

Fusion of Visual and Thermal Face Recognition Techniques: A Comparative Study

Project in Lieu of Thesis

for

Master's Degree

The University of Tennessee, Knoxville

Jingu Heo

October 2003

Acknowledgements

First of all, I would like to thank all IRIS Lab members. I would like to thank my advisor, Dr. Mongi Abidi. His willingness to support my work and his guidance throughout my studies has allowed me to develop my skills as a face recognition researcher within a supportive team environment. Also, I would like to thank, Dr. Besma Abidi, Dr. Seong Kong, and Dr. Joongi Paik. Their advice over the years has been of equal importance.

Abstract

This paper describes how fusion of visual and thermal face recognition can increase the overall performance of face recognition systems. Visual face recognition systems perform relatively reliably under controlled illumination conditions. Thermal face recognition systems are advantageous for detecting disguised faces or when there is no control over illumination. Thermal images of individuals wearing eyeglasses may result in poor performance since eyeglasses block the infrared emissions around the eyes, which are important features for recognition. With taking advantages of each visual and thermal image, the new fused systems can be implemented in collaborating low-level data fusion and high-level decision fusion. This paper contains a comparative analysis of visual, thermal and fused face recognition performance and provides a breakthrough for future intelligent face recognition capable of making decisions based on visual, thermal or fused images depending on different situations.

Contents

1. Introduction.....	1
1.1 Fundamental Issues in Face Recognition.....	2
1.2 Visual Face Recognition.....	6
1.3 Thermal Face Recognition.....	9
1.4 Fusion Approaches to Face Recognition	11
2. Data Acquisition	15
2.1 Image Databases	16
2.2 Image Acquisition Setup.....	17
2.2 Fusion of Visual and Thermal Images.....	21
3. Feature Extraction from Thermal Images.....	22
3.1 Variations.....	22
3.2 Face Detection from Thermal Images	27
3.3 Eyeglass Detection and Removal from Thermal Images.....	31
4. Comparison of Face Recognition Techniques	37
4.1 Feature Based Face Recognition.....	38
4.2 Appearance Based Face Recognition	43
4.3 Decision Fusion	50
4.4 Visual and Thermal Comparison Results	52
4.5 Fusion of Visual and Thermal Face Recognition	54
5. Conclusion	58
References.....	60
Vita.....	68

List of Tables

Table 1: Datasets.....	16
Table 2: Summary of the IRIS face database.....	20
Table 3: Performance for eyeglass detection	36
Table 4: Experimental results for FaceIt® Identification	45
Table 5: Summary of pose test.....	45
Table 6: Test items not included in FaceIt® identification experiment	46
Table 7: Summary of experimental results	48

List of Figures

Figure 1: A framework for the comparison of visual (Vi), thermal (Th), data fusion (Df), decision fusion based on average confidence rates(Fa), and decision fusion based on higher confidence rates (Fh) face recognition system	14
Figure 2: Example of visual and thermal images from Equinox	17
Figure 3: Image acquisition platform.....	19
Figure 4: Example of visual and thermal images taken at the same time	19
Figure 5: Visual and thermal images taken from rotary system	20
Figure 6: A data fusion example.....	22
Figure 7: Problems with edge detection results using thermal images	24
Figure 8: Problems with histogram methods	25
Figure 9: Average thermal images.....	26
Figure 10: Average thermal images without (top)/with (bottom) eyeglasses.....	26
Figure 11: Facial feature variations of average images indicated by histograms	27
Figure 12: Ellipse fitting demo	29
Figure 13: Face detection diagram.....	30
Figure 14: Face detection example	31
Figure 15: Diagram of eyeglass detection algorithm.....	32
Figure 16: Eyeglass detection example.....	34
Figure 17: Performance of eyeglass detection with FAR and FRR.....	35
Figure 18 : The average eye regions template	36
Figure 19: Data fusion after eyeglass removal	37
Figure 20: Geometrical distances on the face	38
Figure 21: Different landmarks.....	39
Figure 22: Features based on face regions.....	39
Figure 23: Eigenface-based face recognition system.....	40
Figure 24: Accuracy results based on shortest Euclidean distance (distance, points)	42
Figure 25: Comparison of our experiments with FaceIt®	42
Figure 26: FaceIt® software templates.....	43

Figure 27: Example images of same individual under different conditions tested with FaceIt® Identification	44
Figure 28: Example images of same individual with different poses	44
Figure 29: Pose test summary	46
Figure 30: Example images of captured faces for FaceIt® Surveillance experiment	47
Figure 31: The effects of DB size and variations	49
Figure 32: The effects of lighting and distance	49
Figure 33: Decision fusion example	52
Figure 34: Performance evaluation when individuals are wearing no eyeglasses.....	53
Figure 35: Fused performance evaluation with eyeglasses off (Probe 1, 2, 3).....	55
Figure 36: Performance comparison when eyeglasses are present (Probe 4, 5, 6).....	56
Figure 37: Performance comparison when eyeglasses are replaced.....	56
Figure 38: Performance comparison of 1 st match success rates	57
Figure 39: Performance comparison of average confidence rates	58

1. Introduction

Face recognition has developed over 30 years and is still a rapidly growing research area due to increasing demands for security in commercial and law enforcement applications. Although, face recognition systems have reached a significant level of maturity with some practical success, face recognition still remains a challenging problem due to large variation in face images.

Face recognition is usually achieved through three steps: acquisition, normalization and recognition. Acquisition includes the detection and tracking of face-like images in a dynamic scene. Normalization includes the segmentation, alignment and normalization of the face images. Finally, recognition includes the representation and modeling of face images as identities, and the association of novel face images with known models. In order to realize such a system, acquisition, normalization and recognition must be performed in a coherent manner.

The performance of face recognition systems varies significantly according to the environment where face images are taken and according to the way user-defined parameters are adjusted in several applications. Recognition based only on the visual spectrum remains limited in uncontrolled operating environments such as outdoor situations and low illumination conditions. Visual face recognition also has difficulty in detecting disguised faces, which is critical for high-end security applications. The thermal infrared (IR) spectrum comprises mid-wave infrared (MWIR) ($3\text{-}5\mu\text{m}$), and long-wave infrared (LWIR) ($8\text{-}12\mu\text{m}$), all longer than the visible spectrum ($0.4\text{-}0.7\mu\text{m}$). Thermal IR imagery is independent of ambient lighting since thermal IR sensors only measure the

heat emitted by objects. The use of thermal imagery has great advantages in poor illumination conditions, where visual face recognition systems often fail.

Then, how to solve variation problems in face recognition? It will be a highly challenging task if we want to solve those problems using visual images only. This paper presents the concept that fusion of visual and thermal face recognition can significantly improve the overall performance of a face recognition system. Sec 1.1 outlines fundamental issues in face recognition technology. We then review basic visual face recognition algorithms and thermal face recognition in Sec 1.2 and Sec 1.3 respectively. In Sec 1.4, we propose a framework for comparison of visual, thermal, data fusion, and decision fusion face recognition systems. This comparison analysis may provide a breakthrough for a future intelligent face recognition system which is capable of making decisions based on visual, thermal, or fused face recognition depending on different situations.

1.1 Fundamental Issues in Face Recognition

Human biometric authentication methods are based on physiological characteristics such as face, fingerprints, veins, iris, voice and ears. Face recognition is regarded as one of the most successful areas of human biometrics. Face recognition offers non-intrusive and perhaps the most natural way of personal identification. Authentication using face recognition is intuitive and does not have to stop user activities. The analysis of frontal or profile face images is often effective without the participants' cooperation or knowledge. A face is a three-dimensional object and can be seen differently according to inside and outside elements. Inside elements are expression, pose, and age that make the face seen

differently. Outside elements are brightness, size, lighting, position, and other surroundings. Robust face recognition requires identifying individuals despite these variations. Much research effort has been concentrated on face recognition tasks in which only a single image or at most a few images of each person are available and a major concern has been scalability to large databases containing thousands of people.

Application areas of face recognition technology include identification for law enforcement, matching of photographs on credit cards or driver's licenses, access control to secure computer networks and facilities such as government buildings and courthouses, authentication for secure banking and financial transactions, automatic screening at airports for known terrorists, and video surveillance usage [1].

Face recognition addresses the problem of identifying or verifying one or more persons of interest in the scene by comparing input faces with the face images stored in a database. While humans quickly and easily recognize faces under variable situations or even after several years of separation, the problem of machine face recognition is still a highly challenging task in pattern recognition and computer vision.

Face recognition in outdoor environments is a challenging machine vision task especially where illumination varies greatly. Performance of visual face recognition is sensitive to variations in illumination conditions [2]. Since faces are essentially 3D objects, lighting changes can cast significant shadows on a face. This is one of the primary reasons why current face recognition technology is constrained to indoor access control applications where illumination is well controlled. Light reflected from human faces also varies significantly from person to person. This variability, coupled with dynamic lighting conditions, causes a serious problem. The use of an artificial

illumination source can reduce light variability, but it will distract the people in the scene and reveal the presence of a surveillance system.

Robust face recognition systems should consider all possible facial variations that might affect the performance of the overall system. This cannot be achieved without acquiring large sized databases. The US Defense Advanced Research Projects Agency (DARPA) and the US Army Research Laboratory (ARL) established the Face Recognition Technology (FERET) program [3]. FERET is designed to measure the performance of face recognition algorithms on a large database in practical settings. The FERET program provides a large database of facial images (14,126) taken from 1,199 individuals and collected between August 1993 and July 1996 to support algorithm development and evaluation. With the collection of large database and evaluation protocols, FERET clarified the state of the art in face recognition and pointed out general directions for future research. The FERET evaluations allowed the computer vision community to assess overall strengths and weaknesses in the field, not only on the basis of the performance of an individual algorithm, but, in addition, on the aggregate performance of all algorithms tested.

The robustness of face recognition system can be evaluated in terms of three tasks: verification, identification, and watch list [4][5][6]. In verification tasks, a person presents his biometric and an identity claim to a face recognition system. The system then compares the presented biometric with a stored biometric of the claimed identity. Based on the results of comparing the new and stored biometric, the system either accepts or rejects the claim. The performance statistics *verification rate* and *false accept rate* characterize the verification performance, the rates that respectively grant legitimate

users and imposters access. An identification task provides a ranked listing of the candidates that best match an unknown person presented to the database. In the watch list task, a face recognition system must first detect if an individual is on the watch list. If the individual is on the watch list, the system then identifies the individual. The statistic for correctly detecting and identifying an individual on the watch list is called the *detection and identification rate*. The *false alarm rate* indicates the rate that the system incorrectly alarms an individual not on the watch list. This report also provides demographic results that indicate males and older people are relatively easier to recognize than females and younger people. Face Recognition Vendor Test (FRVT) 2002 results show that normal changes in indoor lighting do not significantly affect the performance of the top systems.

With the development of face recognition systems, concerns over individuals' privacy have been an issue for real world applications. This is also one of the reasons why current facial recognition systems are limited to security checkpoints or important facilities. Accuracy (face recognition is one of most successful applications of pattern analysis) is still not enough. So far none of the face-recognition systems tested in airports have spotted a single person actually wanted by authorities. They are blamed for serving only to embarrass innocent people. False recognition (false positive or false negative) is very critical. Innocent people can be detained due to a currently unreliable facial recognition system even though they sacrificed their privacy. This is one of the applications of face verification or watch list based face recognition system fails. Since face verification or watch list systems are fully automatic, the responsibility of the operator is to update the database on regular basis and control the environment where faces are taken in order to increase performance. However, it is reasonable that face

identification recognition systems should be regarded as a tool that reduces the searching time and help the operator or face recognition expert. Think about databases that contain millions of faces. The face recognition expert (the best face recognizer among human beings) will take a long time to compare an unknown face against the database. Face recognition systems can reduce that searching time. Here, face recognition experts also can help with the decisions regarding abstract features such as sex, race, and disguises which are challenging tasks for automatic discrimination. These abstract features may also reduce searching time. Utilizing the output of a face identification system, final decision can be made by the experts who are responsible for identity of the faces.

