# FACE RECOGNITION WITH VISIBLE AND THERMAL IR IMAGES

---

A Thesis
Submitted to
the Temple University Graduate Board

---

In Partial Fulfillment
Of the Requirements for the Degree
MASTER OF SCIENCE IN ENGINEERING

---

By
Lei Guan
August, 2010

Thesis Approvals:

Dr. Seong G. Kong, Thesis Advisor, Department of Electrical and Computer Engineering, Temple University
Dr. Li Bai, Committee Member, Department of Electrical and Computer Engineering, Temple University
Dr. Joseph Picone, Committee Member, Department of Electrical and Computer Engineering, Temple University

## ABSTRACT

This thesis describes how the fusion of visible and thermal infrared (IR) images can be used to improve the performance of face recognition techniques, especially when illumination variations and occlusions are involved. Visible images are sensitive to illumination variations, while thermal IR images are robust to them. However, thermal IR images are degraded by occlusions caused from eyeglasses, but visible images can provide detailed information around the eyes even when eyeglasses are present. Fusion techniques, which combine complementary information from both spectrums, generate information that is robust to both illumination variations and occlusions.

Before two images are fused, they must be registered. In this thesis, edge-based mutual information is used to register both visible and thermal IR images taken under different conditions. Following that, eyeglasses (if present) are removed from the thermal IR image, and replaced by eyes that are reconstructed from the visible image. Then, data-level, feature-level, and score-level fusion techniques are applied to the visible and thermal IR images for face recognition. Experimental results using the NIST/Equinox database showed that the fusion of visible and thermal IR images increased the number of first matches by 22% over visible images, and 8% over thermal IR images.

Unfortunately, thermal IR sensors may be cost-prohibitive for many applications. In consideration of this, this thesis explores ways to predict a novelty component from

the visible image. A novelty component is a thermal-like image that can be obtained from information in the visible image. It is later fused with the visible image for face recognition. Experimental results based upon four face recognition algorithms showed that the fusion of visible images and their novelty components increased the number of first matches over visible images by 21% (using the NIST/Equinox database) and 17% (using the Extended Yale Face Database B).

**ACKNOWLEDGEMENT**

I would like to express my gratitude to my advisor Dr. Seong G. Kong for the support and direction that he has given me. He has spent many hours carefully revising and evaluating my research. I wish to thank my committee members Dr. Li Bai and Dr. Joseph Picone for their helpful insights and suggestions. I would like to thank Dr. Choonwo Ryu as well as all the members of the Imaging and Pattern Recognition Lab for their assistance in completion of this thesis. I also extend my sincere thanks to faculty and students from College of Engineering in Temple University for their help.

Finally, I wish to thank my family for their support and encouragement.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

As one of the most successful applications of image analysis and understanding, face recognition has been a rapidly growing research area for many years. Face recognition addresses the problem of identifying or verifying a person by comparing his/her face with the face images stored in a database. General procedures for face recognition include face detection, face normalization, feature extraction, and recognition. Face detection can segment the face from complicated backgrounds. Face normalization is used to normalize the image to ensure that the input face is the same size and at the same position as the faces stored in the database. Feature extraction refers to a process where the normalized face is represented as low-dimensional vectors. Face recognition includes both identification and verification. Face identification is the process of providing a ranked list of candidates that best match the face in the input face. For face verification, an "identity claim" and face are presented to the system, and the system either accepts or rejects the claim by comparing the face with those stored in the database.

Although significant progress has been made in the area of face recognition, it continues to remain a challenging task. For example, face recognition in the visible spectrum is influenced by variations in illumination, pose, facial expression, viewpoint,

and disguise. The thermal infrared (IR) imaging is more robust to illumination variations because it captures the emitted energy from an object in order to generate an image. Therefore, the use of thermal IR imagery has significant advantages over visible imagery for face recognition under varying illumination conditions. However, thermal IR imagery is sensitive to changes in body and ambient temperature, as well as the presence of glass. On the contrary, visible imagery is more robust to these factors. Considering the complementary information contained in visible and thermal IR images, the fusion of both could be used to improve the accuracy of face recognition.

## 1.1 Fundamental Issues in Face Recognition

Human beings have used the face, ears, voice, gait, iris, and fingerprints as biometrics for a long time. Face recognition is one of the most successful and widely used applications of biometric analysis techniques. Unlike fingerprint analysis or iris scans that depend on the cooperation of participants, face recognition is a noninvasive process that can be effective with uncooperative users. A comparative analysis of various biometrics can be found in [1]. Application areas for face recognition include entertainment (e.g., video game and virtual reality), smart cards (e.g., driver's license, passports, and voter registration), information security (e.g., personal device and desktop logon, internet access, and database security), and law enforcement and surveillance (e.g., CCTV control and suspect tracking and investigation) [2].

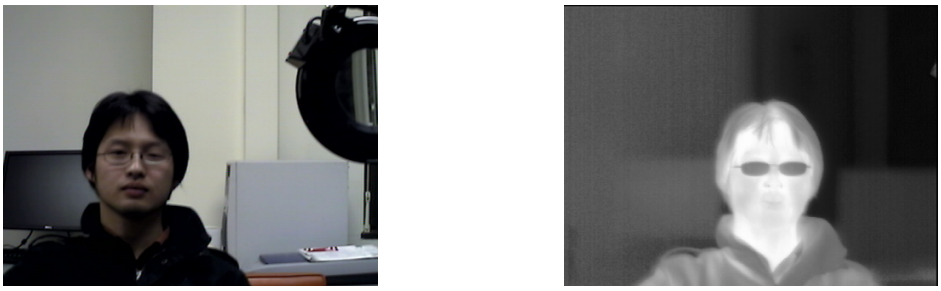
However, face recognition in an uncontrolled environment remains a challenging task. For example, the performance of face recognition algorithms in the visible spectrum

is sensitive to variations in facial expression, pose, viewpoint, and disguise [3-6]. Outdoor environments present a further challenge due to illumination variations resulting from the fact that the face is a 3D object (i.e., light from different directions will cast different shadows on the face). Although the use of an artificial illumination source can reduce such variations, it may distract the person or expose the presence of the surveillance system.

Thermal IR imagery, which captures the energy emitted from an object, is robust to illumination variations. Therefore, this technology offers a promising alternative to visible imaging for face recognition. The thermal IR band, which consists of mid-wave infrared (MWIR, 3-5 $\mu m$) and long-wave infrared (LWIR, 8-14 $\mu m$), is associated with thermal radiation emitted by objects. A thermal IR camera which detects radiation from the human face in both bands can produce a 2D image. Thermal radiation is influenced by vein patterns beneath the skin. Therefore, unlike visible images that only provide reflective information, thermal IR images can provide anatomical information that is less sensitive to outside variations.

However, thermal IR imagery has its limitations. For example, it is degraded by the presence of glass because glass blocks the energy emitted from an object. This partial loss of information presents difficulties when trying to recognize a person wearing eyeglasses or behind a piece of glass (e.g., sitting in a car). Figure 1 shows a pair of visible and thermal IR images of a person wearing eyeglasses. Thermal IR imagery is also influenced by variations in both body (e.g., after exercise or being outdoors for long periods of time) and ambient temperature. Figure 2 illustrates the effect of temperature

fluctuations on thermal IR images. Visible images, however, do not suffer from these limitations.



(a) Visible image (b) Thermal IR image

Figure 1: Influence of eyeglasses on thermal IR images.

(a) Room temperature (b) Outside temperature (c) After 10 minutes' exercise

Figure 2: Thermal IR images of a person with different body and ambient temperature.

Evaluation of new algorithms and systems for face recognition is also very important. A summary of current popular evaluation protocols is provided in [2].

Although different applications require different evaluation methods, these techniques can be grouped in terms of three tasks: identification, verification, and watch list [7]. In identification, the face recognition system provides a ranked listing of candidates whose faces best match with the input face from the unknown person. The performance is often described by the Cumulative Match Characteristic (CMC), which measures the rate at which the input image will be classified at rank *n* or better. In verification, a person presents his/her face to a face recognition system, and an identity claim is made. Then, the system either accepts or rejects the claim by comparing the input face with those stored in the database. In the process, two types of errors can occur: (1) false acceptance (the system falsely accepts the claim by an imposter) or (2) false rejection (the system falsely rejects the claim by the person who actually belongs to the system). The Receiver Operation Characteristic (ROC) is often used to quantify the performance of verification algorithms. The watch list can be viewed as a combination of both verification and identification. For this task, the face recognition system first determines if the individual is on the watch list. If they are, the system identifies the individual. The success rate of detecting and identifying an individual on the watch list are referred to as the detection and identification rate respectively. False recognition can have serious consequences, resulting in innocent people being detained and harassed.

## 1.2 Face Recognition in the Visible Spectrum

Studies in psychology suggest that human beings recognize a face using both holistic and local features. Therefore, face recognition algorithms in the visible spectrum are usually divided into holistic versus feature-based approaches.

### *1.2.1 Holistic Approaches*

Holistic methods use the whole face region as the input to the system. Principal component analysis (PCA), which provides a compact representation of face, is a widely used method for face representation and recognition. A vector $x$ can be expressed as linear combinations of an orthogonal basis $\Psi_i$: $x = \sum_{i=1}^{n} c_i \Psi_i \approx \sum_{i=1}^{m} c_i \Psi_i (m << n)$. The orthogonal basis $\Psi_i$ can be obtained by solving the following problem: $C\Psi = \Psi\lambda$, where $C$ is the covariance matrix for input vector $x$. Advantages of PCA include a compact representation of the original information, as well as reduced noise sensitivity.

The first successful application of PCA for face recognition is Eigenfaces [8]. In this system, every face in the database is projected onto the face space (eigenfaces) to obtain a vector of weights. When a face is presented to the system, it is transformed into a vector of weights by projecting it onto the eigenfaces. Recognition is then accomplished by searching for the image in the database whose weights are the closest to the weights of the probe face.

The eigenfaces method only reconstructs a face from a low dimensional basis, and ignores class-specific information. Therefore, this method may not be optimal in terms of discrimination. The Fisher's Linear Discriminant (FLD) has been proposed as a way to

solve this problem. This method maximizes the ratio of the determinant of the between-class scatter matrix of projected samples to the determinant of the within-class scatter matrix of projected samples. The within-class $S_w$ and between-class $S_b$ scatter matrices are computed as follows:

$$S_w = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - m_i)(x_k - m_i)^T \tag{1}$$

$$S_b = \sum_{i=1}^{c} N_i (m_i - m)(m_i - m)^T \tag{2}$$

where $c$ is the number of classes, $m_i$ is the mean image of class $X_i$, $m$ is the mean image of all sample images, $x_k$ is an image in class $X_i$, and $N_i$ is the number of samples in class $X_i$. Then, a matrix $W_{opt}$ is chosen as the matrix whose orthonormal columns maximize the ratio of the determinants:

$$W_{opt} = \arg\max_{W} \frac{|W^T S_b W|}{|W^T S_w W|} \tag{3}$$

With respect to face recognition, $S_w$ is always singular because the number of training samples is less than the number of pixels. In order to overcome the complication of a singular $S_w$, a method called Fisherface can be used. Fisherface uses PCA to reduce the dimension of the feature space, and then uses the standard FLD (Equation 2) to further reduce the dimension [9]:

$$W_{opt}^T = W_{fld}^T W_{pca}^T \tag{4}$$

$$W_{pca} = \arg\max_{W} |W^T S_T W| \tag{5}$$

$$W_{fld} = \arg\max_{W} \frac{|W^T W_{pca}^T S_b W_{pca} W|}{|W^T W_{pca}^T S_w W_{pca} W|} \tag{6}$$

where $S_T$ is the total scatter matrix of $N$ training images ($x_1, x_2, ....x_N$), and $m$ is the mean image of all training samples:

$$S_T = \sum_{k=1}^{N} (x_k - m)(x_k - m)^T \tag{7}$$

In addition to PCA and FLD, other popular matching methods include Independent Component Analysis (ICA), Support Vector Machines (SVM), Local Binary Patterns (LBP), and Active Appearance Model (AAM). ICA uses the higher-order statistical information generated from a set of basis vectors that possess maximum independence for representation and recognition [10]. SVM attempts to reduce the misclassification error by finding the optimal separating hyper-plane that maximizes the margin of separation [11]. LBP uses histograms in small regions for face representation and recognition [12]. Finally, AAM tries to reconstruct a new face using an integrated statistical model that combines both the shape and texture of a face [13].

### *1.2.2 Feature-based Approaches*

Feature-based approaches try to extract local features (e.g., eyes, nose, and mouth) for classification. One of the most successful algorithms is the Elastic Bunch Graph Matching [14]. This algorithm capitalizes on the idea that all human faces share a similar topological structure. Faces are represented by labeled graphs, with nodes positioned at fiducial points (eyes, noses and etc.). Each node is then described by several complex Gabor wavelet coefficients at different scales and rotations based on fixed

wavelet bases. These wavelet coefficients are robust to illumination changes, translations, distortions, rotations, and scaling. Recognition is based upon these labeled graphs and the distance between two nodes.

### 1.3 Face Recognition in the Thermal IR Spectrum

Face recognition in the visible spectrum is difficult in uncontrolled environments because it is especially sensitive to illumination variations. Thermal IR images, which capture the energy emitted from an object, are almost immune to illumination variations. Objects emit IR energy according to their temperature. Since human temperature is quite uniform (varying from 35.5 °C to 37.5 °C), it provides a consistent thermal IR signature. The thermal patterns associated with human faces are derived primarily from superficial blood vessels under the skin, which are unique to each individual [15]. It has been shown that even similar twins have different thermal patterns. Therefore, face recognition in the thermal IR spectrum tries to capitalize on these characteristics.

The eigenfaces approach has successfully been applied to face recognition in the thermal IR spectrum. It performs much better compared to when it is applied in the visible spectrum under different illumination conditions [16]. Besides the eigenfaces approach, Yoshitomi, et al. also proposed to identify thermal IR face patterns using neural networks. They combine the output from three neural networks using histograms, mosaic images, and shape factors of an image [17] .

