results (%) for Denoised Synthesized NVESD VIS-MWIR and VIS-LWIR datasets using WVU training and selected matcher (LBP DT).

NVESD Denoised Synthesized (NVESD trained) Rank-1 FR Accuracy (%) LBP DT				
Gallery	Probe	Face Detect	WVU	CSU
VIS	Synth. VIS (MWIR)	4.00	**6.00**	**6.00**
Synth. VIS (MWIR)	VIS	6.00	**8.00**	**8.00**
MWIR	Synth. MWIR	0.00	0.00	**6.00**
Synth. MWIR	MWIR	0.00	10.00	**12.00**
VIS	Synth. VIS (LWIR)	2.00	**4.00**	**4.00**
Synth. VIS (LWIR)	VIS	2.00	6.00	**12.00**
LWIR	Synth. LWIR	4.00	**10.00**	4.00
Synth. LWIR	LWIR	2.00	**8.00**	**8.00**

Table C.5: SVM gender classification accuracy for NVESD dataset preprocessed using face detection and trained using leave-one-sample out method.

	Train: WVU MWIR		Train: WVU LWIR	
NVESD (Face Detection) Gender Classification Accuracy (%) using SVM				
	Male	Female	Male	Female
Male	**59.00**	41.00	**73.00**	27.00
Female	7.00	**93.00**	10.00	**90.00**

C.3   Cross-Database FR Experiments after Gender Classification

When training and testing the SVM classifier using cross-datasets, there is still pretty good performance across all pre-processing methods tested. Tables showing classification accuracy per gender, per dataset and training scenario can be seen in Tables C.5, C.6, C.7, C.8, C.9, and C.10.

Irrespective of spectrum of operation and pre-processing methodology, our male subjects are properly classified better than our female subjects. This may be due to the size of the respective subsets, within each datasets, noting fewer subjects that are female. The lowest classification rate we achieve for males is 59.00 % testing with WVU dataset and training with NVESD dataset for face detected images. For females, the lowest classification rate we achieve is 40.00 % for testing with NVESD dataset and training with WVU dataset, for both CSU and WVU face normalized images.

Table C.6: SVM gender classification accuracy for WVU dataset preprocessed using face detection and trained using leave-one-sample out method.

WVU (Face Detection) Gender Classification Accuracy (%) using SVM				
	Train: NVESD MWIR		Train: NVESD LWIR	
	Male	Female	Male	Female
Male	**94.00**	6.00	**40.00**	60.00
Female	44.00	**56.00**	5.00	**95.00**

Table C.7: SVM gender classification accuracy for NVESD dataset preprocessed using WVU face normalization and trained using leave-one-sample out method.

NVESD (WVU) Gender Classification Accuracy (%) using SVM				
	Train: WVU MWIR		Train: WVU LWIR	
	Male	Female	Male	Female
Male	**93.00**	7.00	**100.00**	0.00
Female	13.00	**87.00**	60.00	**40.00**

Table C.8: SVM gender classification accuracy for WVU dataset preprocessed using WVU face normalization and trained using leave-one-sample out method.

WVU (WVU) Gender Classification Accuracy (%) using SVM				
	Train: NVESD MWIR		Train: NVESD LWIR	
	Male	Female	Male	Female
Male	**95.00**	5.00	**96.00**	4.00
Female	44.00	**56.00**	46.00	**54.00**

Table C.9: SVM gender classification accuracy for NVESD dataset preprocessed using CSU face normalization and trained using leave-one-sample out method.

NVESD (CSU) Gender Classification Accuracy (%) using SVM				
	Train: WVU MWIR		Train: WVU LWIR	
	Male	Female	Male	Female
Male	**93.00**	7.00	**100.00**	0.00
Female	17.00	**83.00**	60.00	**40.00**

Table C.10: SVM gender classification accuracy for WVU dataset preprocessed using CSU face normalization and trained using leave-one-sample out method.