1.2 Visual Face Recognition

Face recognition algorithms can be classified into two broad categories: feature-base and holistic methods [7][8]. The *analytic* or feature-based approaches compute a set of geometrical features from the face such as the eyes, nose, and the mouth. The *holistic* or appearance-based methods consider the global properties of the human face pattern. The face is recognized as a whole without using only certain fiducial points obtained from different regions of the face. Feature extraction for face representation is a central issue in face recognition. Feature extraction algorithms aim at finding features from the scenes that distinguish one face from another. Face patterns can have significantly variable image appearances. Therefore it is essential to find techniques that introduce low-dimensional feature representation of face objects with enhanced discriminatory power. Data reduction and feature extraction schemes make the face recognition problem computationally tractable.

A number of earlier face recognition algorithms are based on feature-based methods [9][10][11] that detect a set of geometrical features on the face such as the eyes, eyebrows, nose, and mouth. Properties and relations such as areas, distances, and angles between the feature points are used as descriptors for face recognition. Typically, 35-45 feature points per face were generated. The performance of face recognition based on geometrical features depends on the accuracy of the feature location algorithm. However, there are no universal answers to the problem of how many points give the best performance, what the important features are, and how to extract them automatically.

Appearance-based face recognition algorithms proceed by projecting an image into the subspace and finding the closest point. Linear discriminant analysis (LDA) and principal component analysis (PCA) have been two approaches widely used for dimensionality reduction and feature extraction [12]. Several leading commercial face recognition products use face representation methods based on the PCA or Karhunen-Loeve (KL) expansion techniques, such as Eigenface [13] and local feature analysis (LFA) [14][15][16]. Multispace KL is introduced as a new approach to unsupervised dimensionality reduction for pattern representation and face recognition, which outperform KL when the data distribution is far from a multidimensional Gaussian [17]. Discriminant analysis based on Fisher's linear discriminant function [18][19] has been widely employed in face recognition. LDA determines a set of optimal discriminant basis vectors so that the ratio of the between- and within-class scatters is maximized. The LDA could be operated either on the raw face image to extract the Fisherface [20][21] or on the eigenface to obtain the discriminant eigenfeatures [22][23].

The LFA constructs a family of locally correlated feature detectors based on PCA decomposition. A selection, or sparsification, step produces a minimally correlated and topographically indexed subset of features that define the subspace of interest. Local representations offer robustness against variability due to changes in localized regions of the objects. The features used in the LFA method are less sensitive to illumination changes, easier for estimating rotations, and have less computational burden than the eigenface method. Motivated by the fact that much of the important information may be contained in the high-order relationships, face recognition based on the independent component analysis (ICA) is proposed as a generalization that is sensitive to higher-order statistics, not second-order relationships [24][25]. ICA provides a set of basis vectors that possess maximum statistical independence whereas PCA uses eigenvectors to determine basis vectors that capture maximum image variance.

Face recognition techniques based on elastic graph matching [26] and neural networks [27] showed successful results. Support vector machines (SVMs) [28][29] find the optimal separating hyper-plane that maximizes the margin of separation in order to minimize the risk of misclassification not only for the training samples, but also the unseen data in the test set. Face recognition based on SVM is presented in [30]. The line edge map approach [31] extracts lines from a face edge map as features, based on a combination of template matching and geometrical feature matching. The nearest feature line classifier attempts to extend the capacity covering variations of pose, illumination, and expression for a face class by finding the candidate person owning the minimum distance between the feature point of the queried face and the feature lines connecting any two prototype feature points [32].

1.3 Thermal Face Recognition

Face recognition in the thermal infrared domain has received relatively little attention in the literature in comparison with recognition in visible-spectrum imagery. Identifying faces from different imaging modalities, in particular from infrared imagery becomes an area of growing interest. Despite the success of automatic face recognition techniques in many practical applications, recognition based only on the visual spectrum has difficulties performing consistently under uncontrolled operating environments. Performance of visual face recognition is sensitive to variations in illumination conditions [33]. The performance degrades significantly when the lighting is dim or when it is not uniformly illuminating the face. Even when a face is well lit, differences in the angle of view can affect manual or automatic locating of feature points. Shadows, glint, makeup, and disguises can cause greater errors in locating the feature points and deriving relative distances. Thermal infrared images represent the heat patterns emitted from an object. Since the veins and tissue structure of a face is unique, the infrared images are unique. Thermal IR imagery is independent of ambient illumination since the human face and body is an emitter of thermal energy. The passive nature of the thermal infrared systems lowers their complexity and increases their reliability. The human face and body maintain a constant average temperature of about 36 degrees providing a consistent thermal signature. This is in contrast to the difficulties of face segmentation in the visible spectrum due to physical diversity coupled with lighting, color, and shadow effects. At low resolution, IR images give good results for face recognition.

Thermal face recognition is useful under all lighting conditions including total darkness when the subject is wearing a disguise. Disguised face detection is of particular

interest in high-end security applications. Disguises are meant to cheat the human eye. Various disguise materials and methods that have been developed over the years are impossible or very difficult to be detected in the visible spectrum. Two methods for altering the facial characteristics of a person: artificial materials, e.g., fake nose, make-up, wig, artificial eyelashes, and surgical alteration. Visual identification of individuals with disguises or makeup is almost impossible without prior knowledge. The facial appearance of a person changes substantially through use of simplistic disguise such as fake nose, wig, or make-up. The individual may alter his/her facial appearance via plastic surgery. Both of these issues are critical for the employment of face recognition systems in high security applications. The thermal infrared spectrum enables us to detect disguises under low contrast lighting. Symptoms such as alertness and anxiety can be used as a biometric which is difficult to conceal as redistribution of blood flow in blood vessels causes abrupt changes in the local skin temperature.

Radiometric calibration is an important step for thermal facial imagery. Thermal IR cameras can be radiometrically calibrated using blackbody radiation. Radiometric calibration achieves a direct relationship between the gray value response at a pixel and the thermal emission. This standardizes all thermal data even if taken under different environments. This relationship is called responsivity. The grayvalue response of thermal IR pixels is linear with respect to the amount of incident thermal radiation. The slope of this responsivity line is called the gain and the y-intercept is the offset. Images of a blackbody radiator covering the entire field of view are taken at two known temperatures, and the gains and offsets are computed using the radiant flux for a blackbody at a given temperature [34][35].

Appearance-based approaches are commonly used for IR face recognition systems [36]. In contrast to visual face recognition algorithms that mostly rely on the eye location, thermal IR face recognition techniques present difficulties in locating the eyes. Initial research approaches to thermal face recognition extracts and matches thermal contours for identification. Such techniques include elemental shape matching and the Eigenface method. Elemental shape matching techniques use the elemental shape of thermal face images. Several different closed thermal contours can be observed in each face. Variations in defining the thermal slices from one image to another has the effect of shrinking or enlarging the resulting shapes while keeping the centroid location and other features of the shapes constant. Perimeter, area, x and y coordinates of the centroid, minimum and maximum chord length through the centroid and between perimeter points, and standard deviation of that length are being considered. Automated face recognition using elemental shapes in real time has reported 96% accuracy for cooperative access control applications. A non-cooperative, non-real-time, faces-in-the-crowd version of thermal face recognition achieved 98% accuracy with no false positives for more than 100 people represented in a database of 500 images [37]. Yoshitomi et al [38] also proposed thermal face identification using Neural Networks. They classified based on combining various information such as gray level histograms, mosaic images, and shape factors.

1.4 Fusion Approaches to Face Recognition

The best classifier in the world, human beings, also do not rely on a single modality to recognize objects if they are producing different information. Information fusion

covers any area, which deals with utilizing a combination of different sources of information, either to generate one representational format, or to reach a decision. Information fusion includes consensus building, team decision theory, committee machines, integration of multiple sensors, multi-modal data fusion, and a combination of multiple experts/classifiers, distributed detection and distributed decision making [39]. There are several motivations for using information fusion: utilizing complementary information can reduce error rates; use of multiple sensors can increase reliability; the cost of implementation can be reduced by using several cheap sensors rather than one expensive sensor; sensors can be physically separated, allowing the acquisition of information from different points of view [40].

Information fusion can be divided into three main categories; sensor data level (low level), feature level (intermediate level), and decision level (high level) [41]. **Data fusion (low level)** combines several sources of raw data to produce new raw data that is expected to be more informative and synthetic than the inputs. Typically, in image processing, images presenting several spectral bands of the same scene are fused to produce a new image with all (most) of the information available in the various spectral bands. An operator or an image-processing algorithm could then use this single image instead of the original images. This kind of fusion requires a precise (pixel-level) registration of the available images such as co-registered visual and thermal images. This registration is intrinsic when the various bands come from the same sensor but it is a lot more complicated when several different sensors are used such as visual and thermal, range cameras. This fusion algorithm can be achieved by *weighted summation* or in the form of *average operation*. **Feature level fusion (intermediate level fusion)** combines

various features in feature spaces. Methods of feature fusion, including several features such as edges, corners, lines, texture parameters, are computed and combined in a fused feature map that may then be used for segmentation or detection. Features obtained after dimension reduction such as PCA, LDA, ICA, LFA and other feature extraction methods also can be used for a fused feature map. **Decision Fusion (high level)**, combines decisions coming from several experts. Methods of decision fusion include majority voting, ranked list combination, AND fusion, OR fusion. A detailed analysis of fusion methods can be reviewed at [39][41].

Numerous attempts have been made in the field of face recognition to fuse different modalities. But relatively less publications are available in this area. Ben-Yacoub. et al fused face and speech data enhances personal identification [42]. Gutta. et al [43] fused using multiple neural networks classifiers to classify gender, ethnic origin, and pose of human faces. They improved the overall face recognition system. Fang. et al [44] tried to fuse local features and global features using SVM (Support Vector Machine). Chang, et al [45] proved that multimodal recognition using both the face and the ear significantly improved the overall performance over a single modality. Accuracy using the face, ear and fused recognition was 70.5%, 71.6% and 90.9% respectively. Wilter. et al [46] and authors from Equinox [34][35][36] fused visual and thermal images using the fusion of experts algorithm and showed better performance over single modalities. Challenging problems in visual face recognition can be solved with the use of thermal images to improve overall performance.

Figure 1 shows a general framework for the comparison of face recognition performance with five different cases; visual (Vi), thermal (Th), data fusion (Df), two

decision fusion schemes (average confidence rates (Fa), higher confidence rates (Fh).

Before entering an image into the face recognition system, decisions should be made from the face (eyes and eyeglasses) detection in the visual and thermal images.