Unfortunately, face recognition with thermal IR images suffers from a time lapse problem. Under relatively controlled illumination conditions, the performance of a face

recognition system using visible images outperforms that using thermal IR images when the probe image is acquired substantially later than the gallery images. In [18], the author use PCA-based recognition algorithm on a database that was collected over a 10-week period (one acquisition session per week). The results show that, in the same-session recognition, neither modality is significantly better than the other. However, if acquisitions of gallery and probe images are in different sessions, PCA-based recognition using visible-light images out-performed recognition using thermal IR images.

In an attempt to solve this "time lapse" phenomenon, the innate characteristics of the face under the skin is used for recognition [19]. The authors use a Bayesian framework to segment the face from the background. Then the superficial blood vessel network under skin is extracted by image morphology. The branching points of the skeletonized vascular network (referred as thermal minutia points) are used as features for classification. Experiments show obvious improvements using the database from the University of Notre Dame.

### 1.4 Face Recognition with Visible and Thermal IR Images

Information fusion is widely used in biometrics because the results generated from a single sensor may be unreliable. When looking at one object, different sensors can be used to provide different information. Information fusion tries to improve the performance of a system by integrating complementary information from different sources [20]. As shown before, for face recognition, visible and thermal IR images have their own advantages and disadvantages. For example, visible images are sensitive to

illumination variations, while thermal IR images are almost robust to illumination variations. However, thermal IR images are easily degraded by the presence of glass and variations in temperature, while visible images are more immune to these factors. Therefore, the fusion of visible and thermal IR images provides a way to improve the overall performance of a face recognition system.

Information fusion can be divided into three categories: data-level fusion, feature-level fusion, and score-level fusion [21]. Data fusion combines raw data from different sources to produce new data that is more informative than the original input. Feature fusion tries to combine various aspects of the feature space, like texture or shape information, eigenfaces, etc. Score fusion combines the outputs from different classifiers. Traditional methods include max rule, product rule, sum rule, weighted average rule, etc.

There have been some attempts to fuse visible and thermal IR images for face recognition. One approach involves applying a simple adaptive weighting scheme to fuse the score generated by two classifiers [22]. Genetic Algorithms have also been used to fuse images at the pixel level in both the wavelet domain and eigenspaces domain [23]. A weighted average technique has also been applied to fuse visible and thermal IR images at the data-level [24].

### 1.5 Research Objective

The objective of this research is to improve the performance of face recognition in uncontrolled situations. In order to address this problem, the fusion of visible and thermal IR images is studied. There are four contributions in this research: development

of a method of registering visible and thermal IR images in an uncontrolled environment (i.e., illumination variations, complicated backgrounds and occlusions); development of a method for reconstructing eyes occluded by eyeglasses in thermal IR images using images captured in the visible spectrum; development of a method to predict a novelty component from a visible image (thermal images are not available in many applications); and application of different methods (i.e., data-level, feature-level, and score-level fusion) to fuse visible and thermal IR images for face recognition.

The remainder of this thesis is dedicated to describing new methods for registering, analyzing, and fusing visible and thermal IR images. Chapter 2 proposes a new method called edge-based mutual information (EMI) to register visible and thermal IR images, which is a preliminary step for fusing two images. Chapter 3 introduces a novel approach to remove eyeglasses in thermal IR images. The visible and thermal IR images are then fused at the data-level, feature-level, and score-level. Chapter 4 introduces a method to predict a novelty component from the visible image. The novelty component is then fused with the visible image for face recognition. Chapter 5 evaluates the statistical significance of our experiment results, and Chapter 6 draws the conclusion of this thesis.

# CHAPTER 2

# IMAGE REGISTRATION

Visible and thermal IR images must be registered before they can be fused. Image registration is the process of transforming two images into the same coordinate system. Challenges for registering visible and thermal IR images include significant differences between the two under different illumination conditions, with different backgrounds, and in the presence of occlusions, as well as images taken under different parallaxes. In this chapter, a method called edge-based mutual information (EMI) is proposed to register visible and thermal IR images.

## 2.1 Advantages of Edge-based Mutual Information Image Registration

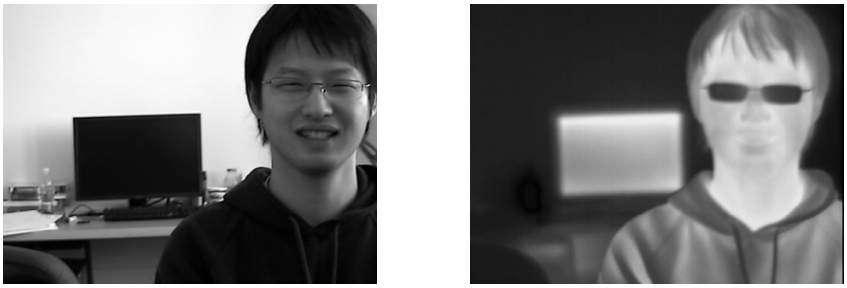
Image registration is the process of overlapping two or more images of the same scene that have been taken at different time, from different viewpoints, and/or by different sensors [25]. Although there exists a system that can produce co-registered visible and thermal IR images, this kind of system may not be readily available or may be too expensive [26]. Therefore, a software approach to register visible and thermal IR images is needed.

To register visible and thermal IR face images, Wang has used the 3D pose of the head, and the relationship between the two cameras [27]. Kong and Heo have designed a Gaussian criterion, and used the edge map of visible and thermal IR images for registration [24]. However, both of these methods can only be used for "clean" backgrounds. When a complicated background is present, feature extraction may not be reliable, and registration will be affected.

To register visible and thermal IR images, a novel method called edge-based mutual information (EMI), is proposed here. EMI is distinguished from other approaches insofar as EMI emphasizes the role of the face in the registration process. This is accomplished by realizing that the foreground and background images will lie on different planes because they are located at different distances from the cameras. Due to different parallaxes, the same scene will appear quite different, even if the cameras are side-by-side. As shown in Figure 3, a face in front of a monitor is captured by a pair of visible and thermal IR cameras that are located side-by-side (Figure 4). In the thermal IR image, we can see that the monitor is next to the face, while in the visible image, it is a distance from the face. This apparent distance discrepancy makes it difficult to find a global transformation that can register both the foreground (face) and the background (monitor) at the same time.

Previous registration approaches have only considered objects in the foreground, and require that the image be captured in front of a clean background. Therefore, those methods are more likely to fail when attempting to register a face in front of complicated backgrounds. EMI, however, can recognize a face in front of a complicated background

and assign more weight to it in the registration process; this helps to ensure that the face is accurately registered. EMI is also robust to illumination variations because it is an intensity-based method, which does not require accurate extraction of facial features. Furthermore, EMI uses image rectification techniques to limit differences between two images to the horizontal axis. This makes it faster to search for the parameter for registration, and avoids the problem of local maximum in the optimization process.



(a) Image captured by visible camera (b) Image captured by thermal IR camera

Figure 3: Different parallaxes.

Figure 4: A pair of visible and thermal IR cameras.

## 2.2 Image Registration Algorithm

A flowchart of the EMI image registration algorithm is presented in Figure 5. First, the face is detected in the visible spectrum using a face detector introduced by Viola and Jones [28]. Then the Canny edge detector is used to extract edges in the face region of the visible image, as well as in the thermal IR image [29]. Pixels that are on the edge are weighted higher in the registration process. This emphasizes the face part in the registration process, which addresses the problem of different parallaxes. It also introduces spatial information into MI, which makes the similarity measure more robust. After the face is detected, both the visible and thermal IR images, as well as their edge maps, are rectified. Image rectification simplifies the optimization process and avoids the problem of reaching a local maximum in the process of searching for transformation parameters. After image rectification, the difference between two images has been confined to the horizontal axis, and the search for the transformation parameter can be limited to one dimension (i.e., the horizontal axis of the image). In the process of searching for this parameter, the EMI objective function is calculated. The parameter with the maximum EMI value is then used to register visible and thermal IR images.

## 2.3 Face Detection

Face detection is used to segment the face from complicated backgrounds in order to ensure accurate registration. To detect the face, we use the method proposed by Viola and Jones, which is well known for its high detection rate and fast speed [28]. The

method uses Haar-like features and the Adaboost algorithm to detect the face. One example of face detection in the lateral illumination condition is shown in Figure 6.

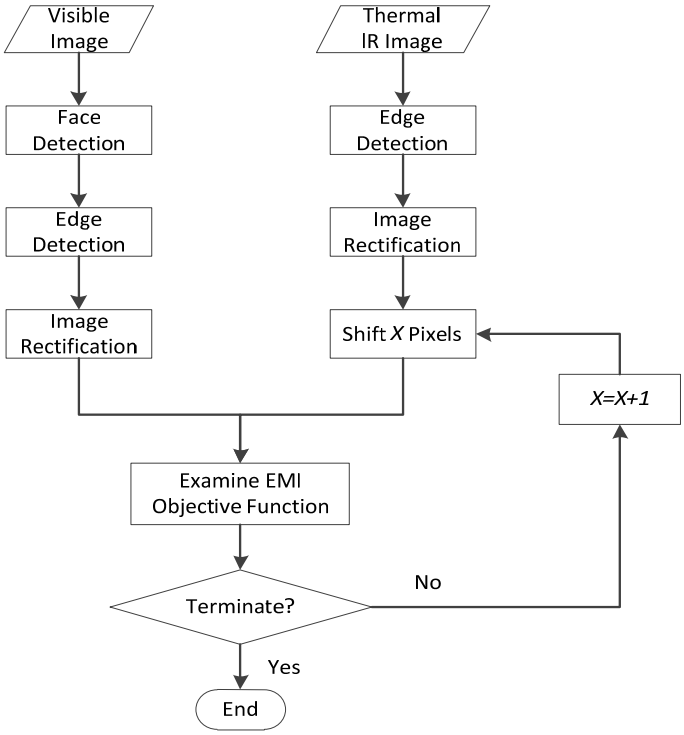



Figure 5: Edge-based mutual information image registration algorithm.

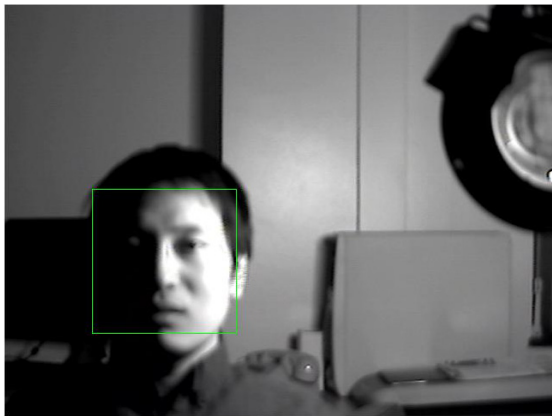


Figure 6: Detected face in the lateral illumination condition.

## 2.4 Image Rectification

Given a pair of images, image rectification is considered to be the transformation of the image planes that make the conjugate epipolar lines of two images become parallel to the horizontal image axis. This enables the disparity-matching to be confined to single-dimensional searching [30].

Before two images can be rectified, it is necessary to describe the intrinsic and extrinsic parameters of the visible and thermal IR cameras. The intrinsic parameter is a matrix that consists of a camera's horizontal ( $\alpha$ ) and vertical ( $\beta$ ) focal lengths, the coordinates of the principle point $(u_0, v_0)$, and a skew coefficient ( $s$ ):

$$A = \begin{bmatrix} \alpha & s & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{8}$$

The extrinsic parameter is another matrix $[R\ t]$, where $R = [r_1\ r_2\ r_3]$ is a $3 \times 3$ rotation matrix, and $t$ is a $3 \times 1$ translation vector. $R$ and $t$ can be related using the coordinate of the optical center of the camera $C$:

$$R = [r_1\ r_2\ r_3] \tag{9}$$

$$t = -RC \tag{10}$$

To obtain these parameters, standard techniques (i.e., camera calibration) can be used. After obtaining camera 1's extrinsic and intrinsic parameters ( $A_1, R_1, t_1$ ) and camera 2's extrinsic and intrinsic parameters ( $A_2, R_2, t_2$ ), image rectification can be performed. This process will be demonstrated for the images captured by camera 1. As shown in Figure 7, a spatial point $M = [X, Y, Z, 1]^T$ is projected onto the camera's image plane as

point $m_1 = [u, v, 1]^T$ using a projection matrix $P_1$. Image rectification finds a transformation matrix $T_1$ that transforms the original projection matrix $P_1$ into a new projection matrix $P_1^{'}$. The new projection matrix $P_1^{'}$ projects the spatial point $M$ to another point $m_1^{'}$, which lies on the plane that is parallel to the base line $C_1C_2$. The original transformation matrix $P_1$ can be obtained from the intrinsic and extrinsic parameters of a camera as shown in Equation (11). After image rectification, the new $X$ axis of the two images is parallel to the baseline $C_1C_2$. Therefore, the new transformation matrix $P_1^{'}$ can be expressed as:

$$P_1 = A_1[r_1 \;\; r_2 \;\; r_3 \;\; t_1] \tag{11}$$

$$P_1^{'} = A^{'}[r_1^{'} \;\; r_2^{'} \;\; r_3^{'} \;\; t_1^{'}] \tag{12}$$

where $r_1^{'}, r_2^{'}, r_3^{'}$ are vectors that represent the $X,Y,Z$ axes of the camera after rectification. The vector $r_1^{'} = (\mathrm{C}_1 - C_2) / \left\| C_1 - C_2 \right\|$ represents the new $X$ axis parallel to the baseline $C_1C_2$. The vector $r_2^{'} = k \times r_1^{'}$ means that the new $Y$ axis is orthogonal to $k$ and $r_1^{'}$. The vector $k$ is an arbitrary unit vector, which can be chosen as the $Z$ unit vector of original coordinate for convenience. Finally, $r_3^{'}$ is orthogonal to $r_1^{'}$ and $r_2^{'}$, which can be expressed as $r_3^{'} = r_1^{'} \times r_2^{'}$. Because $m_1$ and $m_1$' are associated with the same line $MC_1$, we can express $MC_1$ as a set of spatial points $w$ that satisfies the equation (13) and (14).

$$w = C_1 + \lambda_1 (A_1[r_1 \;\; r_2 \;\; r_3])^{-1} m_1 \tag{13}$$

$$w = C_1 + \lambda_1^{'} (A^{'}[r_1^{'} \;\; r_2^{'} \;\; r_3^{'}])^{-1} m_1^{'} \tag{14}$$

$$m_1' = \lambda\,(A'[r_1'\ \ r_2'\ \ r_3'])\,(A_1[r_1\ \ r_2\ \ r_3])^{-1} m_1 \tag{15}$$

Then the rectifying transformation matrix for images taken from camera 1 is:

$$T_1 = (A'[r_1'\ \ r_2'\ \ r_3'])\,(A_1[r_1\ \ r_2\ \ r_3])^{-1} \tag{16}$$

The same method can be applied for finding rectifying transformation matrix for images taken from camera 2.