WVU (CSU) Gender Classification Accuracy (%) using SVM				
	Train: NVESD MWIR		Train: NVESD LWIR	
	Male	Female	Male	Female
Male	**94.00**	6.00	**97.00**	3.00
Female	59.00	**41.00**	43.00	**57.00**

C.4   Cross-Database FR Experiments after Demographic Filtering

With respect to testing the WVU database using NVESD database for training, we achieve a maximum rank-1 identification rate of 9.74% when using CSU normalization for preprocessing and LBP DT matcher after synthesis for WVU VIS-MWIR. For the WVU VIS-LWIR scenario, a rank-1 identification rate of 14.74% for the WVU VIS-LWIR spectrum was achieved using WVU normalization. When testing the NVESD database using WVU database for training, we achieve a maximum rank-1 identification rate of 22.00% when using LBP DT matcher and WVU pre-processing after synthesis for NVESD VIS-MWIR. For the NVESD VIS-LWIR spectrum, a maximum rank-1 identification rate of 20.00%, using CSU normalization . The results for FR of the denoised synthesized images after image denoising and demographic filtering can be seen in Tables C.11 and C.12 for the synthesized WVU and NVESD datasets respectively.

Table C.11: Rank-1 FR results (%) for Cohort Filtered and Denoised Synthesized WVU VIS-MWIR and WVU VIS-LWIR datasets using NVESD training and selected matcher (LBP DT).

WVU Cohort Filtered (NVESD Trained) Rank-1 FR Accuracy (%) LBP DT				
Gallery	Probe	Face Detect	WVU	CSU
VIS	Synth. VIS (MWIR)	1.30	6.49	**7.14**
Synth. VIS (MWIR)	VIS	3.90	5.19	**9.74**
MWIR	Synth. MWIR	5.19	**8.44**	3.90
Synth. MWIR	MWIR	4.55	7.79	**9.09**
VIS	Synth. VIS (LWIR)	2.56	**9.62**	4.49
Synth. VIS (LWIR)	VIS	0.00	7.05	**12.18**
LWIR	Synth. LWIR	2.56	**6.41**	5.77
Synth. LWIR	LWIR	3.21	**14.74**	11.54

Table C.12: Rank-1 FR results (%) for Cohort Filtered and Denoised Synthesized NVESD VIS-MWIR and NVESD VIS-LWIR datasets using WVU training and selected matcher (LBP DT).

NVESD Cohort Filtered (WVU Trained) Rank-1 FR Accuracy (%) LBP DT				
Gallery	Probe	Face Detect	WVU	CSU
VIS	Synth. VIS (MWIR)	6.00	**10.00**	8.00
Synth. VIS (MWIR)	VIS	10.00	10.00	**16.00**
MWIR	Synth. MWIR	4.00	10.00	**14.00**
Synth. MWIR	MWIR	4.00	**22.00**	14.00
VIS	Synth. VIS (LWIR)	4.00	**10.00**	4.00
Synth. VIS (LWIR)	VIS	6.00	8.00	**20.00**
LWIR	Synth. LWIR	2.00	10.00	**12.00**
Synth. LWIR	LWIR	4.00	**16.00**	14.00

Bibliography

[1] A. Jain, R. Bolle, S. Pankanti, Introduction to Biometrics. In Biometrics: Personal Identification in Networked Society, 2002.

[2] T. Bourlai, A. Ross, C. Chen, L. Hornak, A Study on using Middle-Wave Infrared Images for Face Recognition, SPIE, Biometric Tech. for Human Identification IX.

[3] A. K. Jain, P. Flynn, A. Ross, Handbook of Biometrics, Springer, New York, NY, U.S.A, 2007.

[4] A. K. Jain, R. Bolle, S. Pankati, Biometrics: Personal Identification in Networked Security, London, 1999, pp. 103–211.

[5] T. Bourlai, B. Cukic, Multi-spectral face recognition: Identification of people in difficult environments, in: Proc. IEEE Int. Conf. Intell. and Security Informatics (ISI), 2012, pp. 196–201.

[6] D. Socolinsky, A. Selinger, J.Neuheisel, Face recognition with visible and thermal imagery, CVIU 91 (2003) 72–114.

[7] Y. Yoshitomi, T. Miyaura, S. Tomita, S. Kimura, Face identification using thermal image processing, WRHC (1997) 374–379.

[8] D. Yi, S. Liao, Z. Lei, J. Sang, S. Z. Li, Partial face matching between near infrared

and visual images in mbgc portal challenge, in: Proceedings of the Third International Conference on Advances in Biometrics, ICB, Springer-Verlag, Berlin, Heidelberg, 2009, pp. 733–742.