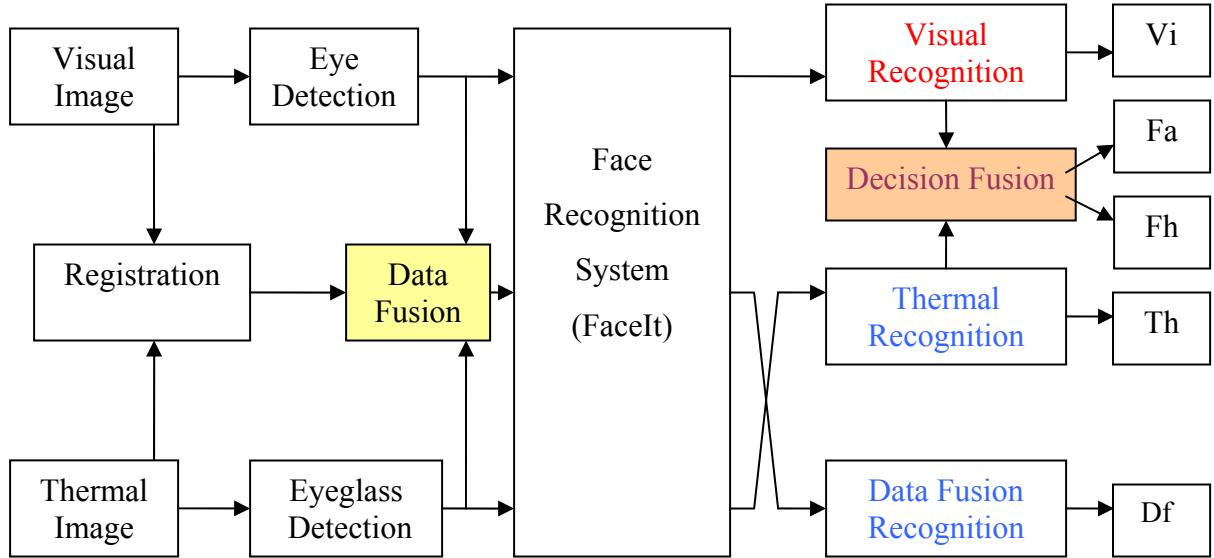


Figure 1: A framework for the comparison of visual (Vi), thermal (Th), data fusion (Df), decision fusion based on average confidence rates(Fa), and decision fusion based on higher confidence rates (Fh) face recognition system

Precise registration is an important step to utilize both visual and thermal images. After registration, eye coordinates from a visual image can be used for fused or thermal face normalization. Under dark illumination conditions, a visual face detection algorithm will not work and only thermal based normalization and recognition should be performed. If people are wearing eyeglasses, which can be easily detected from thermal face detection, it is better to use visual images for face recognition or replace the image with average thermal eyes to minimize the effect of eyeglasses. If both face detection algorithm fails, the faces should not be entered into face recognition system to prevent

false recognition and rejection rates. If the face detection is confident enough in visual images and eyeglasses are detected and replaced in thermal images, we can fuse visual and thermal images for a better recognition system. In other words, this system can be operating with either visual, thermal, or data fusion images depending on different situations. On the other hand, decision fusion systems rely on the output of both visual and thermal recognition results and do not need any sensor fusion.

The remainder of this paper documents the details of our algorithms developed to fuse visual and thermal images or recognition results properly. Chapter 2 deals with image acquisition from public database, our own image acquisition device, and data fusion. In Chapter 3, automatic eyeglasses detection algorithms are proposed with ellipse fitting methods and statistical thermal facial variations. After detecting eyeglasses, eyeglass removal and replacement are also discussed. Chapter 4 contains a comparison of visual, thermal, and fused face recognition, and claims fusion of data or fusion of decisions can increase the overall performance of face recognition systems which may be an important step toward more reliable systems. Finally we conclude in Chapter 5.

2. Data Acquisition

In order to evaluate face recognition algorithms, one should use standard databases so that the evaluation can be reasonable. The FERET database for the visual images is used in comparison with holistic and feature based recognition algorithms in section 4.1 and for the robustness of FaceIt® which we mainly discuss in section 4.2. Regarding visual and thermal images, the Equinox Corporation collects an extensive database of face imagery using co-registered broadband-visible/LWIR, MWIR, and SWIR camera

sensors for experimentation and statistical performance evaluations [47]. Validation trials are being conducted to compare recognition performance of visible and thermal imagery. They developed a spatially co-registered visual and thermal camera which is important in order to achieve precise pixel-based registration.

2.1 Image Databases

We downloaded enough images for the evaluation of face recognition performance. The database consists of co-registered visual and LWIR images of 3,244 (1,622 per modality) face images from 90 individuals. One of each image taken with overhead lighting conditions is used for the gallery images. Probe images were also divided into different conditions as shown in Table 1.

Table 1: Datasets

Dataset	Images	Eyeglasses	Lighting	Expression
Gallery	90	Off	Overhead	Neutral
Probe 1	283	Off	Overhead	Varies
Probe 2	370	Off	Left	Varies
Probe 3	365	Off	Right	Varies
Probe 4	177	On	Overhead	Varies
Probe 5	172	On	Left	Varies
Probe 6	165	On	Right	Varies

Figure 2 shows examples of visual and thermal images acquired from Equinox Company. Original 12-bit gray level thermal images were converted into 8 bit and equalized (only 7 bits contain information) so that the eye regions are roughly seen. The database is used throughout this paper not only for thermal feature extraction but also for fusion of visual and thermal images.



Figure 2: Example of visual and thermal images from Equinox. (a) Visual images, (b) original thermal images, and (c) equalized thermal images

2.2 Image Acquisition Setup

Variations between the images of the same face due to changes in illumination and viewing directions are typically larger than image variations raised from a change in face identity. Other factors such as facial expressions and pose variations further complicate the face recognition task. Simple disguises such as a fake nose or beard substantially change a person's visual appearance. In an effort to solve main factors that degrade the performance of visual face recognition systems such as illumination, expression, rotation, and disguise, we have created our own database considering these variations. In order to provide a comparatively large image gallery for the face recognition project, we consider changing illumination conditions and various expressions. Besides the regular room

lights, we have two extra spot lights locate in the front of the person that we can turn off and on in sequence to obtain face images under different illumination conditions.

The concept here is to perform 180 degree scanning of the human face and then record the process. As this is a 1D scanning process and only scanning degrees are of concern, we put a rotating table to use. An extended shaft is mounted on a rotating table and the cameras are fixed at the end of the shaft. Figure 3 shows the rotary system platform and its dimensions. During scanning, the motor is controlled in rotation from a host computer and motion is conveyed through a central shaft to the rotating bar. When a person is seated directly beneath the shaft, the cameras start rotating and real time video data are sent to a video capturing computer through RCA ports. Two frame grabbers, which are connected to two individual cameras, transfer the video to image frames. After the 180-degree scanning (side to side), multiple views of human face data are acquired.

It is important to capture visual and thermal images at the same time in order to see the variations in the facial images. While a rotary system rotates with a constant speed, visual and thermal images are captured almost at the same time. The frame grabber from Matrox can achieve 30 frames per second. Due to the large size of the images, we limit capture to 5 framers per second with both images having 320*240 image resolution. There is a slight delay (1/30 sec) after capturing one modality; however, there will be no difference in the facial images. Although radiometric calibration is important, we cannot calibrate the thermal camera because of current Raytheon IR camera characteristics and lack of resources such as blackbody source. Figure 3 shows the devices used to capture both visual and uncalibrated thermal images. As shown Figure 3 (a), and (b), visual and thermal cameras are attached in the middle of the bar in the rotary system. Two fixed

lights are used to provide illumination change. The software GUI shown in Figure 3 (c) has been implemented to capture visual and thermal images at the same time. Figure 4 shows an example of visual and thermal images taken at the same time using rotary system.

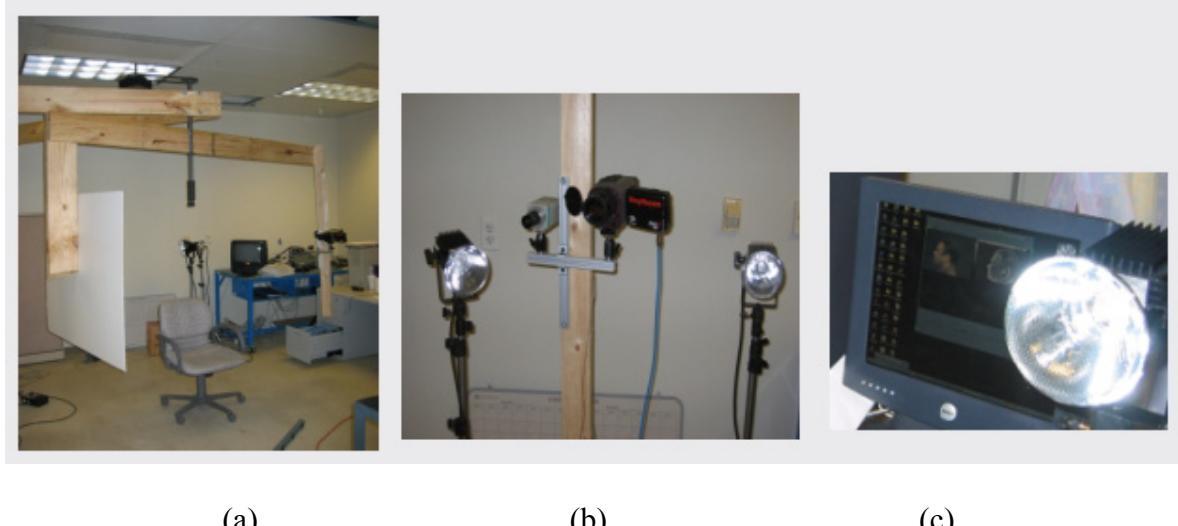


Figure 3: Image acquisition platform. (a) Rotary system, (b) visual and thermal cameras with two lights, and (c) software to capture images at the same time



Figure 4: Example of visual and thermal images taken at the same time

Table 2 shows an overview of the IRIS face database. 32 IRIS lab members have participated in this experiment using 5 different illumination conditions and 3 different expressions while utilizing the rotary system. People who wear eyeglasses had pictures

taken with the eyeglasses off. For testing disguise, 6 people participated in order to check for abnormal temperatures on the faces. The materials used in this test were 2 wigs, 2 sunglasses, 1 goggle, 3 moustaches, and 2 masks.

Figure 5 shows examples of images taken under different illumination and expressions while the rotary system was in operation. This database can be downloaded from <http://imaging.utk.edu/~heo/ECE599>.

Table 2: Summary of the IRIS face database

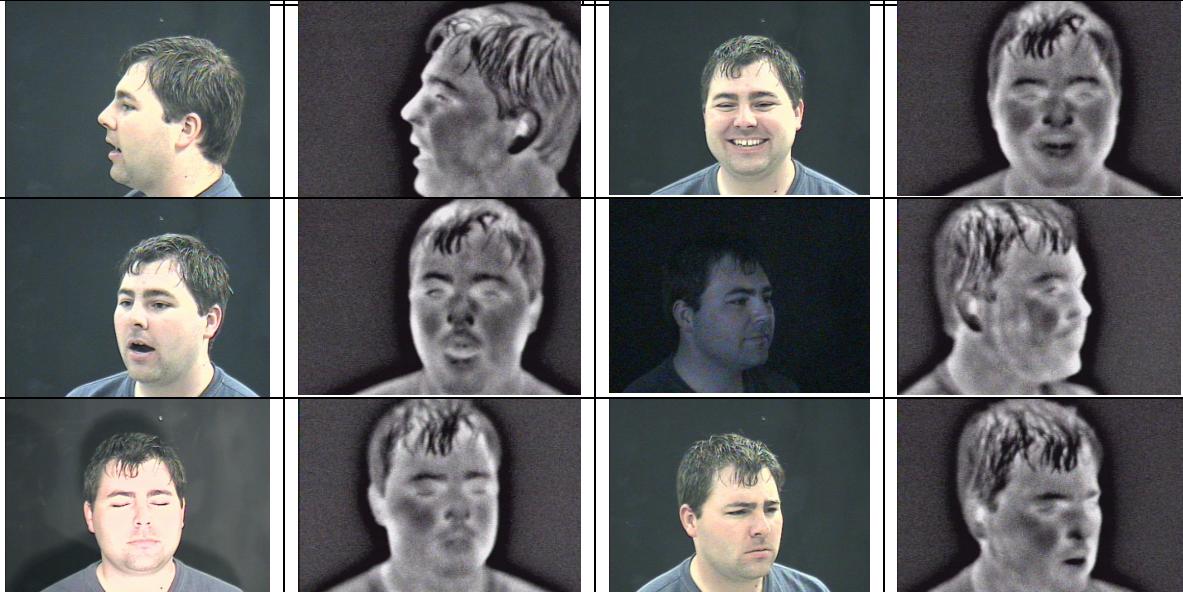
Participants	Number of Images	Variations
32	10,000(5,000/modality, 176~250 images/person, 11 images per rotation)	Illumination (Both on, Left on, Right on, Both off, Dark) Expression (Vowel, Laugh, Angry) Eyeglasses (On, Off) Disguises(Masks, Goggles, Sunglasses, Wigs, Moustaches)
		

Figure 5: Visual and thermal images taken from rotary system

2.2 Fusion of Visual and Thermal Images

With use of the database as shown in Table 1, we fused visual and thermal images. Ideally, the fusion of different sensors can be done by pixel-wise weighted summation of visual and thermal images.

$$F(x, y) = a(x, y)V(x, y) + b(x, y)T(x, y) \quad (1)$$

where $F(x, y)$ is a fused output of a visual image, $V(x, y)$, and a thermal image, $T(x, y)$, while $a(x, y)$ and $b(x, y)$ represent the weighting factor of each pixel. A fundamental problem is: which one has more weight at the pixel level. This can be answered if we know the illumination direction which affects the face in the visual images and other variations which affect the thermal images. Illumination changes in the visual images and facial variations after exercise in the thermal images are also one of challenging problems in face recognition technology. Instead of finding each weight, we make use of the average of both modalities constraining both weighting factors $a(x, y)$, $b(x, y)$ as 1.0. The average of visual and thermal images can compensate variations in each other, although this is not a perfect way to achieve data fusion. Figure 6(c) shows a fused image based on average intensity using (a) visual and (b) thermal images. We have produced the same number of fused images using datasets as described in Table 1. The fused output contains visual and thermal characteristics at the same time, thus, can be regarded to be more informative and synthetic than the original inputs.