Both camera calibration and image rectification are implemented using the Camera Calibration Toolbox [31]. For camera calibration, the toolbox uses control points on input images that feature a visible calibration board. For visible images, this calibration board can be generated simply by placing a checkerboard pattern in front of the camera and then taking a picture. For thermal IR images, 48 ($8 \times 6$) nails are placed on the corner of the checkerboard's grid. Then, the board is heated with a hair dryer so that the 48 control points become visible in thermal IR images. Examples of the checkerboard can be seen in Figure 8. Figure 9 shows example of image rectification. After rectification, the difference between visible and thermal IR images is confined to the horizontal axis.

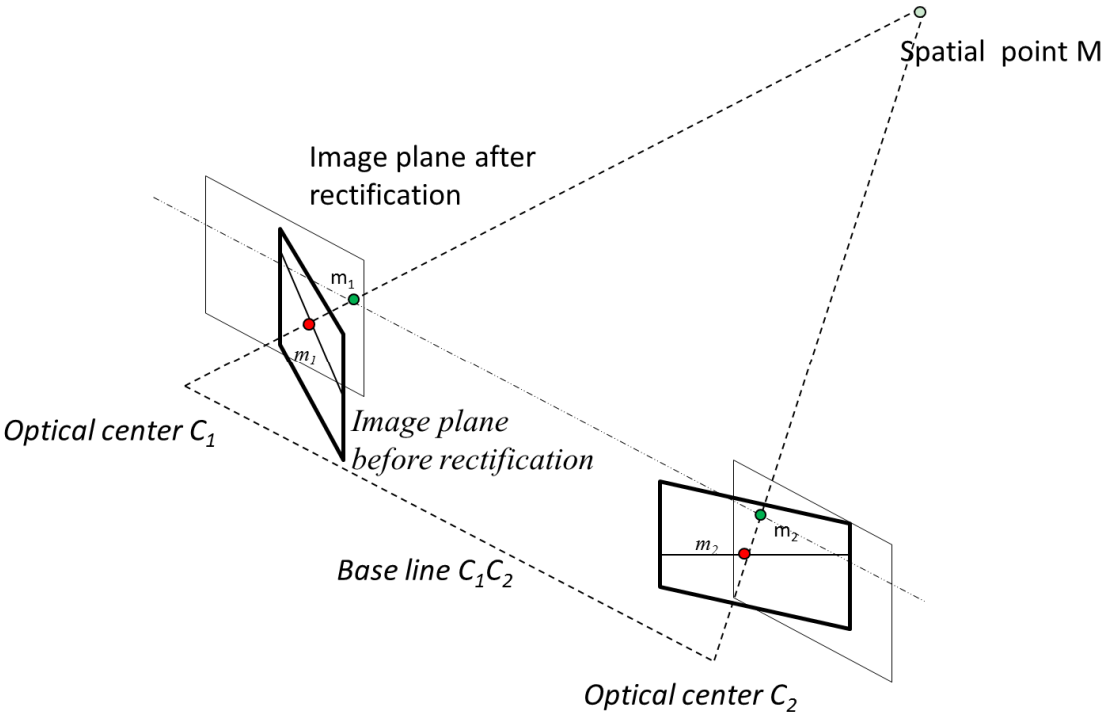
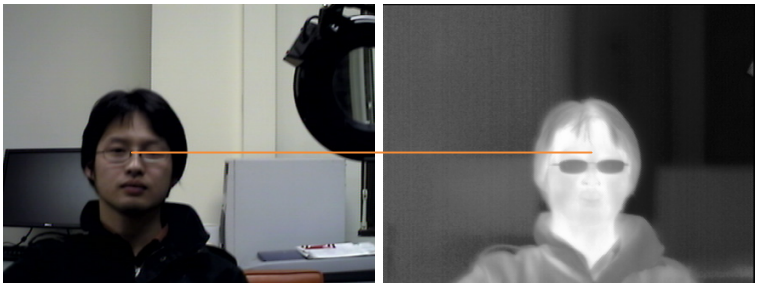


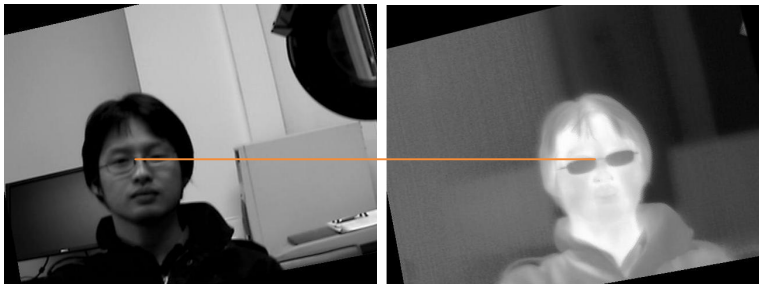

Figure 7: Image rectification.

(a) Visible image (b) Thermal IR image

Figure 8: Checkerboard for camera calibration.



(a) Before rectification



(b) After rectification

Figure 9: Differences of two images before and after rectification.

## 2.5 Edge-based Mutual Information Objective Function

In information theory, mutual information (MI) measures the degree of dependence between two random variables [32]. Assume that two random variables $A$ and $B$, with events $a$ and $b$, have marginal probability $p(a)$ $p(b)$, and joint probability $p(a,b)$. MI measures the degree of dependence between $A$ and $B$ by calculating the distance between the joint probability $p(a,b)$ and $p(a)p(b)$ (i.e., the case when $A$ and $B$ are statistically independent) using the Kullback-Leibler measure [32]:

$$MI = \sum_{a,b} p(a,b) \log(\frac{p(a,b)}{p(a)p(b)}) \tag{17}$$

If we consider two images as random variable $A$ and $B$ and their intensities as events $a$ and $b$, then MI can be considered as a measure of dependence between the two images. The marginal probability and joint probability of both images can be estimated using a histogram and joint histogram of the two images. When two images of the same object are spatially aligned, the dependence between them is at a maximum, and their MI is also at a maximum. Therefore, by maximizing MI, a transformation matrix can be found to register the two images.

One problem that influences the robustness and accuracy of MI is that it only assumes statistical information and ignores spatial information. EMI solves this problem by including edge information. EMI can also accurately register a face in front of a complicated background by assigning more weight to the face in the registration process.

EMI takes advantage of the fact that different pixels, even of the same intensity, have different roles in registering two images. For example, pixels on the face may have the same intensity as those in the background. However, because we are more interested in the face, pixels on the face should play a more important role in the process of registration compared to those in the background. But, when we use the histogram and joint histogram to estimate the marginal probability and joint probability of an image, they are treated equally. EMI overcomes this problem by using edge information. Although the difference between visible and thermal IR images is large, edge information in both images has more information in common because it captures the shape and features (nose, mouth, etc.) of the face. When two images are spatially aligned, the

degree to which the edges overlap is also at the maximum. Based on this, an edge coefficient can be incorporated into the MI.

In order to obtain the EMI objective function, the Canny edge detector is used to determine an edge in the thermal IR image and in the face-region of the visible image [29]. Let $A$ and $B$ denote the overlapping areas of the visible and thermal IR images, $a_1, b_1$ denote the edge points, and $a_0, b_0$ denote non-edge points in $A$ and $B$. Then, the joint probability is calculated by constructing four joint histograms between $a_1\, b_1$, $a_1\, b_0$, $a_0\, b_1$, and $a_0\, b_0$. By dividing the four joint histograms by the total number of pixels in the overlapping areas, the following four joint probabilities can be obtained: $p(a_1 b_1)$, $p(a_1 b_0)$, $p(a_0 b_1)$, $p(a_0 b_0)$. The marginal probability of image $A$ and image $B$, denoted as $p(a_1)$, $p(b_1)$, $p(a_0)$, and $p(b_0)$, can also be estimated from the histograms of the two images. Following that, the EMI can be obtained using the following equations:

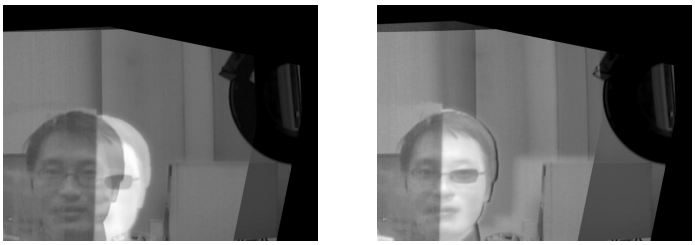
$$EMI = \sum_{a_i, b_j} u(a_i, b_j)\, p(a_i, b_j) \log\left(\frac{p(a_i, b_j)}{p(a_i)\, p(b_j)}\right) \tag{18}$$

$$u(a_i, b_j) = \begin{cases} s \ (s>1) \ \text{if } i=j=1 \\ 1 \ \text{if } i=0 \text{ or } j=0 \end{cases} \tag{19}$$

When $i=1, j=1$, the two corresponding points $a_i, b_j$ are both edge points. When this occurs, more emphasis is placed on this combination by assigning a bigger coefficient $s$ to $u(a_i, b_j)$, which is called the edge coefficient. In our experiment, we chose $s$ as 100. When either $i=0$ or $j=0$, it means that either a point of $A$ or $B$ is not edge point, and both are treated equally by assigning 1 to $u(a_i, b_j)$.

Results of image registration using a partially occluded face (Figure 10), under the lateral illumination condition (Figure 11), and with complicated background (Figure 12) are presented. By using EMI as an objective function, we can register face images more accurately compared to MI.

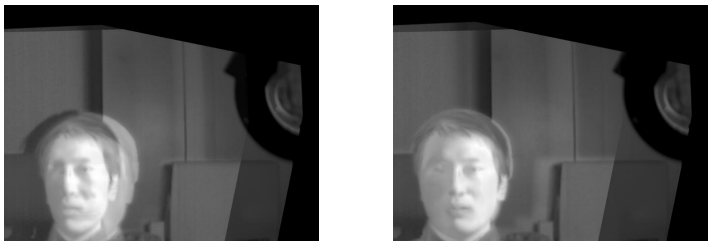
(a)Visible image (b) Thermal IR image (c) Before registration



(d) Registration using MI (e) Registration using EMI

Figure 10: Image registration with partially occluded face.

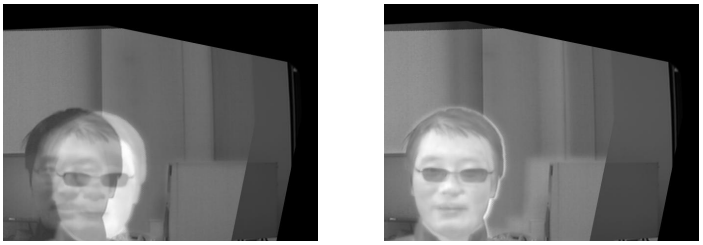
(a)Visible image (b) Thermal IR image (c) Before registration



(d) Registration using MI (e) Registration using EMI

Figure 11: Image registration under the lateral illumination condition.

(a)Visible image (b) Thermal IR image (c) Before registration



(d) Registration using MI (e) Registration using EMI

Figure 12: Image registration with complicated backgrounds.

# CHAPTER 3

# FUSION OF VISIBLE AND THERMAL IR IMAGES

After registering visible and thermal IR images, fusion techniques can be applied to integrate complementary information for the purpose of classification. However, considering the effects of eyeglasses on thermal IR images, it is necessary to remove any occlusions prior to fusion. In this chapter, a novel approach is proposed to remove eyeglasses in the thermal IR image. Different fusion techniques are then applied to the visible and thermal IR images.

With respect to replacing eyeglasses in thermal IR images, an input face image is first classified as either with or without eyeglasses using a support vector machine (SVM) classifier. If the input face contains eyeglasses, the position of the eyeglasses is segmented by thresholding based upon the temperature difference between eyeglasses and the face. Then, the eyes in the thermal IR image are reconstructed using information from the visible eye information using a neural network model.

In order to fuse the visible and thermal IR images, fusion on data-level, feature-level and score-level are applied to the visible and thermal IR images. Experimental results show that facial recognition results are improved when visible and thermal IR images are fused versus using either image independently.

### 3.1 Eyeglasses Detection in Thermal IR Images

In order to replace eyeglasses in thermal IR images, it is necessary to first detect the location of eyeglasses. An SVM classifier can be trained to determine whether or not a person is wearing eyeglasses. SVM is a supervised learning method widely used in pattern recognition [33]. If the person is wearing eyeglasses, thresholding is applied on the face to segment eyeglasses. Considering the significant temperature difference between eyeglasses and the skin, this segmentation is greatly simplified. The diagram of this algorithm is described in Figure 13. The face is first detected in the visible image using the Viola/Jones face detector [28]. Once the visible and thermal IR images are registered, we use the face location in the visible images to obtain the face position in the thermal IR images. Each thermal IR face is then normalized to 32×32 pixels, and this is further converted into a long feature vector (1×1024). These feature vectors are then used to train a SVM classifier that separates face images into two categories—those with eyeglasses and those without eyeglasses. If an image is classified as with eyeglasses, thresholding is applied to the image to segment the eyeglasses.