[9] A. Jain, A. Ross, Bridging the gap: From biometrics to forensics, Philosophical Transactions of The Royal Society B 370 (1674).

[10] D. Maltoni, D. Maio, A. Jain, S. Prabhakar, Handbook of Fingerprint recognition, Springer Verlag, New York, USA, 2003.

[11] A. Jain, A. Ross, S. Prabhakar, An introduction to biometric recognition, IEEE Transactions on circuits and systems for video technology 14 (1).

[12] D. Blackburn, Biometrics 101, version 3.1, Tech. rep., Federal Bureau of Investigation) (March 2004).

[13] A. K. Jain, S. Pankanti, S. Prabhakar, Biometric recognition: Security and privacy concerns, IEEE Security & Privacy.

[14] P. J. Phillips, P. Grother, R. J. Michaels, D. Blackburn, E. Tabassi, M. Bone, Face recognition vendor test 2002: Evaluation report, Tech. rep., http://www.frvt.org (March 2003).

[15] M. M. Sandstrm, Liveness detection in fingerprint recognition systems, Master's thesis (2004).

[16] R. de Luis-Garcia, C. Alberola-Lpez, O. Aghzout, J. Ruiz-Alzola, Biometric identification systems, Signal Processing 83 (12) 2539–2557.

[17] A. Ross, A. Jain, Information fusion in biometrics, Pattern Recognition Letters 24 (2003) 2115–2125.

[18] L. Hong, A. K. Jain, Integrating faces and fingerprints for personal identification, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (12) (1998) 1295–1307.

[19] S. A. C. Schuckers, Spoofing and anti-spoofing measures. information security technical report, Tech. Rep. 4 (December 2002).

[20] S. Li, X. Kanga, L. Fanga, J. Hub, H. Yinc, Pixel-level image fusion: A survey of the state of the art, Information Fusion 33 (2017) 100–112.

[21] F. Al-Wassai, N. Kalyankar, A. Ross, Multisensor images fusion based on feature-level, International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS) 1 (5) (2012) 124–138.

[22] R. Snelick, U. Uludag, A. Mink, M. Indovina, A. Jain, Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (3) (2005) 450–455.

[23] P. Huber, Robust Statistics, Wiley, 1981.

[24] J. Byrnes, Unexploded Ordnance Detection and Mitigation, Springer, 2009, pp. 21–22.

[25] N. K. Dhar, R. Dat, A. K. Sood, Advances in Infrared Detector Array Technology, INTECH Open Access Publisher, 2013.

[26] T. Bourlai, N. Kalka, A. Ross, B. Cukic, L. Hornak, Cross-spectral Face Verification in Short Infrared Band, in: Proc. IEEE, Int. Conf. on Pattern Recognition (ICPR), Istanbul, 2010, pp. 1343–1347.

[27] A. Selinger, D. Socolinksy, Face recognition in the dark, CPRW (2004) 129–134.

[28] J. Wilder, P. Phillips, C. Jiang, S. Wiener, Comparison of visible and infra-red imagery for face recognition, in: Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, 1996, pp. 182–187.

[29] T. Bourlai, Z. Jafri, Eye detection in the Middle-Wave Infrared spectrum: Towards recognition in the dark, IEEE Intl. Workshop on Information Forensics and Security (WIFS) (2011) 1–6.

[30] P. Phillips, J. Beveridge, B. Draper, G. Givens, A. O'Toole, D. Bolme, J. Dunlop, Y. M. Lui, H. Sahibzada, S. Weimer, An introduction to the good, bad and ugly face recognition challenge problem, in: Automatic Face Gesture Recognition, 2011.

[31] A. K. Jain, B. Klare, U. Park, Face matching and retrieval: Applications in forensics, Vol. 19, IEEE Multimedia, 2012, pp. 20–28.

[32] A. K. Jain, B. F. Klare, U. Park, Face recognition: Some challenges in forensics, in: Int. Conference on Automatic Face and Gesture Recognition, 2011.

[33] B. Klare, A. Jain, Heterogeneous face recognition: Matching nir to visible light images, ICIP.