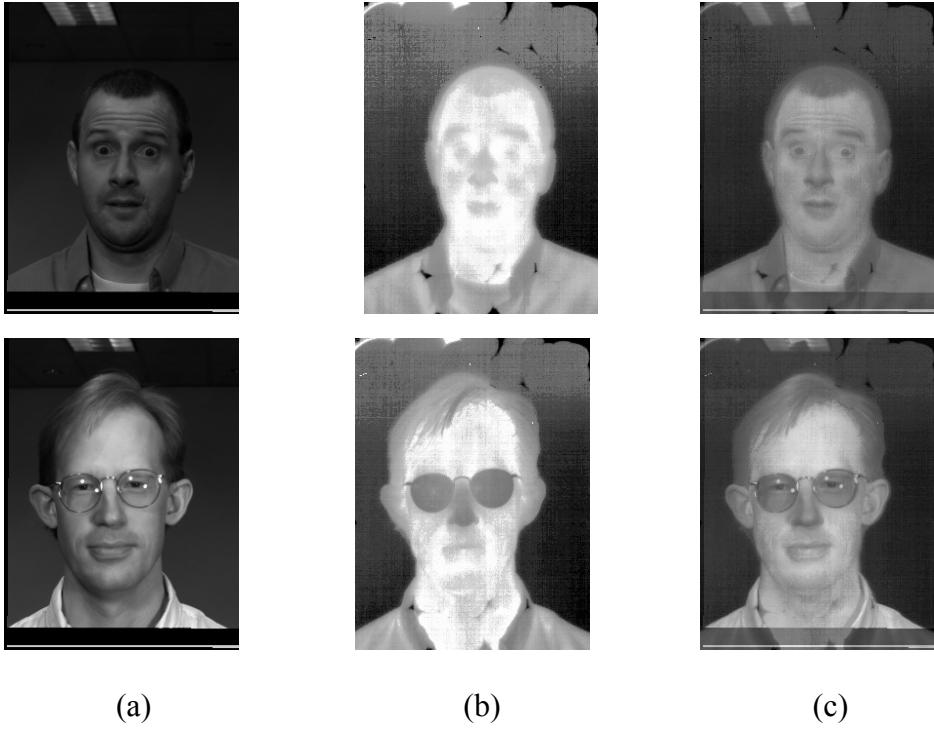


Figure 6: A data fusion example. (a) Visual image, (b) thermal image, and (c) fused image based on average intensity

3. Feature Extraction from Thermal Images

3.1 Variations

It is important to find the most robust features in the thermal images. These robust features can be used for thermal face normalization. The normalization is performed the face extracted from an image is scaled and rotated such that a fixed size results. The eyes in visual images are proven to be a salient method for the normalization of the face. If we use a co-registered camera, the eye locations of the visual images can be used for the thermal images. This means no further processing is required for the normalization of thermal images. If we use a thermal camera only, are the eyes still a salient feature?

Since eyes in visual images have strong edges, and contain eyeballs inside and clear corners, it is relatively easy to process detection of the eyes. Although researcher and scientists have developed eye detection algorithms for visual images, it is still questionable if the eyes (the center of the eyeballs) can be detected when individuals close their eyes, are affected by reflection of lights, or are wearing sunglasses. The eyes (especially the eyeballs) in the visual images are relatively robust features more than any other features such as the mouth, nose, or the size of the head. Jeffrey Huang et al [48] proposed an eye detection method using natural selection (Genetic Algorithm), learning (Decision Trees), and their interactions. Lam et al [49] used a corner detection scheme which can provide information on corner orientations and angles in addition to corner location. Mariani [50] proposed a subpixellic eye detection scheme to reduce greatly the number of possible scales used during the face recognition process. Huang et al [51] proposed a scheme for face location and accurate eyes detection, which is based on multiple evidence, including facial component structure, texture similarity, component feature measurement, the Hough transform and contour matching. After detecting faces, they proposed a precise eye location combining contour and region information extracted from a zoomed image. Jeffrey Huang. et al [52] proposed an approach for the eye detection task using optimal wavelet packets for eye representation and the Radial Basis Function (RBF) for subsequent classification of facial areas as eye vs. non-eye regions. Entropy minimization is the driving force behind the deviation of optimal wavelet packets.

Edge detection methods such as Canny, Sobel, Prewitt, and other edge detection schemes in Matlab 6.5 with optimal thresholds yield unsatisfactory results for further thermal images processing as shown in Figure 7.

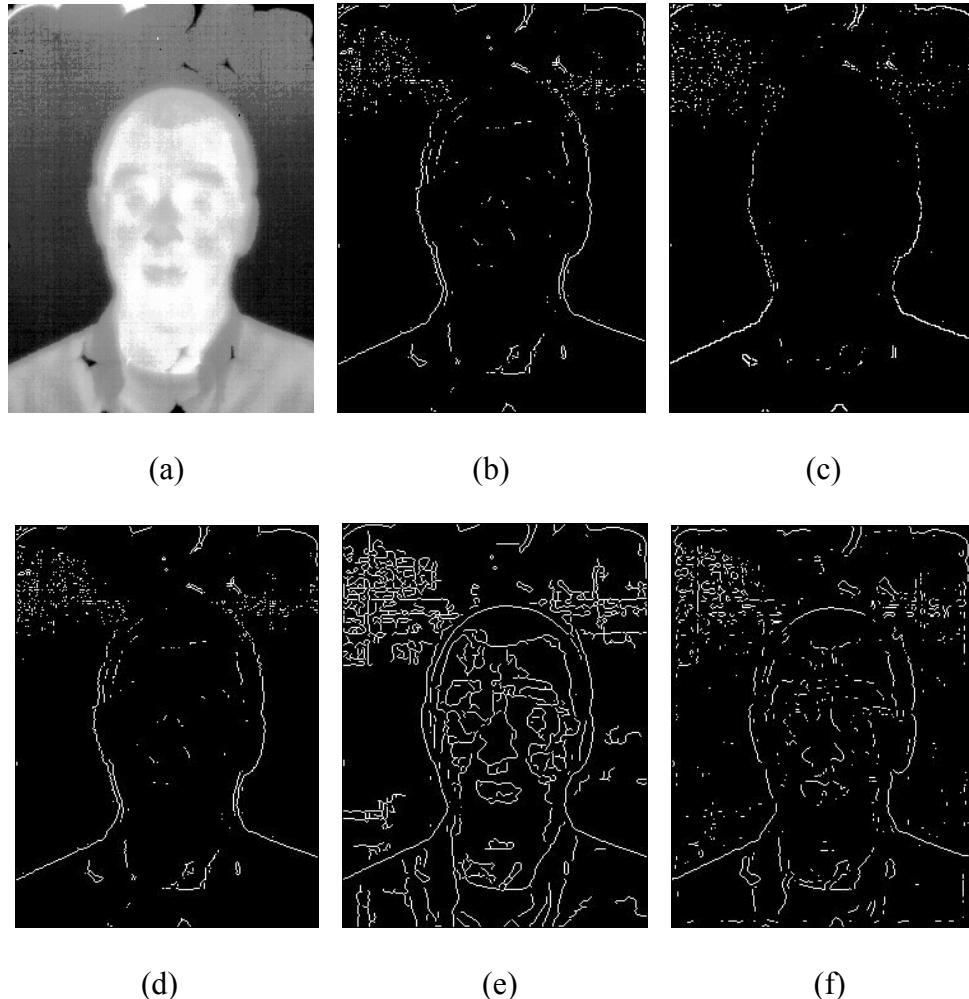


Figure 7: Problems with edge detection results using thermal images; (a) original thermal image, (b) Prewitt, (c) Roberts, (d) Canny, (e) Laplacian, and (f) zero-cross edge detection results

Histogram based automatic thresholds make it even harder to find the optimal threshold for segmenting facial components. Figure 8 shows the histograms of an original and equalized thermal image. Since the optimal threshold, which can be derived from the histogram, does not contain any regional information, the approach to solve local minima or local maximum has its own limitations.

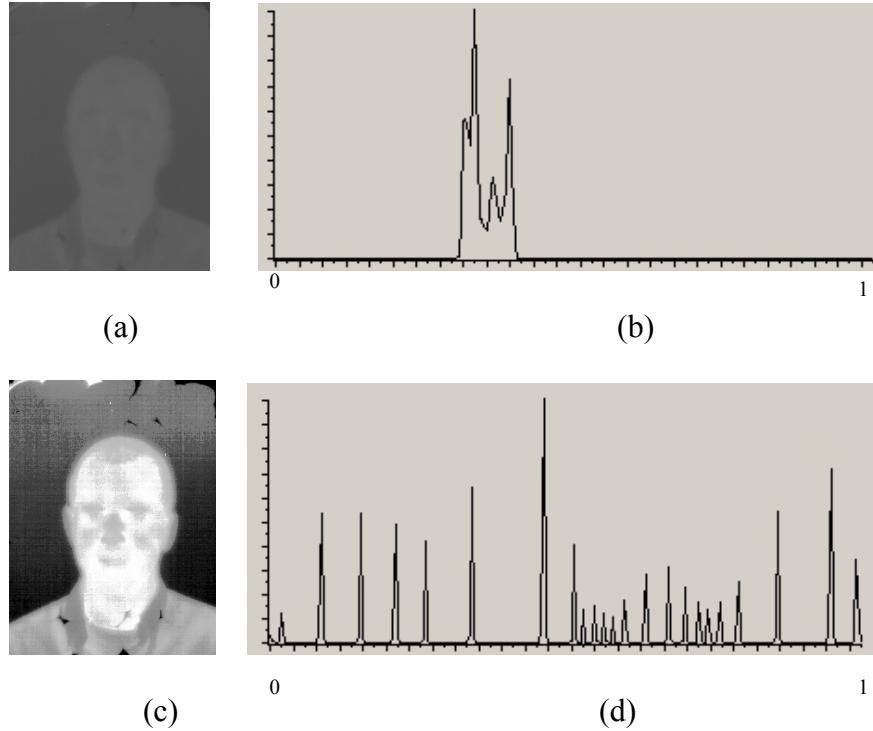


Figure 8: Problems with histogram methods (a) original thermal image, (b) histogram of (a), (c) equalized thermal image, (d) histogram of (b)

Therefore, we make use of average thermal images to measure variation over thermal face images instead of observing each individual thermal face. Figure 9 shows the average of thermal images generated from 60 images without eyeglasses (a) and with eyeglasses (b). Figure 9(c) was generated from 200 images from the FERET database. All images are normalized using manual eye positioning.