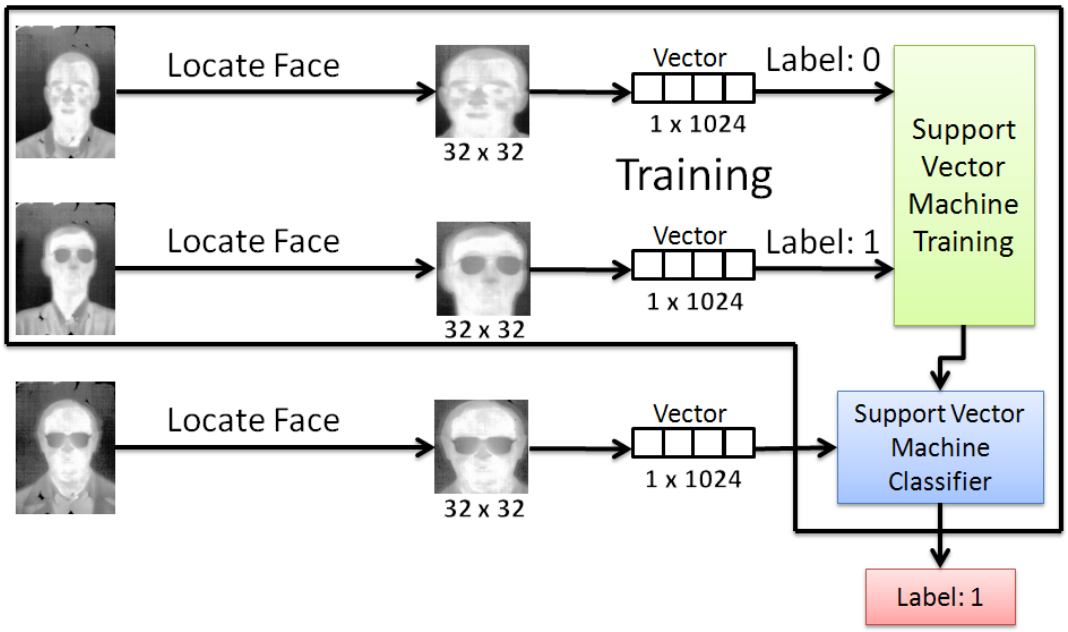



Figure 13: Diagram of eyeglass detection algorithm.

### 3.2 Eyeglasses Replacement in Thermal IR Images

After eyeglasses are detected in the thermal IR images, they are reconstructed using the eyes in the visible image. This is accomplished by assuming that, since the visible and thermal IR images are of the same person, some relationship might exist between the visible and thermal IR eye regions. If this relationship can be described modeled, it is possible to predict the lost information.

The relationship between visible and thermal IR images is highly complex and difficult to describe using any closed-form mathematical model. Therefore, a nonlinear neural network model can be used. The neural network can be trained to learn any nonlinear input-output relationship from a set of training data [34]. In the training

process, the internal weights of a neural network are adjusted to minimize the estimation error over a set of examples. For this application, we selected to use a multi-layer feed-forward neural network. The neural network was trained to learn the relationship between the eyes in thermal IR images and those in visible images.

A diagram of proposed algorithm is presented in Figure 14. Eye regions are first extracted from visible training images with eyeglasses and thermal IR training images without eyeglasses using the visible eye coordinates. Then, principal component analysis (PCA) is performed on both of the extracted eye regions for dimension reduction. The purpose of the PCA operation is to represent each eye region as a low dimensional feature vector, as well as to obtain eigenvectors from which thermal IR eyes can be reconstructed. The procedure for using PCA for dimension reduction and reconstruction is described as follows:

**Dimension Reduction using PCA:**

- obtain images $I_1, I_2, ..., I_M$ (training images)
- represent every image $I_i$ as a vector $\Gamma_i$
- compute the average image vector $\Psi$ :

$$\Psi = \frac{1}{M}\sum_{i=1}^{M}\Gamma_i \tag{20}$$

- subtract the mean face from each vector to get the normalized face:

$$\Phi_i = \Gamma_i - \Psi \tag{21}$$

- compute the covariance matrix $C$ :

$$C = \frac{1}{M}\sum_{i=1}^{M}\Phi_i\Phi_i^T = AA^T \tag{22}$$

$$A = [\Phi_1, \Phi_2 ..... \Phi_M] \tag{23}$$

- compute the eigenvectors $u_j$ of $C$
- keep only *K* eigenvectors (corresponding to the *K* largest eigenvalues)
- project each training face vector onto eigenspace:

$$\Phi_i = \sum_{j=1}^{K} w_j u_j, (w_j = u_j^T \Phi_i) \tag{24}$$

- each normalized training face $\Phi_i$ is represented in this basis by a vector:

$$\Omega_i = \begin{bmatrix} w_1^i \\ w_2^i \\ \dots \\ w_K^i \end{bmatrix}, i = 1, 2 ... M \tag{25}$$

**Image Reconstruction:**

- using eigenvectors $u_j$, the weight $w_j$ and the average image $\Psi$ to approximate the original image:

$$\Gamma_{estimated} = \sum_{j=1}^{K} w_j u_j + \Psi \tag{26}$$

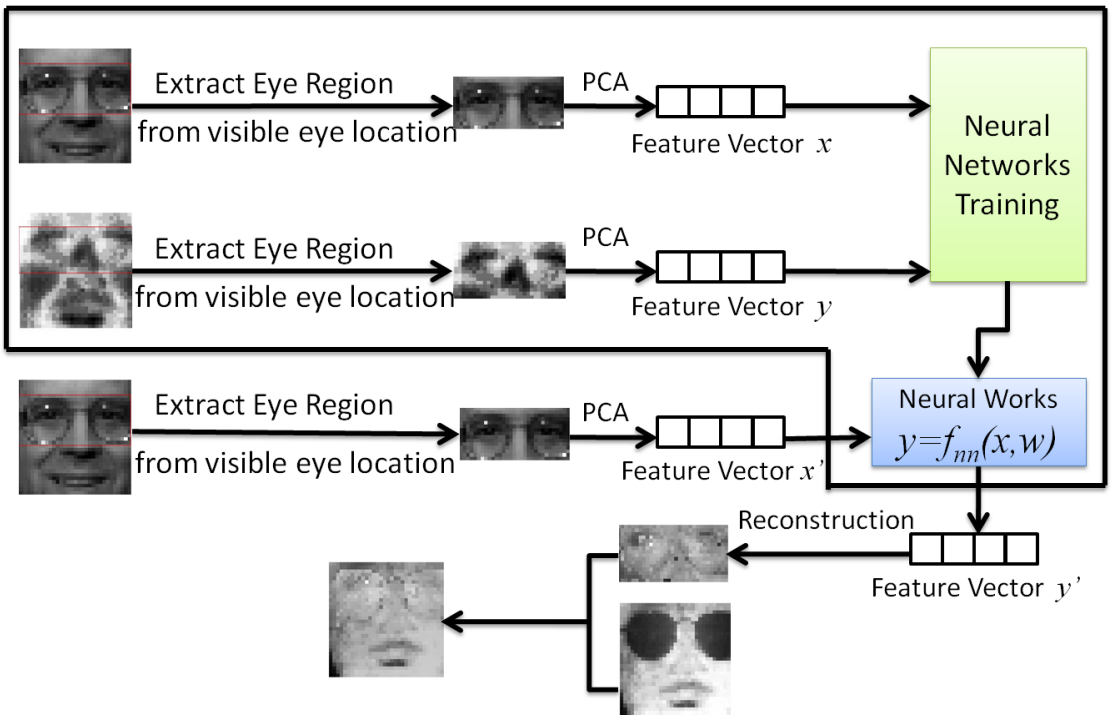



Figure 14: Diagram of eyeglass replacement algorithm.

After each image has been represented by a lower dimensional vector, these are used as inputs to train a neural network. A neural network that describes the relationship between visible and thermal IR eye regions can be obtained after training. When a pair of visible and thermal IR images is first presented, the visible eye region is extracted and represented by a vector using PCA. This vector is presented to the trained neural network. The trained neural network is then used to predict the feature vector from which the thermal IR eye regions can be reconstructed. The reconstructed eye regions are then used to replace eyeglasses in thermal IR images.

The proposed algorithm was tested on the NIST/Equinox database (see Table 1). The NIST/Equinox database consists of 1619 pairs of registered visible and thermal IR

images from 88 individuals [29]. Visible and thermal IR images were registered to within 1/3-pixel accuracy using a hardware setting. The spatial resolution is 320 × 240 pixels, and the gray-scale resolution is 8 bits for visible images and 12 bits for thermal IR images. The database was collected under varying illumination conditions, and subjects were asked to present different expressions and wearing or not wearing eyeglasses. The database was divided into three parts: Gallery, Probe, and Training. Gallery denotes images in the database with known identity (i.e., one image for each person was taken with frontal lighting condition and a neutral facial expression). "Probe" indicates the images presented to the system for identification. Probe images were divided according to the conditions under which the image was captured (i.e., the presence of eyeglasses, lighting conditions, and facial expression). "Training" refers to those images used for training the SVM classifier to identify eyeglasses, the neural network for predicting thermal IR eye regions, and the eigenfaces for face recognition. The training images contained 30 people who did not exist in either the "Probe" or "Gallery" images. Although this setup may reduce the prediction accuracy, it is useful to eliminate any bias that might be introduced into the face recognition algorithm, as well as make the experimental result more objective.

For this database, the images were preprocessed by first cropping the face from the original image based upon the eye coordinates. The eye coordinates were obtained by manually pointing to the location of the eyes on the visible image. Then, the image was resampled to 32 × 32 pixels using bilinear interpolation. Histogram equalization was then performed on the face region to reduce the effect of variations on the image.

In the experiment, an SVM classifier with an RBF kernel was trained to classify an input face as being with or without eyeglasses. A feed-forward neural network was constructed with two hidden layers (6 nodes per layer), and a log-sigmoid transfer function. This neural network was then trained to transform the visible eyeglasses region into the thermal IR eye regions. Examples of training samples can be seen from Figure 15.

When a new face was presented to the system, it is first fed into the trained SVM classifier. If it was classified as a face with eyeglasses, the eyeglass regions were converted into a lower dimension vector (dimension 10) using PCA. This vector was then fed into the neural network to predict a new vector. Thermal eye regions were then reconstructed from the predicted vector, and this was used to replace eyeglasses on the input face. Examples of faces after replacing eyeglasses can be seen from Figure 16.

In order to evaluate the benefit of the eyeglass replacement algorithm on face recognition using thermal IR images, the probe sets with eyeglasses (probe 1, 2, 5, 6) were used as testing images. One of the popular face recognition algorithms (i.e., Eigenfaces) was used [8]. For this algorithm, the dimension of the feature vector was chosen as 500, and the whitened cosine distance was used for matching because it has previously been shown to work well [31]. Figure 17 shows the top 10 matches and Figure 18 shows the ROC curve for face verification. As can be seen from these figures, identification accuracy was improved by replacing the eyeglasses.

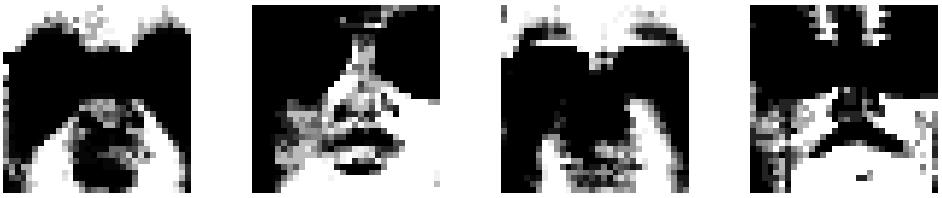
Table 1. NIIST/Equinox database of visible and thermal IR images

| Dataset | Visible (Thermal) | Eyeglasses | Lighting | Expression |
|---|---|---|---|---|
| Gallery | 58 (58) | Off | Frontal | Neutral |
| Probe 1 | 27 (27) | On | Frontal | Neutral |
| Probe 2 | 85 (85) | On | Frontal | Various |
| Probe 3 | 36 (36) | Off | Frontal | Neutral |
| Probe 4 | 165 (165) | Off | Frontal | Various |
| Probe 5 | 57 (57) | On | Lateral | Neutral |
| Probe 6 | 139 (139) | On | Lateral | Various |
| Probe 7 | 121 (121) | Off | Lateral | Neutral |
| Probe 8 | 349 (349) | Off | Lateral | Various |
| Training | 582 (582) | On+Off | Frontal+Lateral | Neutral+Various |

(a) Thermal IR images

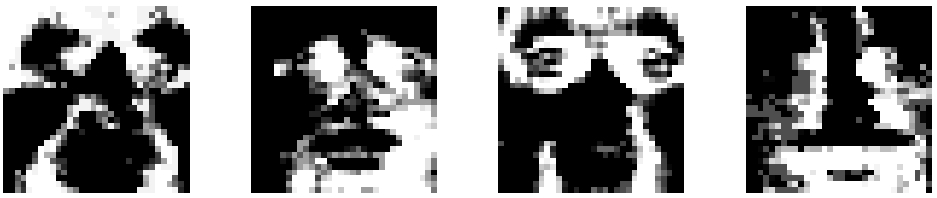
(b) Visible images

Figure 15: Examples of training images.



(a) Images before replacing eyeglasses

(b) Images after replacing eyeglasses



(c) Thermal IR Images of the corresponding people without eyeglasses

Figure 16: Examples of images before and after eyeglasses replacement.

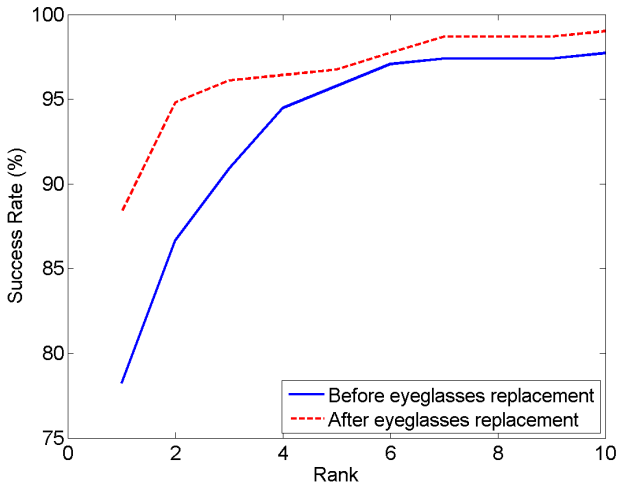



Figure 17: Top 10 matches before and after eyeglasses replacement.