[34] B. F. Klare, M. Burge, Assessment of h.264 video compression on automated face recognition performance in surveillance and mobile video scenarios, in: SPIE, Biometric Technology for Human Identification, Vol. VII, 2010.

[35] M. A. Goodale, A. Milner, Separate visual pathways for perception and action, Trends in Neurosciences 15 (1) (1992) 20–25.

[36] N. Kanwisher, J. McDermott, M. M. Chun, The fusiform face area: A module in human extrastriate cortex specialized for face perception, Journal of Neuroscience 17 (11) (1997) 4302–4311.

[37] R. Mauro, M. Kubovy, Caricature and face recognition, Memory & Cognition 20 (4) (1992) 433–440.

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

[39] J. C. Brigham, The influence of race on face recognition, Aspects of face processing (1986) 170–177.

[40] J. Sadr, I. Jarudi, P. Sinha, The role of eyebrows in face recognition, Perception 32 (2003) 285–293.

[41] P. Viola, M. Jones, Robust real-time face detection, Journal of Computer Vision 57 (2) (2004) 137–154.

[42] J. Parris, M. Wilber, B. Helfin, H. Rara, A. E. Barkouky, A. Farag, J. Movellan, M. Santana, J. Lorenzo, M. N. Teli, S. Marcel, C. Atanasoaei, Face and eye detection on hard datasets, in: Int. Joint Conference on Biometrics, 2011.

[43] J. Short, J. Kittler, K. Messer, A comparison of photometric normalisation algorithms for face verification, 2004, pp. 254–259.

[44] A. Reza, Realization of contrast limited adaptive histogram equalization (clahe) for real-time image enhancement, VLSI Signal Processing 38 (2004) 35–44.

[45] V. Struc, Inface: A toolbox for illumination invariant face recognition, http://luks.fe.uni-lj.si/sl/osebje/vitomir/face_tools/INFace/ (2009).

[46] H. Wang, S. Li, Y. Wang, J. Zhang, Self quotient image for face recognition, in: ICIP, 2004, pp. 1397–1400.

[47] S. Kong, J. Heo, B. Abidi, J. Paik, M. Abidi, Recent advances in visual and infrared face recognition - a review, Comput. Vis. and Image Understanding 97 (1) (2005) 103–135.

[48] P. Belhumeur, J. Hespanha, D. J. Kriegman, Eigenfaces vs. fisherfaces: Recognition using class specic linear projection, IEEE Trans. Patt. Anal. Mach. Intell. 19 (7) (1997) 711–720.

[49] L. Sirovich, M. Kirby, Application of the karhunen-loeve procedure for the characterization of human faces, IEEE Trans. Patt. Anal. Mach. Intell. 12 (1) (1990) 103–108.

[50] M. Turk, A. Pentland, Eigenfaces for recognition 3 (1) (1991) 71–86.

[51] J. Chirillo, S. Blaul, Implementing biometric security, Wiley Publishing, 2003.

[52] W. Zhao, R. Chellappa, P. J. Phillips, A. Rosenfeld, Face recognition: A literature survey, ACM Computing Surveys 35 (4).

[53] J. D. W. Jr., N. M. Orlans, P. T. Higgins, Biometrics, McGrawHill/Osborne, New York, 2003.

[54] I. B. Group, Facial scan technology: How it works, Tech. rep., International Biometrics Group) (2002).

[55] L. Wiskott, J. Fellous, N. Kruger, C. Malsburg, Face recognition by elastic bunch graph matching, IEEE Trans. Patt. Anal. Mach. Intell. 19 (7) (1997) 775–779.

[56] J. Wayman, A. Jain, D. Maltoni, D. Maio, Biometric Systems: Technology, Design and Performance Evaluation, Springer-Verlag London Limited, 2005.

[57] J. Zhang, Y. Yan, M. Lades, Face recognition: Eigenface, elastic matching, and neural nets, Vol. 85, 1997, pp. 1423–1435.

[58] D. Bolme, J. Beveridge, M. Teixeira, B. Draper, The csu face identification evaluation system: Its purpose, features and structure, in: Proc. International Conference on Vision Systems, 2003, pp. 301–311.

[59] B. D. andS. Fischer, J. Bigun, Face authentication with gabor information on deformable graphs, Transactions on Image Processing 8 (4).