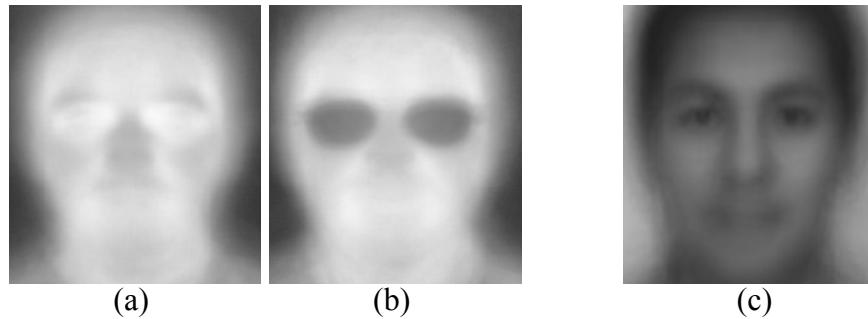


Figure 9: Average thermal images; (a) without eyeglasses, (b) with eyeglasses and
(c) average visual image

From the average thermal images of Figure 9, we measured statistical variations of the thermal images. Figure 10(b) shows the histogram of each thermal image without and with eyeglasses and (c) shows the color-coded thermal images. Skin regions are a relatively hotter area than the background and the nose or eyeglasses regions are the coldest areas in thermal images. From the histogram Figure 10(b), one can observe that it is not simple to use automatic thresholds based on the histogram itself. The maximum threshold was set to 1.

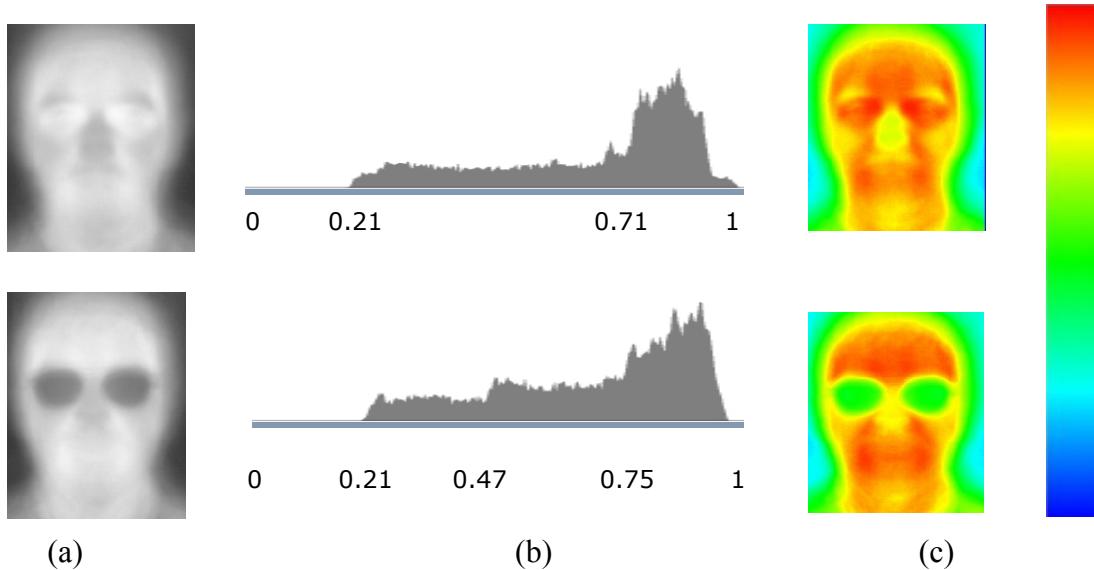


Figure 10: Average thermal images without (top)/with (bottom) eyeglasses, gray
level histograms and color-coded images

In order to see variations among the faces, we cropped each region of the average images.

Figure 11 shows each feature variation among thermal face images. The eyes are relatively the hottest regions. The eyebrows, cheek, and the mouth are all almost in the same range temperature. The nose is the coldest area if individuals are not wearing eyeglasses while eyeglass regions are the coldest area if individuals are wearing eyeglasses.

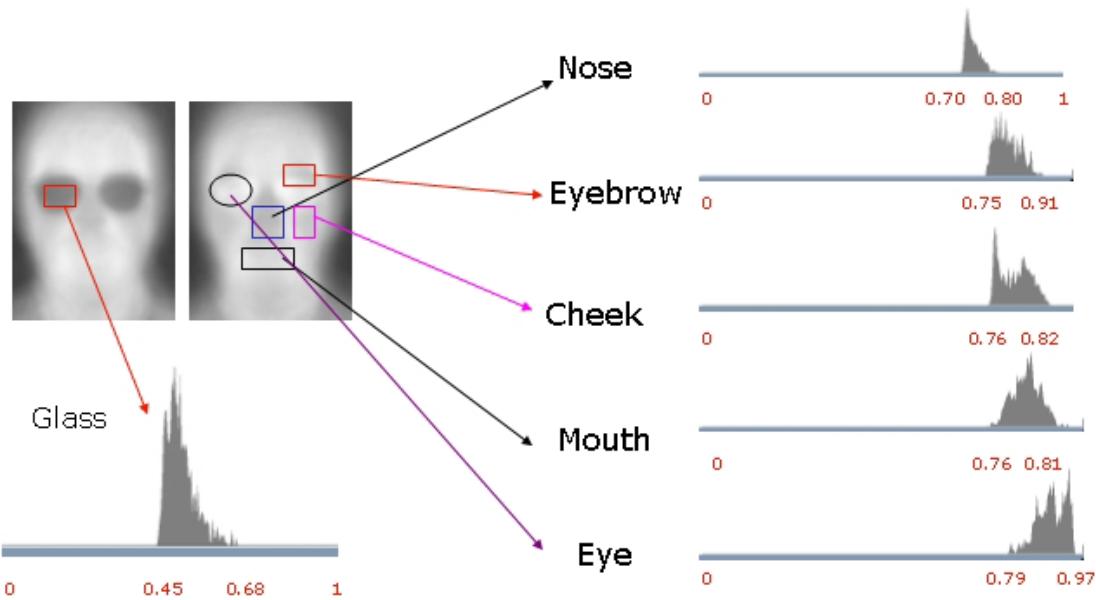


Figure 11: Facial feature variations of average images indicated by histograms

3.2 Face Detection from Thermal Images

Ellipses are often used in face-related technology such as face tracking, face detection, head tracking and other facial component analysis. Since thermal faces are integrated with several features comprised of different blobs, the use of an ellipse can be a powerful representation of certain features around the faces in the thermal images. The general equation of a conic can be represented as

$$F(A, T) = AT = ax^2 + bxy + cy^2 + dx + ey + f = 0 \quad (2)$$

where $A = [a, b, c, d, e, f]$ and $T = [x^2, xy, y^2, x, y, 1]^T$. Commonly used conic fitting methods minimize the algebraic distance in terms of least squares

$$\overset{A}{A} = \arg \min_{A} \left\{ \sum_{i=1}^N (F(A, T_i))^2 \right\}. \quad (3)$$

Researchers suggested different methods to fit conics while applying constraints over $\overset{A}{A}$, such as $a + c = 1, f = 1$, and $a^2 + \frac{1}{2}b^2 + c^2 = 1$. Bookstein [53] showed that if a quadratic constraint is set on the parameters, the minimization can be solved by a generalized eigenvalue system which can be denoted as

$$D^T D A = S A = \lambda C A. \quad (4)$$

where $D = [x_1, x_2, \dots, x_n]^T$ is called the design matrix, $S = D^T D$ is called the scatter matrix and C is a constant matrix. Least squares conic fitting was commonly used for fitting ellipses, but it can lead to other conics. Fitzgibbon et al [54] proposed direct least conic fitting algorithms while applying constraints

$$A^T C A = I. \quad (5)$$

They introduced a non-iterative ellipse-fitting algorithm that yields the best least square ellipse fitting method. Their method has a low eccentricity bias, is affine-invariant, and is extremely robust to noise.

Figure 12 shows a demo for fitting ellipses. We make use of OpenCV which provides high-level image processing libraries to check the robustness of ellipse fitting functions. The original image is processed using a user-specified threshold, then each

component is connected based on a chain coding scheme; finally, ellipses are fitted based on each connected component. Although face regions are not well representative, eye blobs are well fitted. This ellipse fitting method may be helpful to reduce the search time for detecting the eyes even in visual images.

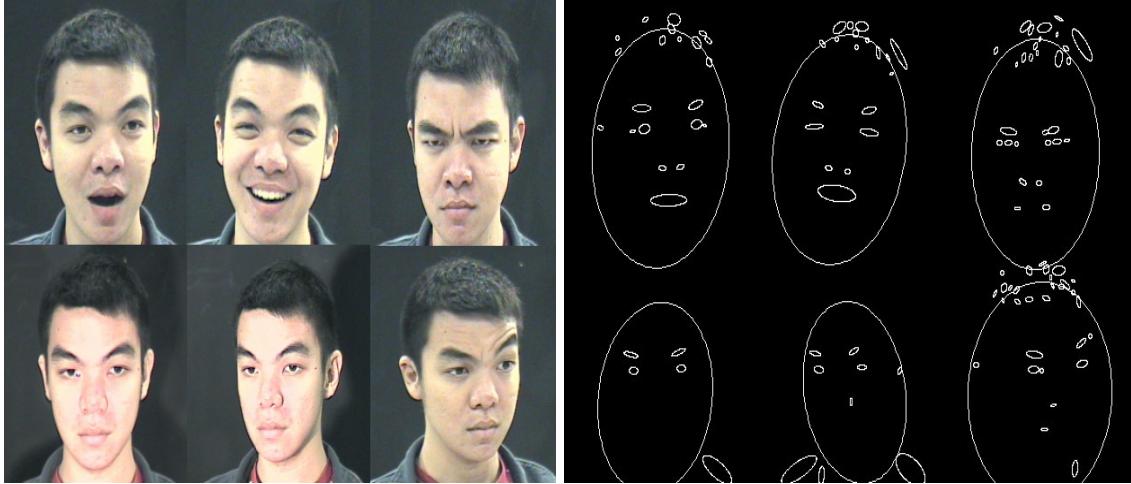


Figure 12: Ellipse fitting demo

Detecting and tracking of face-like objects in cluttered scenes is an important preprocessing stage of overall automatic face recognition. The goal of face detection is to segment out face-like objects from cluttered scenes. Recent survey papers on face detection techniques in visual images can be found in [55][56]. The detection of faces in thermal images can be achieved by applying thresholding in which most skin areas are covered.

Figure 13 shows a basic flow diagram of the face detection algorithm. Original images are binarized with a statistically driven threshold and processed by morphological (opening & closing) to reduce small blobs inside of the faces. Then the results are linked using a Freeman chain-coding algorithm. A small number of connected components is

ignored to achieve better performance in fitting the ellipses. After fitting ellipses for each connected component, only the largest ellipse is considered to be a face candidate. The face candidate will be processed for further analysis of facial features for the normalization of faces.

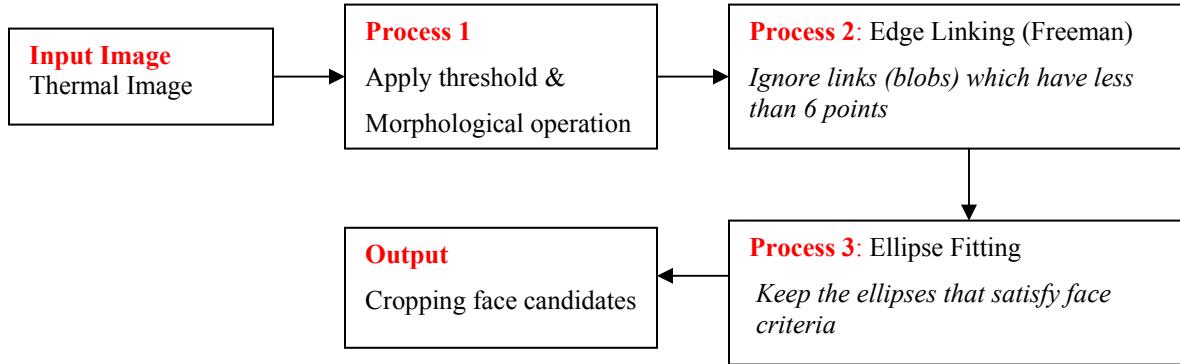


Figure 13: Face detection diagram

Figure 14 shows a face detection example based on the diagram as shown in Figure 13. Input thermal images as shown in Figure 14 (a), binarized in (b), and processed by morphological operations in (c). Connected components (d) are fitted with ellipses (e). Finally, face regions are cropped (e) for further analysis. We achieved fairly high performance with almost every face captured with the threshold of 0.70, which is derived from the minimum temperature of the nose from an average thermal image as shown Figure 11, though the cropped face regions have multiple candidates from dividing the upper head region and lower head region in some cases. These inexact face candidates should be processed more in the next stage by using nose and eyebrow detection to verify the faces.