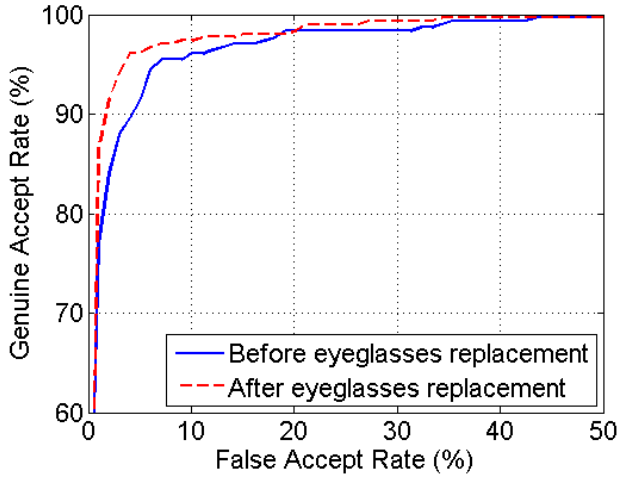



Figure 18: ROC curve for face verification.

### 3.3 Fusion of Visible and Thermal IR Images for Face Recognition

After removing any occlusions caused by eyeglasses in the thermal IR images, fusion techniques can be applied to integrate complementary information from both the visible and thermal IR spectrums. In order to fuse the images, three techniques were explored: data-level fusion, feature-level fusion, and score-level fusion.

For fusion at the data-level, a multi-scale fusion scheme with a discrete wavelet transform (DWT) is used. Both images are transformed into a set of approximation coefficients $W_{\varphi}(m,n)$ and details coefficients $W_{\Psi}(i,j)$ using DWT. By using a weighted average combination of visible and thermal IR wavelet coefficients, the wavelet coefficients of the fused image can be obtained from the following:

$$W_{\varphi}^{Fused}(m,n) = \alpha_1 W_{\varphi}^{Visible}(m,n) + \beta_1 W_{\varphi}^{Thermal}(m,n) \tag{27}$$

$$W_{\psi}^{Fused}(i,j) = \alpha_2 W_{\psi}^{Visible}(i,j) + \beta_2 W_{\psi}^{Thermal}(i,j) \tag{28}$$

In order to determine the weight coefficients, a global optimization method called particle swarm optimization (PSO) is used. PSO is a population-based stochastic optimization technique that is inspired by the social behavior of birds flocking or fish schooling [35]. PSO learns the weights of the visible and thermal IR images from a set of training images by maximizing an objective function (i.e., the top matching rate). After combining the weighted coefficients from the visible and thermal IR images, a fused image is reconstructed using inverse DWT (*IDWT*):

$$I_{Fused}(x,y) = IDWT[W_{\psi}^{Fused}(i,j), W_{\varphi}^{Fused}(m,n)] \tag{29}$$

For fusion at the feature-level, both visible and thermal IR images are first converted to long vectors. Then, fusion is applied by concatenating them together to generate a fused-feature vector (Figure 19). This fused feature vector is then used as the input to the face recognition algorithm for matching.

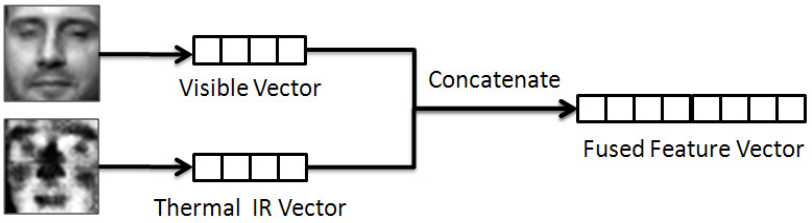



Figure 19: Feature level fusion of visible and thermal images.

For fusion at the score-level, visible and thermal IR images are sent to the face recognition algorithm separately in order to generate matching scores. Then, a fused score can be generated using different rules. In this research, the max rule (Equation (30)), the sum rule (Equation (31)), and the product rule (Equation (32)) are used:

$$Score_{Fusion} = Max(Score_{Visible}, Score_{Thermal}) \tag{30}$$

$$Score_{Fusion} = Score_{Visible} + Score_{Thermal} \tag{31}$$

$$Score_{Fusion} = Score_{Visible} \times Score_{Thermal} \tag{32}$$

In order to test the performance of the methods used to fuse the visible and thermal IR images, the NIST/Equinox database and Eigenfaces algorithm were again used. For this experiment, we used all of the probe images for testing (probe 1-8). For the thermal IR images, the eyeglasses were first removed the using algorithm introduced

before, and the image was then fused with the visible image. The performance of the face recognition algorithm using fused images was compared with using each image type independently. Figure 20 shows the top 10 matches for face identification using visible images, thermal IR images, and different fusion techniques. Figure 21 shows the ROC curve for face verification. As can be seen from these graphs, face recognition with fusion techniques consistently outperforms face recognition with only a single image.

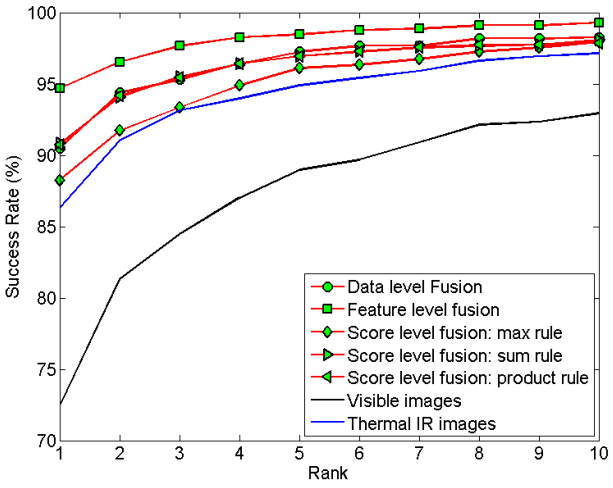



Figure 20: Top 10 matches for face identification.

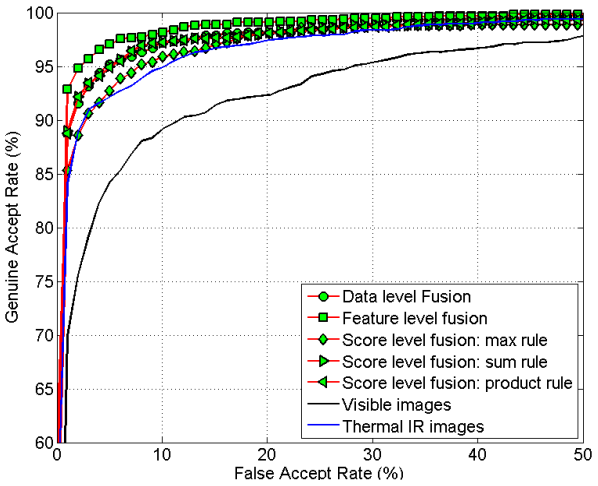



Figure 21: ROC curve for face verification.

# CHAPTER 4

# FACE RECOGNITION WITH A NOVELTY COMPONENT

In the previous chapter, it was shown that the fusion of visible and thermal IR images can improve the performance of face recognition algorithms. However, fusion techniques require both a visible and thermal IR camera, the later which may not be readily available considering the high cost and availability of such technology. On the contrary, visible cameras are easily accessible. Motivated by this, a method to predict a novelty component from the visible image is proposed in this chapter. The novelty component is the corresponding thermal IR component of a visible image. By fusing visible images and the novelty component, the performance of face recognition can be improved.

## 4.1 Algorithm Description

Although fusion of the visible and thermal IR images can improve the performance of a face recognition system, it also requires the presence of two types of

cameras. Thermal IR cameras may not be readily available because they are relatively expensive. On the other hand, visible cameras are almost everywhere. Motivated by this, a method to predict a thermal IR component from a visible image is proposed here. This thermal IR component is called the novelty component of an image. It is shown that by fusing this novelty component with a visible image, the performance of face recognition using only visible images can be improved.

In order to transform between the visible and thermal IR images, the relationship between them must be interdependent. Considering that the visible and thermal IR sensors are detecting the same person, this condition is met. Furthermore, research has shown that sensors can be considered interdependent if the level of uncertainty between an object and a system can be reduced when using all of the sensors together versus using only one sensor [36]. By fusing the visible and thermal IR images, the performance of the face recognition system was improved. In this case, the improvement in performance can be viewed as a reduction in uncertainty between the person and the system. Therefore the relationship between visible and thermal IR cameras is interdependent. Registration of the visible and thermal IR images also supports the interdependent relationship between them (i.e., this process maximizes the interdependence between the images [37]).

In the literature, relatively few attempts have been made to convert between visible and thermal IR images. Reiter et al. uses canonical correlation analysis (CCA) to

generate near IR (NIR) images from the visible images [38]. The NIR images capture the reflected photons from subjects, which is similar to how visible images work. Therefore, the correlation between visible and NIR images is higher than between visible and thermal IR images, which makes it easier for image mapping. Jun et al. uses CCA and Markov random fields to generate a visible face from a thermal IR face [39]. However, when using the generated image for face recognition, the performance is not satisfactory. The best performance they reported is only 50.06% accuracy.

The method for generating the novelty component from a visible image is described in Figure 22. The novelty component is the thermal IR image that has been predicted from the visible image. After that, a fused image is generated by fusing the visible image with its novelty component in the discrete wavelet transform (DWT) domain. This fused image is used for face recognition.

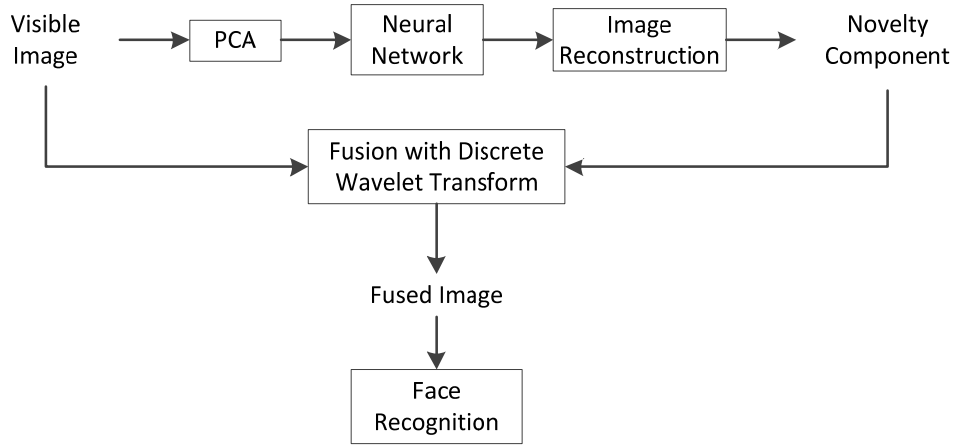



Figure 22: Diagram of face recognition using novelty component.

## 4.2 Novelty Component of a Visible Image

The relationship between visible and thermal IR images is highly complex and difficult to describe using a closed-form mathematical model. Therefore, an artificial neural network was selected to learn this relationship. Artificial neural networks, which are inspired by the fact that the human brain can learn from observations and generalize by abstraction, are ideal for modeling non-linear, highly complex relationships. It has been proven that any Boral measurable function can be approximated by a multi-layer feed-forward artificial neural networks to any desired degree of accuracy [34].

In order to describe the relationship between visible and thermal IR images, a neural network was built to transform visible information into thermal IR information:

$$t = f_{NN(v-t)}(v, c_{(v-t)}) \tag{33}$$

Here, $v$ and $t$ denote vectors that represent visible and thermal IR images separately, and $c_{(v-t)}$ denotes a parameter vector of internal connections of the neural network that can transform visible information to thermal IR information.

In order to obtain such a neural network, coregistered visible and thermal IR images are used for training. The visible and thermal IR images must be registered in order to maximize the overlapped information between them. The predicted image can be co-registered with the input image, which is necessary for fusion [40]. The training process for finding such a neural network is described in Figure 23. First, PCA is performed on visible and thermal IR training images. Given a set of *N* training images $\{I_1, I_2, I_3 ... I_N\}$ all in $R^n$, PCA finds a linear transformation $W^T$, mapping the original $n$ dimensional image $I_k$ into an $m$ dimensional feature space $v_k$, where $m < n$:

$$v_k = W^T I_k, k = 1, 2, ..., N \tag{34}$$

$W$ is chosen as $W_{opt}$ which maximizes the determinant of the total scatter matrix of projected samples:

$$W_{opt} = \arg\max_{W} |W^T S_T W| = [w_1, w_2, ..., w_m] \tag{35}$$

$$S_T = \sum_{k=1}^{N} (I_K - \mu)(I_K - \mu)^T \tag{36}$$

where $w_i$ $(i = 1, 2, ..., m)$ are $n$ dimensional eigenvectors of $S_T$ corresponding to $m$ largest eigenvalues, which are also called eigenfaces. $N$ is the number of training images, $\mu \in R^n$ is the mean image of training images. After applying PCA, each image will be represented as an $m$-dimensional vector, where $m$ is smaller than the original dimension of the image.

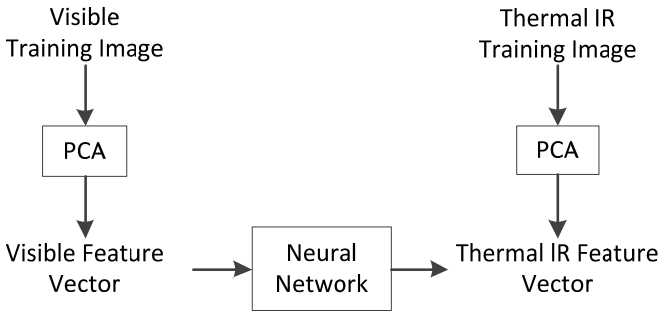



Figure 23: Training neural network.