[60] R. Brunelli, T. Poggio, Face recognition: Features versus templates, Transactions on Pattern Analysis and Machine Intelligence 15 (10) (1993) 1042–1052.

[61] I. Craw, H. Ellis, J. Lishman, Automatic extraction of face features, Pattern Recognition Letters 5 (1987) 183–187.

[62] R. Baron, Mechanisms of human facial recognition, International Journal of Man Machine Studies 15 (1981) 137–178.

[63] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, Trans. Img. Proc. 19.

[64] R. Singh, M. Vatsa, A. Noore, Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition, Pattern Recognition 41 (3) (2008) 880–893.

[65] S. Kong, J. Heo, F. Boughorbel, Y. Zheng, B. Abidi, A. Koschan, M. Yi, M. Abidi, Multiscale fusion of visible and thermal ir images for illumination-invariant face recognition, International Journal of Computer Vision 71 (2) (2007) 215–233.

[66] A. Gyaourova, G. Bebis, I. Pavlidis, Fusion of infrared and visible images for face recognition, Vol. 3024, 2004, pp. 456–468.

[67] P. Buddharaju, I. Pavlidis, C. Manohar, Face Recognition Beyond the Visible Spectrum, Advances in Biometrics: Sensors, Algorithms and Systems.

[68] H. Mendez, C. Martin, J. Kittler, Y. Plasencia, E. Reyes, Face recognition with lwir imagery using local binary patterns, in: Proc. Int. Conf. Advances in Biometrics (ICB), Springer-Verlag, Berlin, 2009, pp. 327–336.

[69] D. Socolinsky, L. Wolff, J. Neuheisel, C. Eveland, Illumination Invariant Face Recognition Using Thermal Infrared Imagery, in: Proc. IEEE CS Conf. Comput. Vis. Patt. Recogn. (CVPR), 2001, pp. 527–534.

[70] L. Trujillo, G. Olague, R. Hammoud, B. Hernandez, Automatic Feature Localization in Thermal Images for Facial Expression Recognition, in: Proc. IEEE CS Conf. Comput. Vis. Patt. Recogn. (CVPR), Vol. 14, 2005.

[71] X. Chen, P. Flynn, K. Bowyer, Pca-based Face Recognition in Infrared Imagery: Baseline and Comparative Studies, in: Proc. IEEE Int. Workshop on Anal. and Modeling of Faces and Gestures (AMFG), 2003, pp. 127–134.

[72] T. Elguebaly, N. Bouguila, A bayesian method for infrared face recognition, Machine Vision Beyond Visible Spectrum 1 (2011) 123–138.

[73] A. Srivastana, X. Liu, Statistical hypothesis pruning for recognizing faces from infrared images, Image Vis. Comput. 21 (7) (2003) 651–661.

[74] T. Mandal, A. Majumdar, Q. Wu, Face recognition by curvelet based feature extraction, in: ICIAR, 2007, pp. 806–817.

[75] Z. Xie, S. Wu, G. Liu, Z. Fang, Infrared face recognition based on radiantenergy and curvelet transformation, in: International Conference on Information Assurance and Security (IAS), Vol. 2, 2009, pp. 215–218.

[76] S. Wu, W. Song, L. Jiang, S. Xie, F. Pan, W. Yau, S. Ranganath, Infrared face

recognition by using blood perfusion data, in: International Conference on Audio and Video-Based Biometric Person Authentication, 2005, pp. 320–328.

[77] Z. Xie, G. Liu, S. Wu, Z. Fang, Infrared face recognition based on blood perfusion and fisher linear discrimination analysis, in: IST, 2009, pp. 85–88.

[78] P. Buddharaju, I. Pavlidis, P. Tsiamyrtzis, M. Bazakos, Physiology-based face recognition in the thermal infrared spectrum, IEEE Trans. Patt. Anal. Mach. Intell. 29 (4) (2007) 613–626.

[79] O. Nakamura, S. Mathur, T. Minami, Identification of human faces based on isodensity maps, in: IEEE Proc. Patt. Recogn., Vol. 24, 1991, pp. 263–272.

[80] D. Lowe, Object recognition from local scale-invariant features, Vol. 2, 1999, pp. 1150–1157.