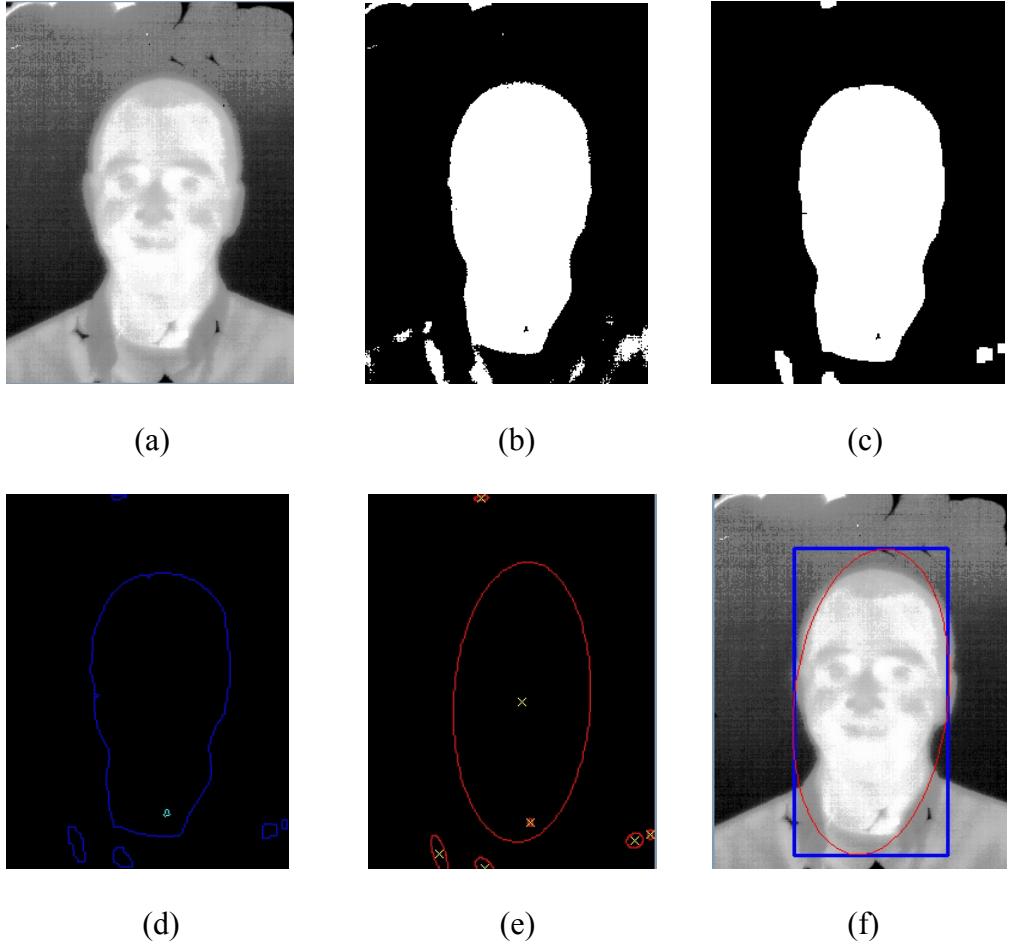


Figure 14: Face detection example: (a) thermal image, (b) thresholded image, (c) after morphological operation, (d) connected components, (e) fitting ellipses, and (f) cropped face region

3.3 Eyeglass Detection and Removal from Thermal Images

One of the disadvantages of using thermal images for face recognition will be the fact that eyeglasses block thermal emissions from the face area. Automated eyeglass detection is thus important in implanting further processing for recognition. Eyeglass regions are hotter than the background and cooler than any of the facial components. After detecting face regions eyeglasses can be detected in the same manner using face

detection algorithm as described above. Figure 15 shows classification of eyeglasses in thermal images. A thermal image is thresholded and ellipses are fitted. Based on the ellipses, each blob is processed to check similarity. In case where one blob is detected, we determine there are no eyeglasses since we are using only frontal images. Total comparison for N blobs will be $N(N - 1)/2$. We computed similarity of two blobs in terms of α , which represents the ratio of major and minor axis of a ellipse, β , which represents the ratio of size (major*minor) of two ellipses, θ which represents the orientation of two blobs, γ , which represents the distance of two blobs compared to the width of the face, and κ , which represents the location of ellipses relative to the height of the face. After checking the similarity, only the best two blobs will remain for claiming the presence of eyeglasses.

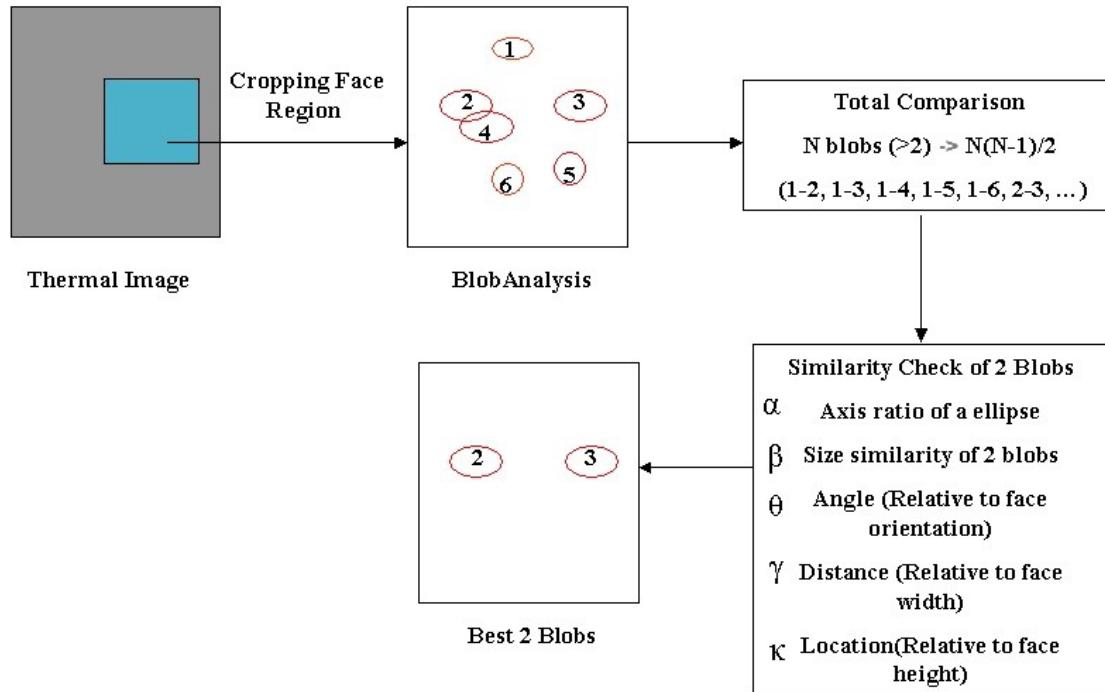
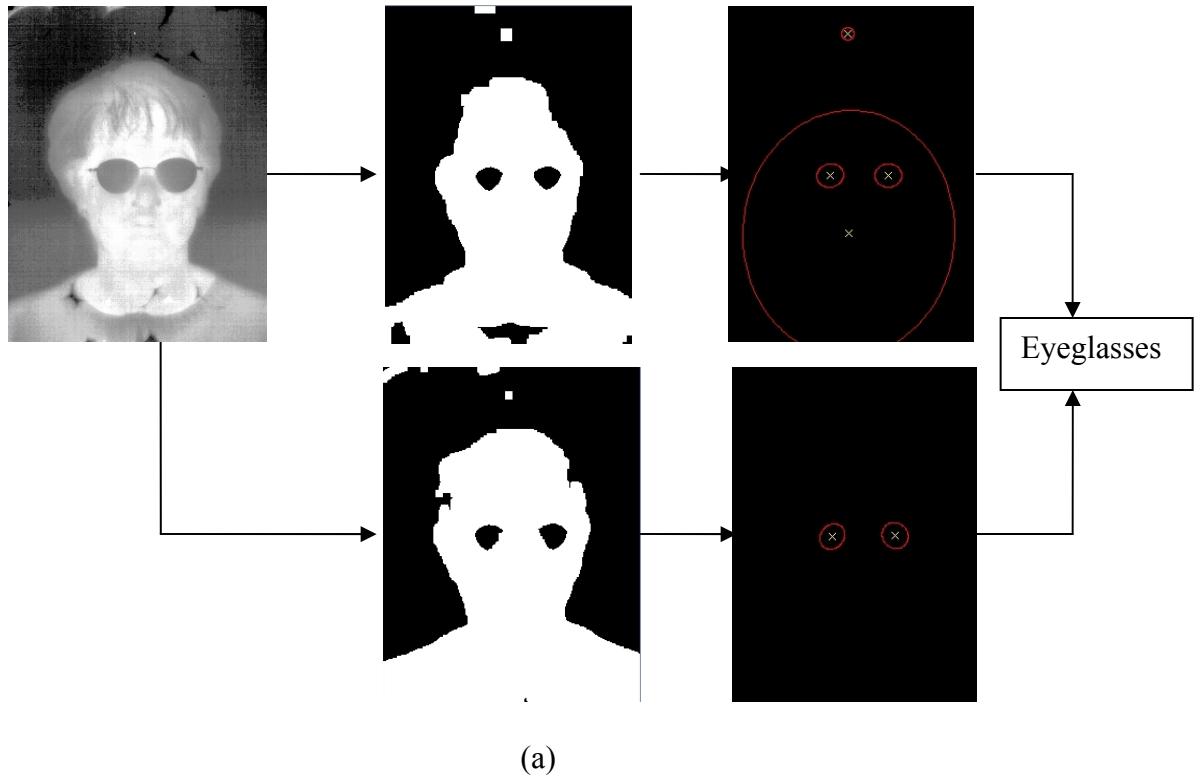
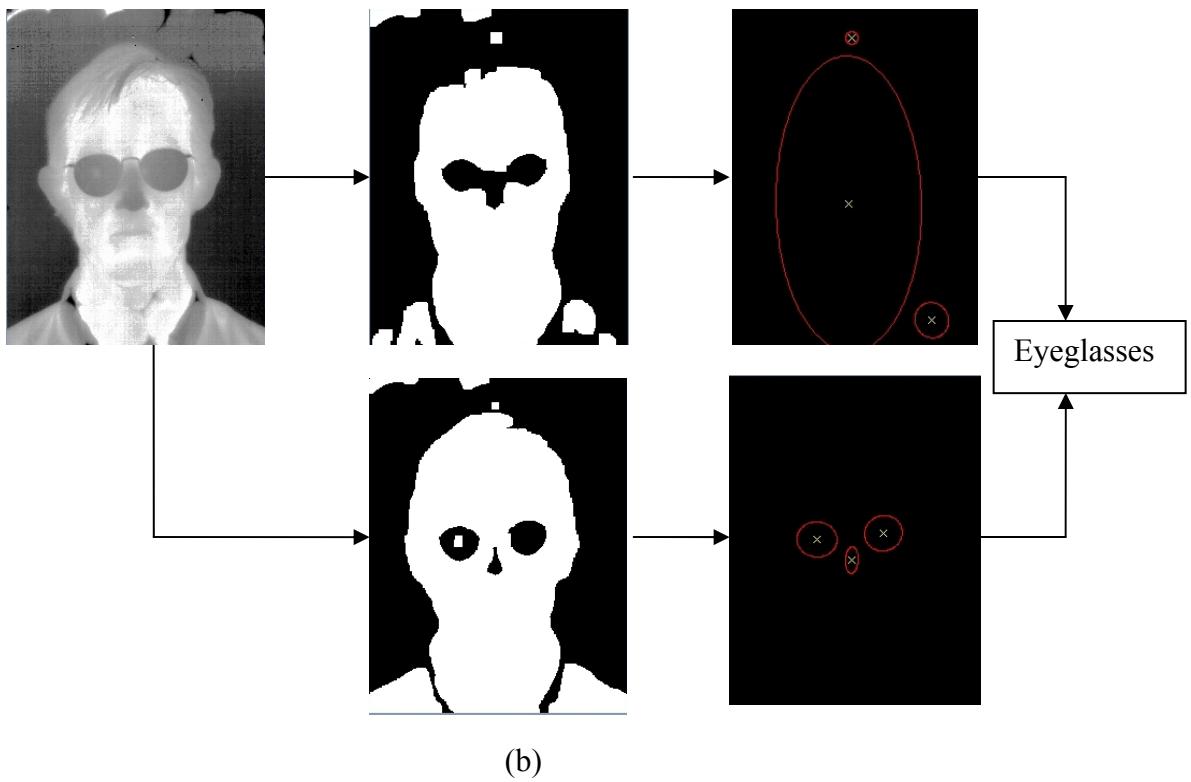