For obtaining a neural network that can transform a visible vector to a thermal IR vector, the visible vector $v$ is used as the input to the neural network, and the

corresponding thermal IR vector $t$ is used as desired output. The output of each layer of the neural network is computed using a nonlinear activation function, which is comprised of a weighted sum of inputs from the previous layer:

$$y_i = h(\sum_{k=1}^{n} c_{ik} v_k + c_{i0}) \tag{37}$$

where $h$ denotes a nonlinear activation function, such as the logistic sigmoid function.

During the training process, the neural network tries to minimize the estimation error by adjusting the internal connections (or weights) using several training samples. Traditional training algorithms, such as back-propagation, can be used. This algorithm, which is based upon the gradient descent search method, tries to determine the weights to use by minimizing the errors $E(c)$ between the desired thermal IR feature vector $t$ and the actual output $y$ :

$$E(c) = \frac{1}{2} \sum_{i} (t_i - y_i)^2 \tag{38}$$

After training, PCA is performed on an input visible image $I_v$ to generate a low dimension vector $v$. This vector is used as an input to a trained neural network to predict a thermal IR vector $t$. A thermal IR image, $I_t$, can be reconstructed using the predicted vector:

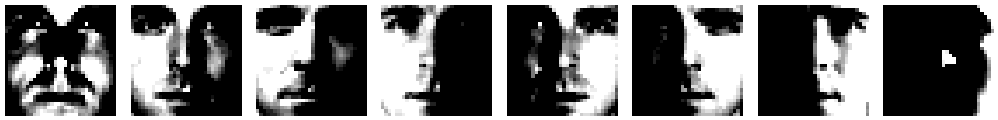
$$I_t = \mu + \sum_{i=1}^{m} w_i t_i \tag{39}$$

where $w_i$ is the $ith$ eigenvector obtained from thermal IR training images, $\mu$ is mean image of thermal IR training images and $t_i$ is the $ith$ entry of predicted thermal IR feature vector.

Figure 24 shows examples of reconstructed novelty components using different numbers of principal components. When a limited number of principal components are used, the reconstructed novelty component looks similar to the ground truth. Figure 25 shows the novelty components reconstructed from visible images under different illumination conditions. Variations between the visible and thermal IR images are determined by computing the standard deviation. When compared to the original visible images (standard deviation: 10.07), the predicted thermal IR components have fewer variations (standard deviation: 1.04).

(a) Visible image (b) 50 PCs (c) 150 PCs (d) 250 PCs (e) Ground truth

Figure 24: Novelty component reconstructed with principal components (PCs).



(a) Visible images (standard deviation: 10.07)

(b) Novelty components of corresponding visible images (standard deviation: 1.04)

Figure 25: Variations of visible images and their novelty components

### 4.3 Fusion of Visible Image with Novelty Component

Many approaches for fusing visible and thermal IR images at the data-level, feature-level, score-level, or decision-level have been reported in the literature. In order to demonstrate that the novelty component can improve face recognition performance, the predicted image is fused with the original image at the data-level to generate a new image. Fusion at the data-level enables many existing face recognition algorithms to be directly applied to the fused image. Therefore, any performance improvement can be attributed to use of the novelty component versus complicated processing. It should be noted that other fusion methods (i.e., feature-level, score-level, or decision-level) can also be applied to the visible image and its novelty component for face recognition.

For the fusion at the data level, a multiscale fusion scheme can be used considering the different resolution between visible and thermal IR images. Both images are first transformed into a set of approximation coefficients $W_{\varphi}(m,n)$ and details coefficients $W_{\Psi}(i,j)$ using DWT. The wavelet coefficients for the fused image are obtained using a weighted average combination of the corresponding wavelet coefficients of the visible image and its novelty component:

$$W_{\varphi}^{Fused}(m,n) = \alpha_1 W_{\varphi}^{Visible}(m,n) + \beta_1 W_{\varphi}^{Thermal}(m,n) \tag{40}$$

$$W_{\psi}^{Fused}(i,j) = \alpha_2 W_{\psi}^{Visible}(i,j) + \beta_2 W_{\psi}^{Thermal}(i,j) \quad (41)$$

A global optimization method called particle swarm optimization (PSO) can be used to learn the weights from a set of training images by maximizing an objective function (i.e., the top one matching rate). After combining the weighted coefficients, the fused image is reconstructed using inverse DWT (*IDWT*):

$$I_{Fused}(x,y) = IDWT[W_{\psi}^{Fused}(i,j), W_{\varphi}^{Fused}(m,n)] \quad (42)$$

## 4.4 Experiment Results

In order to show that the novelty component can improve the performance of face recognition, four popular face recognition methods were performed on both the fused and the visible images. These methods are Eigenfaces, Fisherfaces, kernel direct discriminant analysis (KDDA), and Laplacianfaces [9, 41, 42, 8]. The performance of the face recognition algorithms was assessed in terms of both face identification and verification using two public databases (i.e., the NIST/Equinox database and The Extended Yale Face Database B).

### *4.4.1 NIST/Equinox Database*

The NIST/Equinox database was divided into three parts: (1) training images, (2) probe images, and (3) gallery images. With respect to the training set, there were 582 pairs of visible and thermal IR images. These images were of 30 individuals, and were used to develop the neural network and train the face recognition algorithms. There were 979 pairs of probe images. These images of 58 individuals were presented to the system for identification and verification. The gallery images consisted of 58 pairs of images of 58 individuals with known identities. Each gallery image was taken under frontal illumination condition. Furthermore, the subjects were instructed to maintain a neutral expression and remove eyeglasses.

It should be noted that there is no overlap between the images in any of the sets, which describes the worst-case scenario in terms of success rate. However, this is a good practice to eliminate any bias that might be introduced in the testing process in order to make the evaluation of a face recognition system more objective. The single gallery image setup also makes recognition more difficult. Face recognition systems often use at least three, good-quality images of each person, which are captured in a controlled environment. However, many law enforcement applications may not satisfy these conditions.

All of the images were pre-processed before the experiment. The face was cropped from the original image based upon the eye coordinates. The images was further

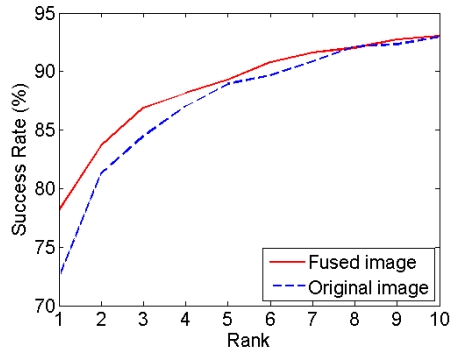
re-sampled to 32 × 32 pixels using bilinear interpolation. Then, histogram equalization was performed on the face region to reduce variations in the image.

In order to train the neural network to transform the visible face to the thermal IR face, 337 pairs of images (without eyeglasses) were chosen from the 582 pairs of training images. Images without eyeglasses were selected in order to avoid the effects of occlusions in the thermal IR images. The selected images were reduced to feature vectors of dimension 30 using PCA. These visible feature vectors were used as the training input, and the thermal IR feature vectors were used as the training target. The neural network used in this research was composed of one input layer, two hidden layers, and one output layer. The input and output layers contained 30 nodes, and the hidden layers contained 6 nodes. After obtaining the neural network, all visible images (i.e., training, probe, and gallery) were converted into feature vectors using PCA. These feature vectors were fed into the neural network in order to predict thermal IR feature vectors. The novelty components were reconstructed with the thermal IR eigenfaces.

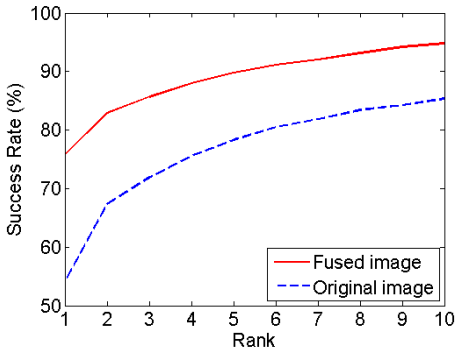
All the training, galley, and probe images were fused using a weighted average combination of the visible images and the novelty components in the level-2 DWT domain. The weights for the visible image and its novelty component were learned by using PSO to maximize the rank one matching rate over the training samples (described previously).

In order to evaluate performance, four popular face recognition algorithms were used: Eigenfaces, Fisherfaces, KDDA and Laplacianfaces. The dimension of the eigenvector for eigenfaces was set to 500. The RBF kernel-based KDDA was used. For Eigenfaces, the whitened cosine distance was selected as the distance measure because it has previously been shown to perform well for this application [31]. For LDA and Laplacianfaces, the cosine distance was used, and the Euclidian distance was used for KDDA.

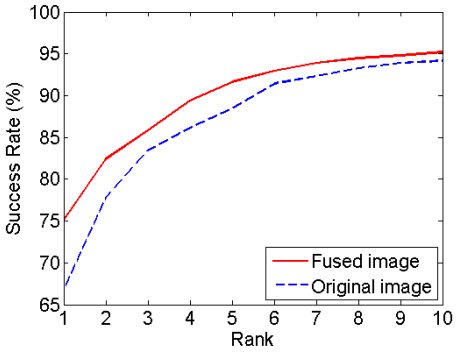
The performances of face recognition with visible images and fused images are compared in Figure 26 and Figure 27. Figure 26 shows the cumulative match characteristic (CMC) for top 10 face matching, and Figure 27 shows the receiver operation characteristic (ROC) for face verification. As can be seen in these graphs, face recognition using the fused visible image with its novelty component outperforms recognition using only the visible image.



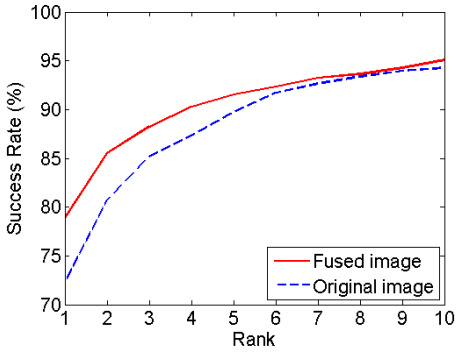

(a) Eigenfaces




(b) Fisherfaces

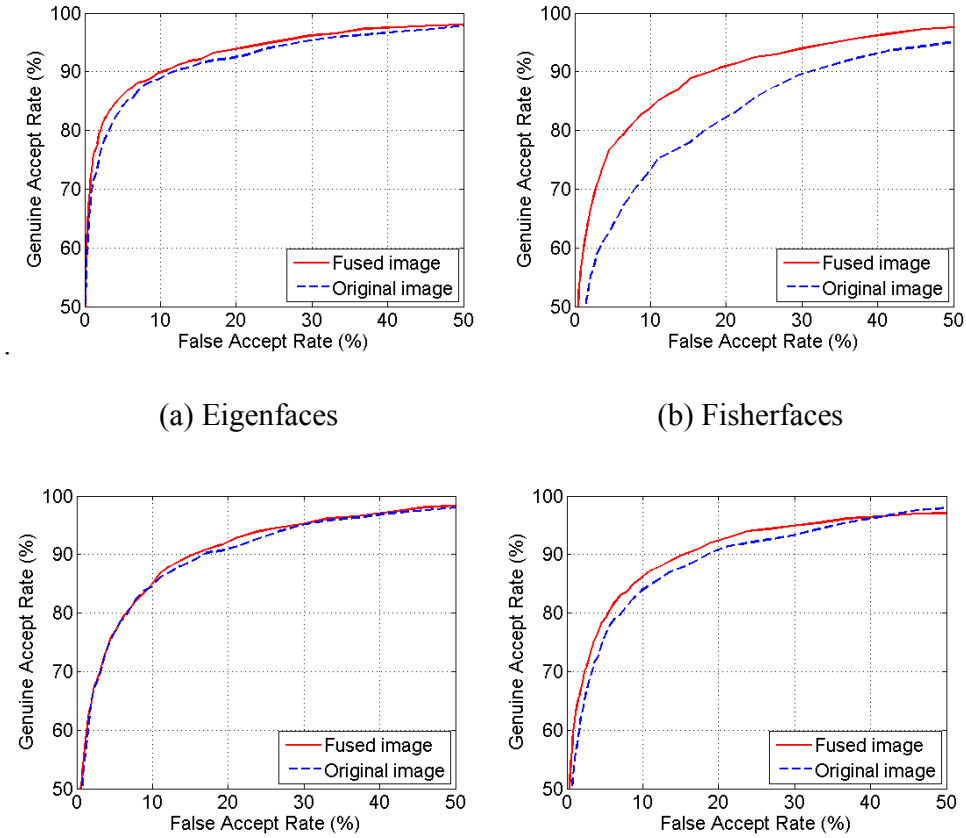



(c) KDDA




(d) Laplacianfaces

Figure 26: Top 10 matches using fused images and visible images




(a) Eigenfaces (b) Fisherfaces

(c) KDDA (d) Laplacianfaces

Figure 27: ROC curve for verification using fused images and visible images.

### *4.4.2 The Extended Yale Face Database B*

The Extended Yale Face Database B consists of 2414 visible frontal images of 38 different individuals (there are approximately 64 images of each person). These images were taken under 64 illumination conditions, with spatial resolutions of 192 × 168 pixels [43, 44]. This database was also divided into training, probe, and gallery images. There were 576 training images of 9 individuals. These were used for training the face recognition algorithms. There were 1809 probe images of 29 individuals, which were presented to the system for identification and verification. There were 29 gallery images of 29 individuals (with known identity). Each gallery image was taken under frontal illumination conditions. As with the NIST/Equinox database, there is no overlap between images in any of the three categories.