[81] H. Bay, A. Ess, T. Tuytelaars, L. Gool, Surf: Speeded up robust features, Comput. Vis. and Image Und. (CVIU) 110 (3) (2008) 346–359.

[82] P. Perona, J. Malik, Scale space and edge detection using anisotropic diffusion, IEEE Tran. Patt. Anal. Mach. Intell. 12 (7) (1990) 629–639.

[83] W. Zhili, Fingerprint recognition, Ph.D. thesis, Honk Kong Baptist, University (2002).

[84] A. K. Jain, K. Nandakumar, A. Ross, Score normalization in multimodal biometric systems, Patt. Recogn. 38 (12) (2005) 2270–2285.

[85] A. P. Devijver, J. Kittler, Pattern Recognition: A Statistical Approach, Prentice-Hall, 1982.

[86] M. Teixeira, The bayesian intrapersonal/extrapersonal classifier, Ph.D. thesis, Colorado State University (2003).

[87] M. Pietikinen, Image analysis with local binary patterns, in: Proc. Scandinavian Conf. Image Anal., 2005, pp. 115–118.

[88] N. Osia, T. Bourlai, Holistic and Partial Face Recognition in the Mwir Band using Manual and Automatic Detection of Face-based Features, in: IEEE Int. Conf. Tech. for Homeland Security (HST), 2012, pp. 273–279.

[89] X. Lu, R. Hsu, A. Jain, B. Kamgar-Parsi, B. Kamgar-Parsi, Face recognition with 3d model-based synthesis, in: Proc. Internat. Conf. on Biometric Authentication (ICBA), Honk Kong, 2004, pp. 139–146.

[90] D. Yi, R. Liu, R. Chu, Z. Lei, S. Li, Partial face matching between near infrared and visual images in mbgc portal challenge, in: ICB, 2007.

[91] C. Li, G. Su, Y. Shang, Y. Li, Y. Xiang, Face Recognition Based on Pose-Variant Image Synthesis and Multi-level Multi-feature Fusion, AMFG (2007) 261–275.

[92] A. Majumdar, R. K. Ward, Single image per person face recognition with images synthesized by non-linear approximation, in: International Conference on Image Processing, 2008, pp. 2740–2743.

[93] A. Ansari, M. Mahoor, M. Abdel-Mottaleb, Normalized 3d to 2d model-based facial image synthesis for 2d model-based face recognition, in: IEEE GCC Conf. and Exhibition (GCC), 2011, pp. 178–181.

[94] M. Sarfraz, R. Stiefelhagen, Deep perceptual mapping for thermal to visible face recognition, in: British Machine Vision Conference (BMVC), 2016.

[95] X. Tang, X. Wang, Face sketch recognition, IEEE Trans. Circuits and Systems for Video Technology 14 (1) (2004) 50–57.

[96] Q. Liu, X. Tang, H. Jin, H. Lu, S. Ma, A nonlinear approach for face sketch synthesis and recognition, Vol. 1, 2005, pp. 1005–1010.

[97] J. L. W. Liu, X. Tang, Bayesian tensor inference for sketch-based facial photo hallucination, in: IJCAI, 2007, pp. 2141–2146.

[98] X. Wang, X. Tang, Face photo-sketch synthesis and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 31 (11) (2009) 1955–1967.

[99] Z. Lei, Q. Bai, R. He, S. Li, Face shape recovery from a single image using cca mapping between tensor spaces, 2008.

[100] Z. Lei, S. Li, Coupled spectral regression for matching heterogeneous faces, CVPR.

[101] W. Zhang, X. Wang, X. Tang, Lighting and pose robust face sketch synthesis, Vol. 6, 2010, pp. 420–433.

[102] J. Chen, D. Yi, J. Yang, G. Zhao, S. Li, M. Pietikainen, Learning mappings for face synthesis from near infrared to visual light images, in: CVPR, 2009.

[103] S. Li, R. Chu, S. Liao, L. Zhang, Illumination invariant face recognition using near-infrared images, PAMI.

[104] W. Zhang, X. Wang, X. Tang, Coupled information-theoretic encoding for face photo-sketch recognition, 2011, pp. 513–520.