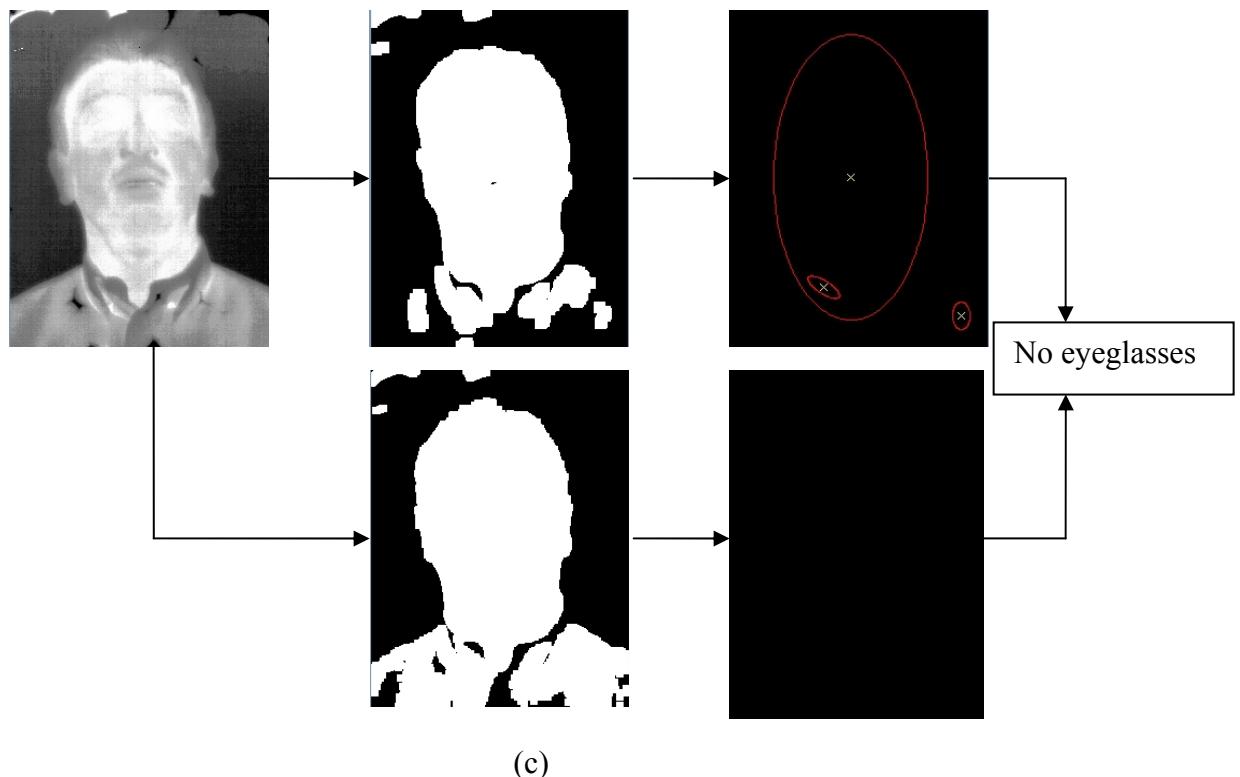
Figure 15: Diagram of eyeglass detection algorithm

Figure 16 shows an example of detecting eyeglasses. In Figure 16, the first threshold is used for face detection. The detection of faces is important to restrict the presence of eyeglasses and reduce the searching area. After applying the first threshold, 70% of eyeglasses are detected at the same time. In some cases as shown in Figure 16(b), the threshlded images in the 1st row are grouped together which result in incorrect eyeglass detection. Although it fails in the 1st threshold, the 2nd threshold can detect the presence of eyeglasses as shown in Figure 16(b) 2nd row images. The optimal threshold can be found after evaluation of the performance of the glass detection algorithm in terms of False Acceptance Rate (FAR) and False Rejection Rate (FRR) as shown in Figure 17.





(b)



(c)

Figure 16: Eyeglass detection example; (a), (b) eyeglasses are detected, (c)
eyeglasses are not detected

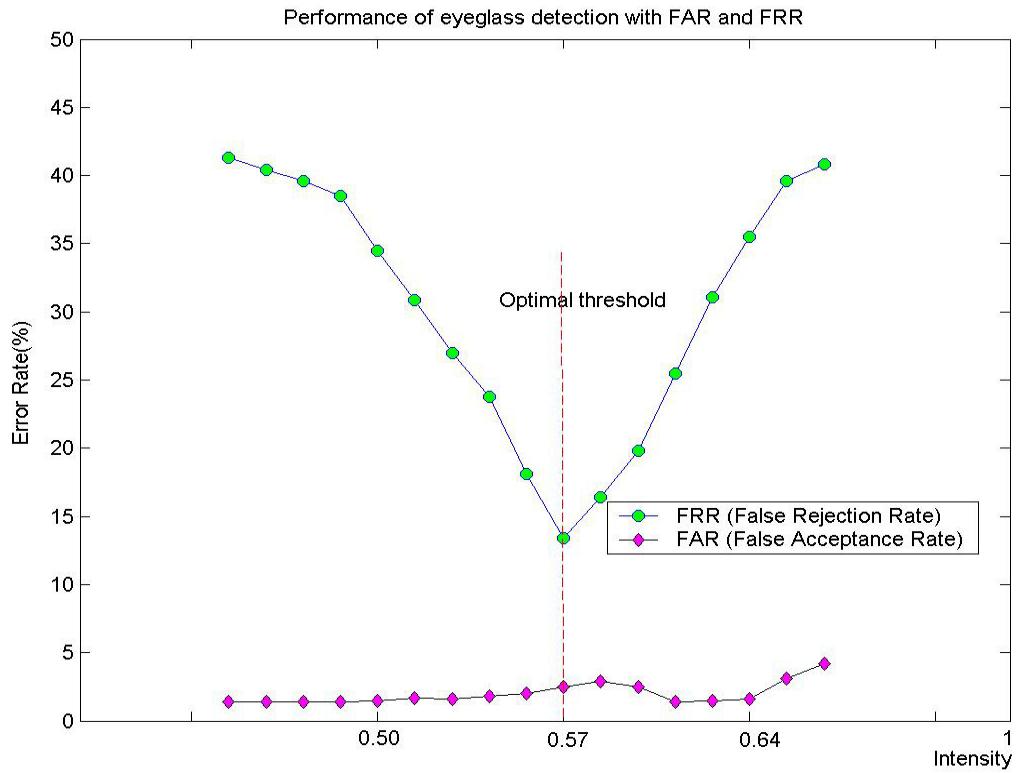


Figure 17: Performance of eyeglass detection with FAR and FRR

Table 3 shows the performance of the eyeglass detection algorithm with two fixed thresholds (face-0.70, eyeglass -0.57) as shown in Figure 17. Correct detection when eyeglasses are worn is 86.6% and when no eyeglasses are worn is 97.1%. The parameters used in eyeglass detection are $0.5 < \alpha < 1.5$, $0.8 < \beta < 1.2$, $\theta < |0.2(\text{radian})|$, $0.3 < \gamma < 0.9$, and $0.4 < \kappa < 0.8$. These parameters can be adjusted. If we apply the parameters more restrictively, correct detection will be decreased while incorrect detection will be increased.

Table 3: Performance for eyeglass detection

Descriptions	Matched /Total Images	Accuracy
True Positive (Eyeglass-> Eyeglass)	445/ 514	86.6%
True Negative (No Eyeglass -> No Eyeglass)	1096/1129	97.1%
False Positive (No Eyeglass -> Eyeglass)	33/1129	2.9%
False Negative (Eyeglass ->No eyeglass)	85/514	13.4%

After detecting eyeglasses, it will be better to replace the eyeglass regions with average eye templates in the thermal images to minimize the effect of blocked information. Figure 18 shows the average eye template to be replaced with eyeglasses regions and Figure 19 shows example images after eyeglasses removal in the thermal images and fused images. Here we used only one eye template but it will be better to use different types of eye templates depending on the size, shape and coverage regions of the eyeglasses.

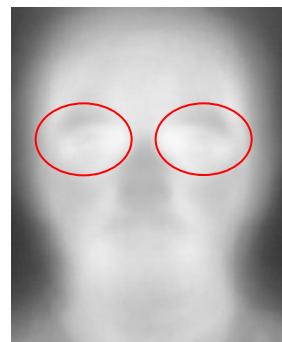


Figure 18 : The average eye regions template

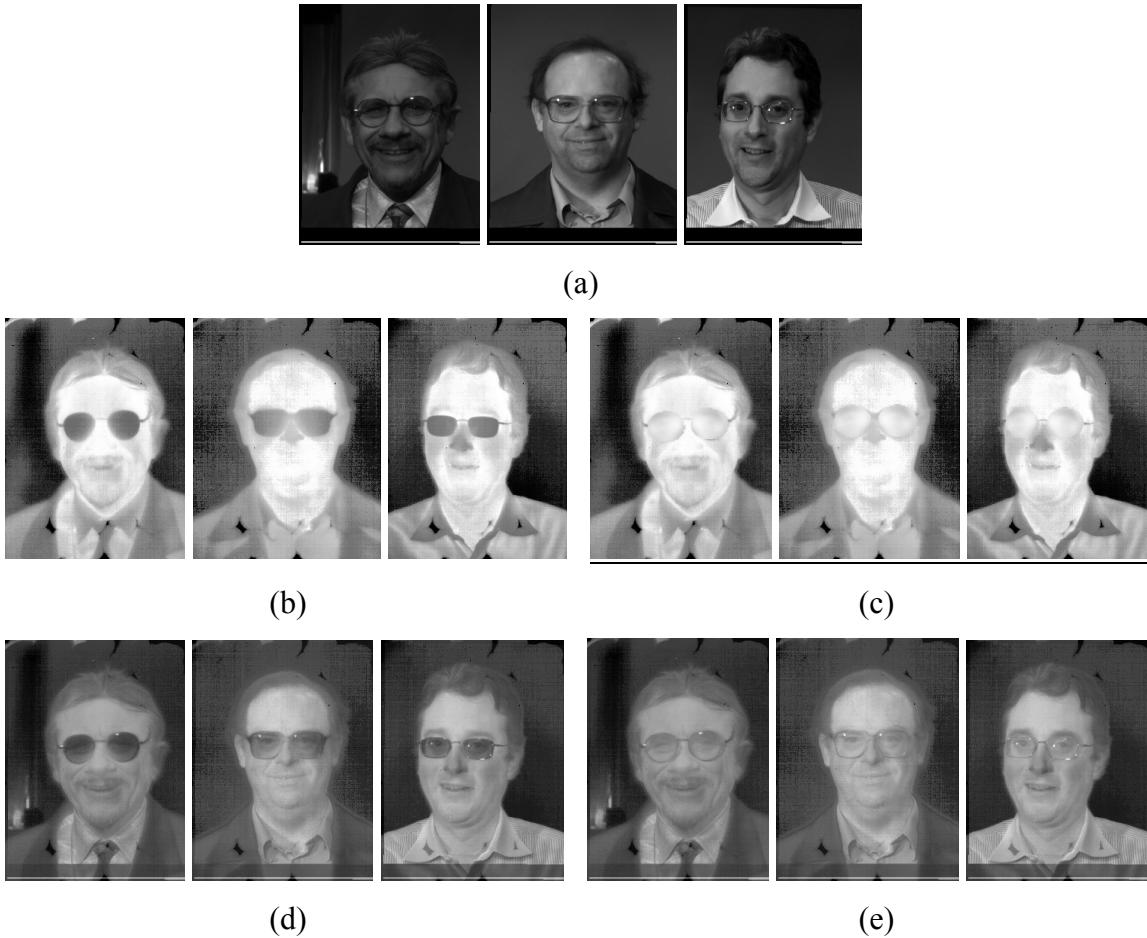


Figure 19: Data fusion after eyeglass removal; (a) original visual images, (b) original thermal images with glasses, (b) thermal images after eyeglass region replacement, (d) original fused images, and (e) fused images after eyeglass region replacement