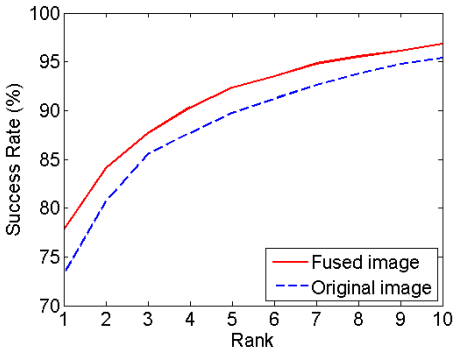
The same pre-processing techniques used previously on the NIST/Equinox database were also used in this experiment. Before predicting the novelty component from a visible image, the neural network that transforms the visible image to a thermal IR image was built. This neural network was built using the 1024 pairs of images (without eyeglasses) from the NIST/Equinox database. Because there is no overlap between the two databases, this trained neural network will not introduce any bias.

In order to construct the neural network, the visible images and thermal IR images from the NIST/Equinox database were first converted to feature vectors with dimension 30. Then, the visible feature vectors were used as training inputs, and the thermal IR

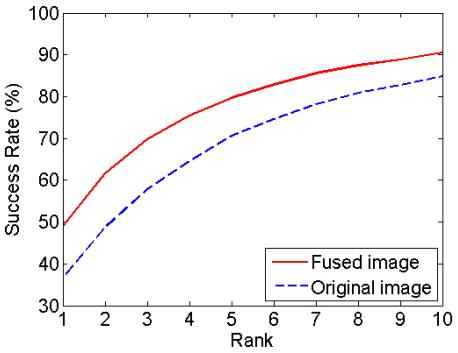
feature vectors were used as training targets. The neural network was again built using one input layer, two hidden layers, and one output layer. The hidden layers had 8 and 6 nodes respectively, and there were 30 input and output nodes.

After obtaining the neural network, all of the visible images from The Extended Yale Face Database B were used to predict the corresponding novelty components. Then, fused images were generated using a weighted average combination of the visible images with their novelty components in the level-2 DWT domain. These fused images were then used as input to the face recognition algorithms.

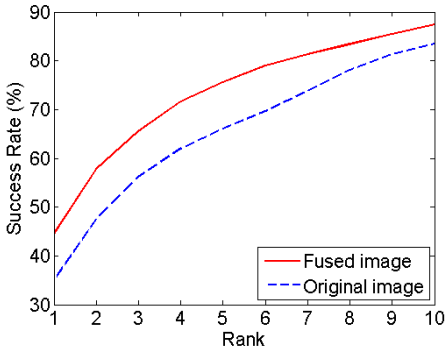
The performance of the face recognition algorithms was evaluated using both the fused images versus only the original visible images as inputs. The results of the top 10 matches for face identification and verification on The Extended Yale Face Database B are shown in Figure 28 and Figure 29. Figure 28 shows the top 10 matches for face identification, and Figure 29 shows the ROC curve for face verification. As can be seen from these graphs, the performance of the face recognition algorithms using the fused visible image with its novelty component is consistently better compared to using the visible image only.



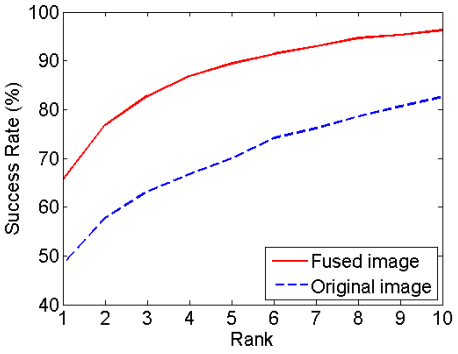

(a) Eigenfaces




(b) Fisherfaces

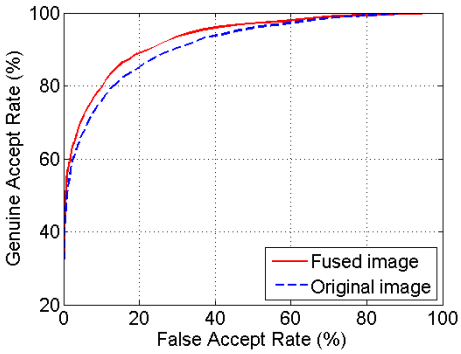



(c) KDDA

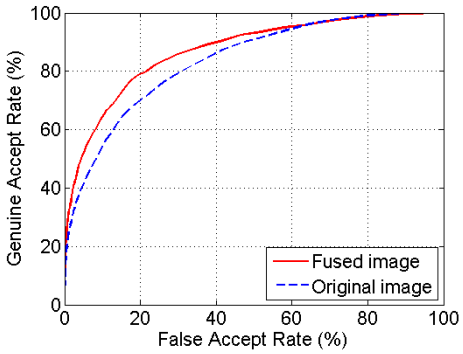



(d) Laplacianfaces

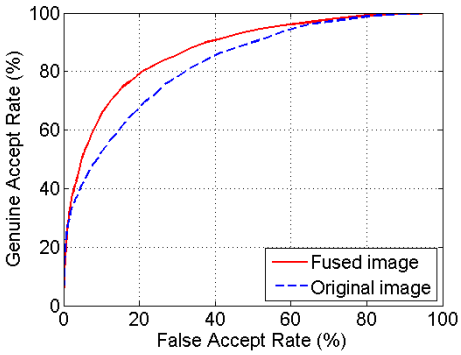
Figure 28: Top 10 matches using fused images and visible images.



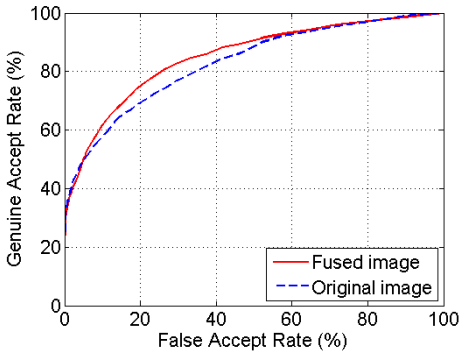

(a) Eigenfaces




(b) Fisherfaces




(c) KDDA




(d) Laplacianfaces

Figure 29: ROC curve for verification using fused images and visible images

.

# CHAPTER 5

# STATISTICAL SIGNIFICANCE

In this chapter, the statistical significance of the experiment results is discussed. Statistical measures are employed to lend confidence that one system is truly better than the other one. In order to evaluate statistical significance, the z-statistic and the right tailed test were used.

## 5.1 Z-Statistic and Right Tailed Test

Because we sometimes lack of knowledge of the whole population for certain experiments, we can only make inferences using a finite sample set. Therefore, uncertainty exists in inferences. In statistics, this uncertainty is referred as the probability of making an error. In order to measure the uncertainty, statistical testing can be used. Statistical testing is often implemented as a problem of hypothesis testing. For two experiments with error rates $P_1$ and $P_2$, it is first assumed that these experiments have the same performance (i.e., the same error rate). Therefore, a null hypothesis can be constructed as:

$$Null\ Hypothesis\ H0: \ P_1 = P_2 \tag{43}$$

In order to prove that the second experiment is significantly better than the first one, the null hypothesis must be rejected at a given significance level. In order to reject this null hypothesis, the normalized z-statistic is used:

$$Z = \frac{(P_1 - P_2)}{\sqrt{\frac{P_1(1-P_1)}{N} + \frac{P_2(1-P_2)}{N}}} \tag{44}$$

According to the Central Limit Theorem, the distribution of the z-statistic is approximately normal if the sample size is large enough. The right-tailed test is used to reject the null hypothesis. For example, let us assume that $P_1$ is the error rate of a baseline experiment with $N$ trials, and our desired statistical significance level is 5%. If we want to show that our experiment with $N$ trials is significantly better than the baseline experiment, we need to specify the upper bound of $P_2$ (i.e., the maximum error rate). In this bound, all of the $Z$-values calculated are larger than $Z_{5\%}$=1.6449 (estimated from normal distribution). By using this constraint, the upper bound of $P_2$ can be found.

## 5.2 Statistical Significance of Experiments

The statistical significance of experiments was applied to our face recognition experiments. These experiments included: face recognition with eyeglass replacement, face recognition with the fusion of visible and thermal IR images, and face recognition of

fusion of a visible image and its novelty component. For all experiments, a statistical significance level of 5% was used to calculate the upper bound of the error rate $P_2$. In order to make the claim that our system is significantly better than the other system, the error rate of our system should be lower $P_2$ (i.e., the success rate should be higher than 1-$P_2$).

Figure 30 shows the statistical significance of the experiments investigating the performance of the face recognition using the eyeglass replacement algorithm. With respect to the top 3 matching images, performance is significantly better using the thermal IR images with the eyeglasses replaced compared to the one using the original thermal IR images. For the top 4 to top 10 matches, no conclusions can be drawn because more data is needed.

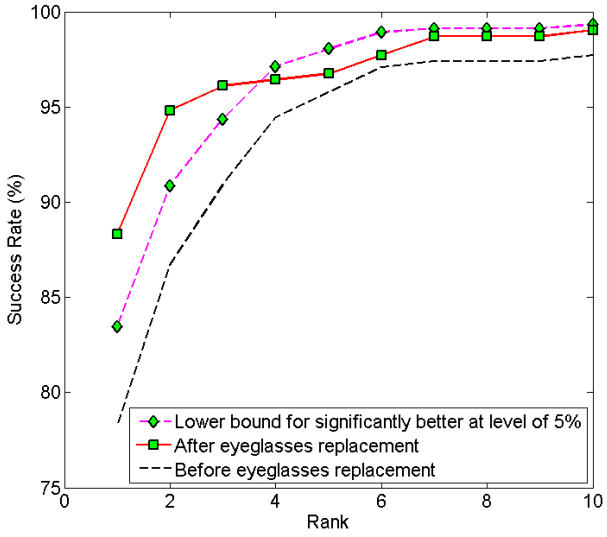



Figure 30: Statistical significance for the eyeglass replacement experiments.

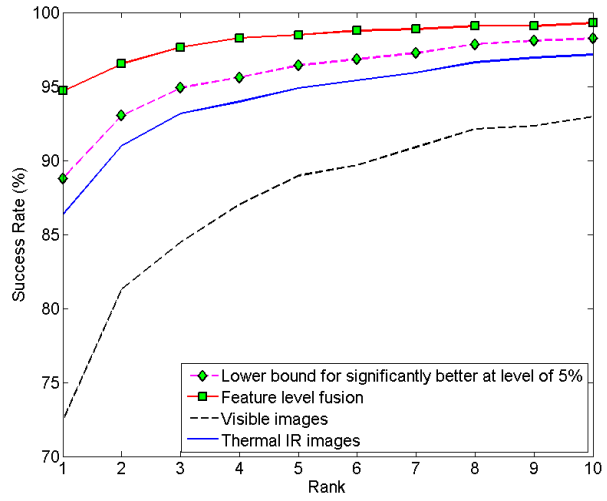
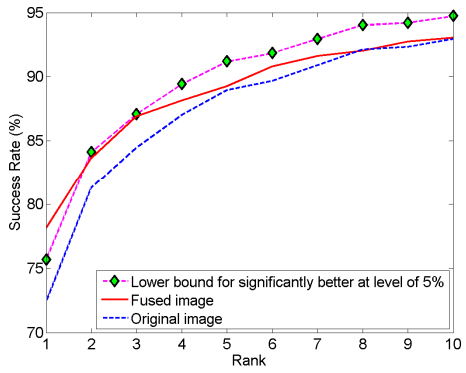
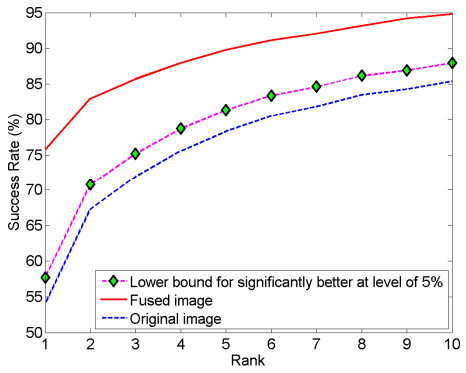



Figure 31: Statistical significance for fusion of visible and thermal IR images.

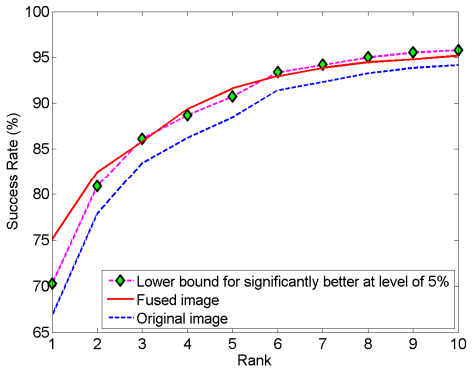
Figure 31 shows the statistical significance of the experiments comparing performance using the fusion of the visible and thermal IR images along with using either image type independently. As can be seen from the figure, face recognition using fusion techniques is significantly better than using a single modality.



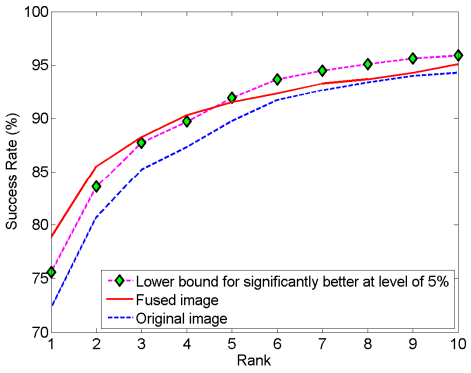

(a) Eigenfaces




(b) Fisherfaces




(c) KDDA




(d) Laplacianfaces

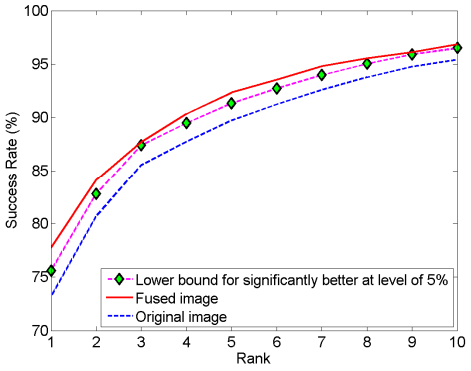
Figure 32: Statistical significance (NIST/Equinox database).