[105] B. Klare, Z. Li, A. Jain, Matching forensic sketches to mug shot photos, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (3) (2011) 639–646.

[106] B. Klare, A. Jain, Heterogeneous face recognition using kernel prototype similarities, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (6) (2013) 1410–1422.

[107] D. Lin, X. Tang, Inter-modality face recognition, in: Proceedings of the European Conference on Computer Vision, Vol. 3954, 2006, pp. 13–26.

[108] Z. Lei, S. Liao, A. Jain, S. Li, Coupled discriminant analysis for heterogeneous face recognition, IEEE Transactions on Information Forensics and Security 7 (6) (2012) 1707–1716.

[109] H. Shuowen, N. Short, P. Gurram, K. Gurton, C. Reale, FR Across the Imaging Spectrum, 2016, Ch. 4.

[110] C. Chen, A. Ross, Matching thermal to visible face images using hidden factor analysis in a cascaded subspace learning framework, Pattern Recognition Letters 72 (2016) 25–32.

[111] R. Wang, J. Yang, D. Yi, S. Li, An analysis-by-synthesis method for heterogeneous face biometrics, in: ICB, 2009.

[112] Z. Zhang, Y. Wang, Z. Zhang, Face synthesis from near-infrared to visual light via sparse representation, ICB.

[113] M. Dou, C. Zhang, P. Hao, J. Li, Converting thermal infrared face images into normal gray-level images, ACCV.

[114] H.Chang, D. Yeung, Y. Xiong, Super-resolution through neighbor embedding, CVPR.

[115] W. Fan, D. Yeung, Image hallucination using neighbor embedding over visual primitive manifolds, CVPR.

[116] D. Beymer, T. Poggio, Image representation for visual learning, Science.

[117] P. Belhumeur, J. Hespanha, D. J. Kriegman, Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, IEEE Trans. Patt. Anal. Mach. Intell. 19 (7) (1997) 711–720.

[118] T. Melzer, M. Reiter, H. Bischof, Appearance Models Based on Kernel Canonical Correlation Analysis, Pattern Recognition 36 (2003) 1961–1971.

[119] J. Tenenbaum, Mapping a manifold of perceptual observations, Advances in Neural Information Processing Systems 10 (1998) 682–688.

[120] J. Tenenbaum, V. de Silva, J. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.

[121] S. Roweis, L. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.

[122] L. Saul, S. Roweis, Think globally, fit locally: unsupervised learning of low dimensional manifolds, Journal of Machine Learning Research 4 (2003) 119–155.

[123] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, Advances in Neural Information Processing Systems (14) (2002) 585–591.

[124] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Computation 15 (6) (2003) 1373–1396.

[125] K. Byrd, Preview of the newly acquired nvesd-arl multimodal face database, 2013, pp. 8734–34.

[126] S. K. Mohideen, S. A. Perumal, M. M. Sathik, Image de-noising using discrete wavelet transform, Int. J. Comput. Sci. Netw. Security 8 (1) (2008) 213–216.

[127] D. Donoho, I. Johnstone, Ideal spatial adaptation via wavelet shrinkage, Biometrika 81 (1994) 425–455.

[128] R. R. Coifman, D. L. Donoho, Translation-invariant de-noising, Wavelets and Statistics 103 (2012) 125–150.

[129] T. Bourlai, N. Kalka, A. Ross, B. Cukic, L. Hornak, Cross-spectral Face Verification in the Short Wave Infrared (SWIR) Band, 2010.

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

[131] K. Muller, S. Mika, G. Ratsch, K. Tsuda, B. Scholkopf, An introduction to kernel-based learning algorithms, in: IEEE Transactions on Neural Networks, 2001, pp. 181–201.

[132] N. Vapnikv, The Nature of Statistical Learning Theory, Springer, 1995.

[133] B. Decann, A. Ross, Can a poor verification system be a good identification system? a preliminary study, 2012.

[134] U. o. N. D. The Computer Vision Laboratory, Biometrics database distribution, `http://www.nd.edu/~cvrl` (2002).

[135] R. Hammoud, Augmented Vision Perception in Infrared, Springer, 2009.

[136] F. Mosteller, A k-sample slippage test for an extreme population, Annals of Mathematical Statistics 19 (1948) 58–65.