Figure 32 shows the statistical significance of the experiments comparing performance using the fusion of the visible image with its novelty component versus using only the visible image (i.e., for the NIST/Equinox database). Performance for all four face recognition algorithms was statistically better for rank one matching. For other ranks, further research is needed to draw more accurate conclusions.

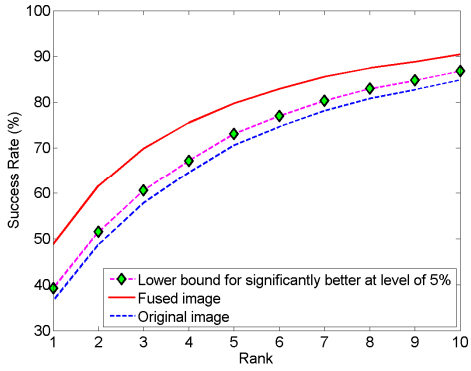
Figure 33 shows the statistical significance of the experiments comparing performance using fusion of the visible image with its novelty component versus using only the visible image (i.e., for the Extended Yale Face Database B). Performance for all four face recognition algorithms was statistically better when using the fusion of the visible image with its novelty component (versus only the visible image).

In conclusion, application of the techniques introduced in the previous chapters improves the performance of a face recognition system. For top one matches (one of the most important statistics in evaluating a face recognition system), the improvement is statistically significant compared to the baseline algorithms. For top 10 matches, performance of the new techniques is not always statistically better, but there is a trend towards improvement (i.e., more data is needed). Furthermore, the error rate of top $N$ matches using the baseline algorithm drops dramatically as $N$ increases. If the error rate is too low (i.e., less than 10%), evaluation of statistical significance becomes unreliable

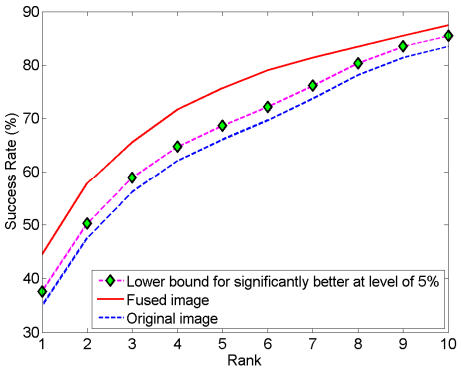
because there is limited space for improvement, as well as a limited number of samples available for evaluation (i.e., which is our case). In order to be able to draw more definite conclusions in terms of statistical evaluation for top $N$ matches ($N > 3$), research on collecting databases is needed.



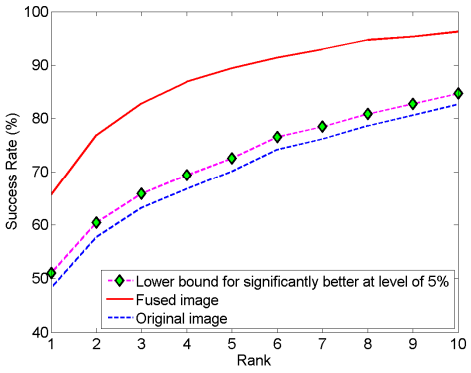

(a) Eigenfaces




(b) Fisherfaces




(c) KDDA




(d) Laplacianfaces

Figure 33: Statistical significance (Extended Yale Face Database B).

# CHAPTER 6

# CONCLUSIONS

In this thesis, face recognition with visible and thermal infrared (IR) images was investigated. Visible images and thermal IR images have their respective advantages and disadvantages for this application. Visible images are widely available and provide many details needed for accurate face recognition. However, they are very sensitive to illumination variations, and performance is degraded if the illumination conditions change. Thermal IR images, which measure the energy emitted from a face, are more robust to illumination variations and therefore provide better performance for recognizing a face in varying illumination conditions. However, as thermal IR is opaque to glass, the presence of eyeglasses on the face results in occlusions that can affect the performance of face recognition.

Considering that visible and thermal IR images provide complementary information, this thesis explored the use of fusion techniques for integrating information

from both spectrums. Prior to fusion, the visible and thermal IR images need to be registered. In this research, an edge-based mutual information (EMI) method is used to register the visible and thermal IR images. Compared to traditional approaches, EMI can register faces more accurately under uncontrolled environments (i.e., illumination variations, complicated backgrounds, and occlusions).

After the visible and thermal IR images have been registered, the eyeglasses (if present) are replaced in thermal IR images. This is accomplished by first detecting the eyeglasses using a support vector machine classifier, and then replacing them with the reconstructed eyes. The eyes are reconstructed from the visible eye information using a neural network model. Experimental results showed that by replacing the eyeglasses in the thermal IR images, the performance of face recognition was improved.

After the images have been registered, they can be fused. This thesis explored the use of data-level, feature-level, and score-level fusion techniques. Experimental results using the NIST/Equinox database showed that face recognition with the fusion of visible and thermal IR images outperformed face recognition using either visible or thermal IR images alone. The fusion of visible and thermal IR images increased the number of first matches by 22% over visible images, and 8% over thermal IR images.

Considering the difficulties of availability and affordability of a thermal IR camera, the fusion of visible and thermal IR images for recognition may be limited to

certain applications. Therefore, a method for predicting a novelty component from a visible image is explored in this thesis. The novelty component is considered to be the thermal IR component of a visible image. The novelty component is reconstructed using principal component analysis and a neural network model. Experimental results using two public databases and four face recognition algorithms show that face recognition performance can be improved by fusing the visible image with its novelty component (versus using only the visible image).

Future research for this project includes developing new methods for face detection in thermal IR images. Current techniques are not robust to variations in temperature or the presence of complicated backgrounds. Additionally, many of the face recognition algorithms that have been developed for the thermal IR spectrum are taken from those used in the visible spectrum. Therefore, an algorithm that incorporates temperature properties is critical for improving the performance of face recognition using thermal IR images. Moreover, additional advanced fusion techniques need to be developed to take advantage of both visible and thermal IR images.

## REFERENCES


[1] A. K. Jain, A. Ross, and S. Prabhakar, "An Introduction to Biometric Recognition," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 14, no. 1, pp. 4-20, January 2004.

[2] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey," *ACM Computing Surveys,* vol. 35, no. 4, pp. 399-458, December 2003.

[3] Y. Adini, Y. Moses, and S. Ullman, "Face Recognition: The Problem of Compensating for Changes in Illumination Direction," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 721-732, July 1997.

[4] J. Ben-Arie and D. Nandy, "A Volumetric/Iconic Frequency Domain Representation for Objects with Application for Pose Invariant Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 20, no. 5, pp. 449-457, May 1998.

[5] I. Pavlidis and P. Symosek, "The Imaging Issue in an Automatic Face/Disguise Detection System," *Proceedings of the IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, pp. 15-24, Hilton Head, South Carolina, USA, June 2000.

[6] Y. I. Tian, T. Kanade, and J. F. Cohn, "Recognizing Action Units for Facial Expression Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 23, no. 2, pp. 97-115, February 2001.

[7] S. Z. Li, R. Chu, S. Liao, and L. Zhang, "Illumination Invariant Face Recognition Using near-Infrared Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 29, no. 4, pp. 627-639, April 2007.

[8] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience,* vol. 3, no. 1, pp. 71-86, January 1991.

[9] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces Vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 711-720, July 1997.

[10] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face Recognition by Independent Component Analysis," *IEEE Transactions on Neural Networks,* vol. 13, no. 6, pp. 1450-1464, November 2002.

[11] B. Heisele, P. Ho, and T. Poggio, "Face Recognition with Support Vector Machines: Global Versus Component-Based Approach," *Proceedings of the Eighth IEEE International Conference on Computer Vision*, pp. 688-694 Vancouver, Canada, July 2001.

[12] K. I. Chang, W. Bowyer, and P. J. Flynn, "Multiple Nose Region Matching for 3D Face Recognition under Varying Facial Expression," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28, no. 10, pp. 1695-1700, October 2006.

[13] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active Appearance Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 23, no. 6, pp. 681-685, June 2001.

[14] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *Proceedings of the International Conference on Image Processing*, pp. 26-29, Santa Barbara, California, USA, October 1997.

[15] S. G. Kong, J. Heo, B. R. Abidi, J. Paik, and M. A. Abidi, "Recent Advances in Visual and Infrared Face Recognition--a Review," *Computer Vision and Image Understanding,* vol. 97, no. 1, pp. 103-135, January 2005.

[16] D. A. Socolinsky, L. B. Wolff, J. D. Neuheisel, and C. K. Eveland, "Illumination Invariant Face Recognition Using Thermal Infrared Imagery," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 527-534, Kauai, Hawaii, USA, 2001.

[17] Y. Yoshitomi, T. Miyaura, S. Tomita, and S. Kimura, "Face Identification Using Thermal Image Processing," *Proceedings of the Sixth IEEE International Workshop on Robot and Human Communication*, pp. 374-379, Sendai, Japan, 1997.

[18] X. Chen, P. J. Flynn, and K. W. Bowyer, "IR and Visible Light Face Recognition," *Computer Vision and Image Understanding,* vol. 99, no. 3, pp. 332-358, September 2005.

[19] P. Buddharaju, I. T. Pavlidis, P. Tsiamyrtzis, and M. Bazakos, "Physiology-Based Face Recognition in the Thermal Infrared Spectrum," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 29, no. 4, pp. 613-626, April 2007.

[20] A. Ross and A. Jain, "Information Fusion in Biometrics," *Pattern Recognition Letters,* vol. 24, no. 13, pp. 2115-2125, September 2003.

[21] J. Llinas and D. L. Hall, "An Introduction to Multi-Sensor Data Fusion," *Proceedings of the 1998 IEEE International Symposium on Circuits and Systems*, pp. 537-540, Monterey, California, USA, 1998.

[22] D. A. Socolinsky, A. Selinger, and J. D. Neuheisel, "Face Recognition with Visible and Thermal Infrared Imagery," *Computer Vision and Image Understanding,* vol. 91, no. 1-2, pp. 72-114, July-August 2003.

[23] G. Bebis, A. Gyaourova, S. Singh, and I. Pavlidis, "Face Recognition by Fusing Thermal Infrared and Visible Imagery," *Image and Vision Computing,* vol. 24, no. 7, pp. 727-742, July 2006.

[24] S. G. Kong, J. Heo, F. Boughorbel, Y. Zheng, B. R. Abidi, A. Koschan, M. Yi, and M. A. Abidi, "Multiscale Fusion of Visible and Thermal IR Images for Illumination-Invariant Face Recognition," *International Journal of Computer Vision,* vol. 71, no. 2, pp. 215-233, February 2007.

[25] B. Zitova, "Image Registration Methods: A Survey," *Image and Vision Computing,* vol. 21, no. 11, pp. 977-1000, October 2003.

[26] D. A. Socolinsky and A. Selinger, "A Comparative Analysis of Face Recognition Performance with Visible and Thermal Infrared Imagery," *Proceedings of the 16th International Conference on Pattern Recognition*, pp. 217-222, Quebec City, Canada, August 2002.

[27] J.-G. Wang, E. Sung, and R. Venkateswarlu, "Registration of Infrared and Visible-Spectrum Imagery for Face Recognition," *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 638-644, Seoul, Korea, May 2004.

[28] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision,* vol. 57, no. 2, pp. 137-154, May 2004.

[29] J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 8, no. 6, pp. 679-698, November 1986.

[30] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, New York, New York, USA, 2000.

[31] J. R. Beveridge, D. Bolme, B. A. Draper, and M. Teixeira, "The CSU Face Identification Evaluation System," *Machine Vision and Applications,* vol. 16, no. 2, pp. 128-138, February 2005.

[32] I. Vajda, *Theory of Statistical Inference & Information*, Springer-Verlag, New York, New York, USA 1989.

[33] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery,* vol. 2, no. 2, pp. 121-167, June 1998.

[34] K. Hornik, M. Stinchcombe, and H. White, "Multilayer Feedforward Networks Are Universal Approximators," *Neural Networks,* vol. 2, no. 5, pp. 359-366, March 1989.

[35] Y. S. Russell C. Eberhart "Particle Swarm Optimization: Developments, Applications and Resources," *Proceedings of the 2001 Congress on Evolutionary Computation*, pp. 81-86 Seoul, Korea, May 2001.

[36] A. C. S. Chung and H. C. Shen, "Entropy-Based Markov Chains for Multisensor Fusion," *Journal of Intelligent and Robotic Systems,* vol. 29, no. 2, pp. 161-189, October 2000.

[37] R. Singh, M. Vatsa, and A. Noore, "Integrated Multilevel Image Fusion and Match Score Fusion of Visible and Infrared Face Images for Robust Face Recognition," *Pattern Recognition,* vol. 41, no. 3, pp. 880-893, March 2008.

[38] M. Reiter, R. Dormer, G. Langs, and H. Bischof, "3d and Infrared Face Reconstruction from Rgb Data Using Canonical Correlation Analysis," *Proceedings of the 18th International Conference on Pattern Recognition*, pp. 425-428, Hong Kong, China, August 2006.

[39] J. Li, P. Hao, C. Zhang, and M. Dou, "Hallucinating Faces from Thermal Infrared Images," *Proceedings of the 15th IEEE International Conference on Image Processing*, pp. 465-468, San Diego, California, USA, 2008.

[40] P. Viola and I. William M. Wells, "Alignment by Maximization of Mutual Information," *International Journal of Computer Vision,* vol. 24, no. 2, pp. 137-154, September 1997.

[41] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face Recognition Using Laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, no. 3, pp. 328-340, March 2005.

[42] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face Recognition Using Kernel Direct Discriminant Analysis Algorithms," *IEEE Transactions on Neural Networks,* vol. 14, no. 1, pp. 117-126, January 2003.

[43] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 23, no. 6, pp. 643-660, June 2001.

[44] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring Linear Subspaces for Face Recognition under Variable Lighting," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, no. 5, pp. 684-698, May 2005.