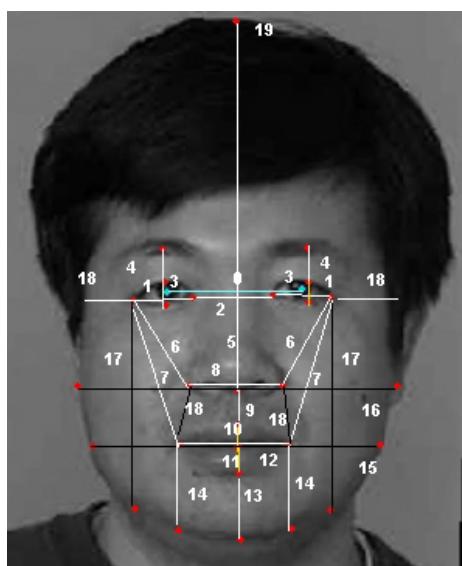
4. Comparison of Face Recognition Techniques

In this chapter, comparisons of visual face recognition have been performed with feature based methods and appearance based methods using FERET database. The results then are compared with FaceIt® which uses local feature algorithm. The performance of Visual and thermal face recognition are also evaluated using FaceIt® under different fusion methods.

4.1 Feature Based Face Recognition

We used 610 faces (200 gallery images (fa images)), 410 subject images (fb images)) from the FERET database to evaluate analytic and feature based face recognition algorithms. We did not use all of the images provided by FERET but selected only those suitable for this experiment. The 2-letter codes (fa, fb, and etc) indicate the type of imagery. For example, fa indicates a regular frontal image. Detailed naming conventions can be seen at http://www.itl.nist.gov/iad/humanid/feret/feret_master.html.

It is important to extract this distance information considering their significance in representing a face, and reliability in extracting faces automatically. In addition, it is important to weight distance information, because each works differently in the recognition process. We used geometric distances, which were extracted from 30 landmark points as shown in Figure 20, and different sets of landmarks as shown in Figure 21.



0. Iris Distance – For Normalization
1. Eye width (average of two eye lengths)
2. Eye separation
3. Eye height (average)
4. Eye-eyebrow distance(average)
5. Horizontal nose position
6. Eye-nose distance(average)
7. Eye-mouth distance(average)
8. Nose length
9. Nose-mouth distance
10. Top-lip thickness
11. Bottom lip thickness
12. Mouth width
13. Mouth-chin distance
14. Mouth corner- chin distance(average)
15. Face width1(based on mouth)
16. Face width2(based on nose)
17. Eye-chin distance(average)
18. Eye-side distance(average)
19. Face height

Figure 20: Geometrical distances on the face [image from FERET]

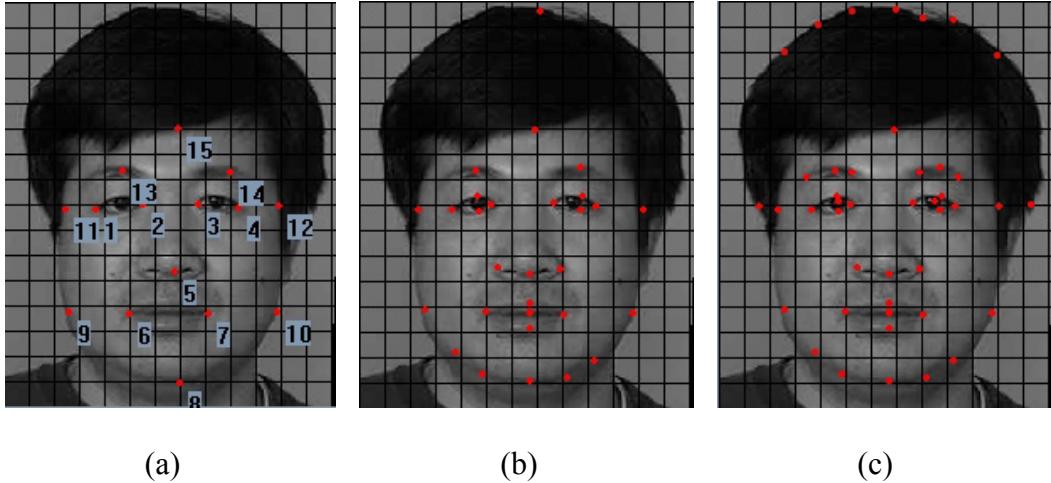


Figure 21: Different landmarks; (a) 15 points, (b) 30 points, and (c) 44 points. [images from FERET]

It is apparent that the eyes, nose, and mouth regions are also useful for recognition. Humans also recognize faces using these features. We also extracted these regions based on landmark points around them for face recognition as shown in Figure 22. Here, we need to also find out how to perform optimum weighting since those regions work differently in the recognition stage.

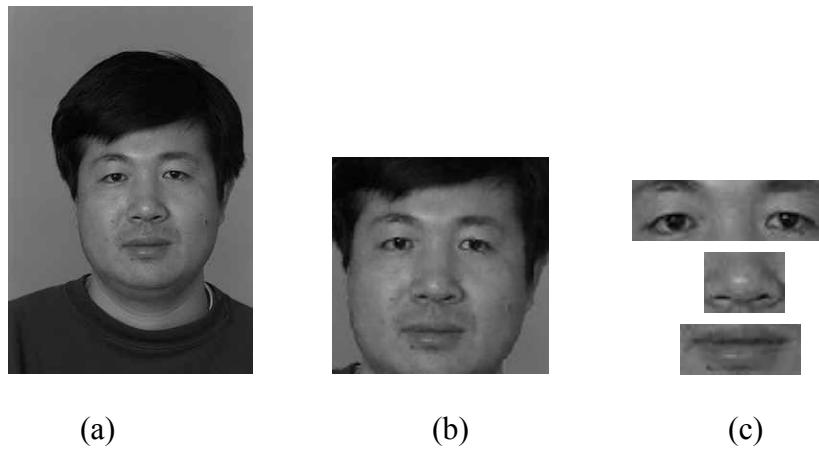


Figure 22: Features based on face regions. (a) Original face image (256*384 pixels), (b) normalized face image (200*200), and (c) extracted eye (30*95), nose (30*40), and mouth (25*60) regions.

In addition to feature based methods, we use the Eigenface algorithm which is a part of holistic based face recognition algorithms as shown in Figure 23.

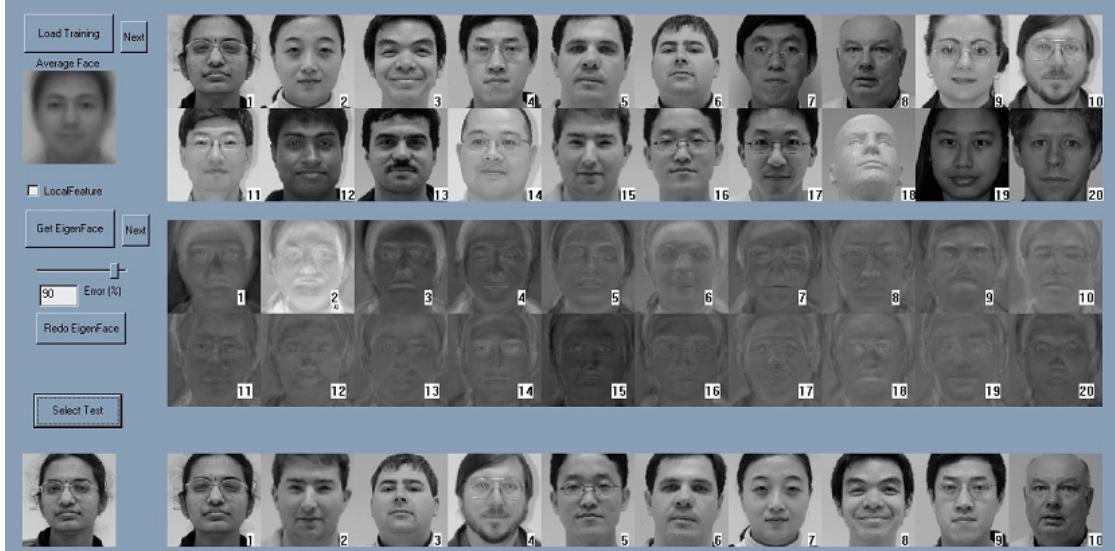


Figure 23: Eigenface-based face recognition system

A detailed description of Eigenfaces is as follows. Each face image is a column vector in a matrix. Here we use a 100*100 image size, then, the column matrix X_i becomes 1*10000. Let $T = [X_1, \dots, X_N]$, where N is the number of faces. After normalizing each

matrix with an average face ($A_i = X_i - m; i = 1, \dots, N$) where $m = \sum_{n=0}^{N-1} X_n$), the covariance

matrix $C = AA^T$ (10000*10000) can be calculated. Finding the eigenvectors of the matrix with 10000*10000 dimensions is an intractable task, we compute $L = A^T A$ first, then use the eigenvector equation; $A^T A v_i = \lambda_i v_i$, where λ_i represents the eigenvalues and v_i represents the eigenvector of L. Finally, the eigenvectors U (called Eigenfaces) can be achieved by multiplying the normalized training set A with eigenvectors generated from L using these equations; $AA^T A v_i = \lambda_i A v_i$ and $U = AV (U = [u_1, \dots, u_M], V = [v_1, \dots, v_M])$.

In face recognition, the shortest distance between the gallery and subject can be measured and the first matched for verification or the first 10 or 100 can be matched for identification. The Euclidean distance, D , is used to compare two feature sets,

$$D = \sum_{i=1}^N \| P_i - Q_i \|^2, \quad (6)$$

where N is the number of the features and P_i and Q_i is the matrix of the feature sets.

Since the features of the faces have different degrees of significance and reliability for representing a face, the weighted Euclidean distance, D_w , should be used to compare two feature sets. If we choose different feature sets, they will contribute differently in the recognition process. The main problem is how to find each weight.

$$D_w = \sum_{i=1}^N k_i \| P_i - Q_i \|^2, \quad (7)$$

where k_i is the weighting factors for P_i and Q_i .

Figure 24 shows the performance of different sets of landmark points and geometrical distances, while Figure 25 shows FaceIt® outperform our experimental results. This indicates that the features what we extract is not a good features for face recognition than FaceIt® has. A plot of probabilities of correct match versus the number of best similarity scores called a cumulative match characteristic curve (CMC) is mainly used in this paper for the evaluation of performance of face identification. Another performance method called a receiver characteristic curve (ROC), a plot of numerous false acceptance rate and false rejection rate combinations, which often is used for the evaluation of face verification, will be included in the future.

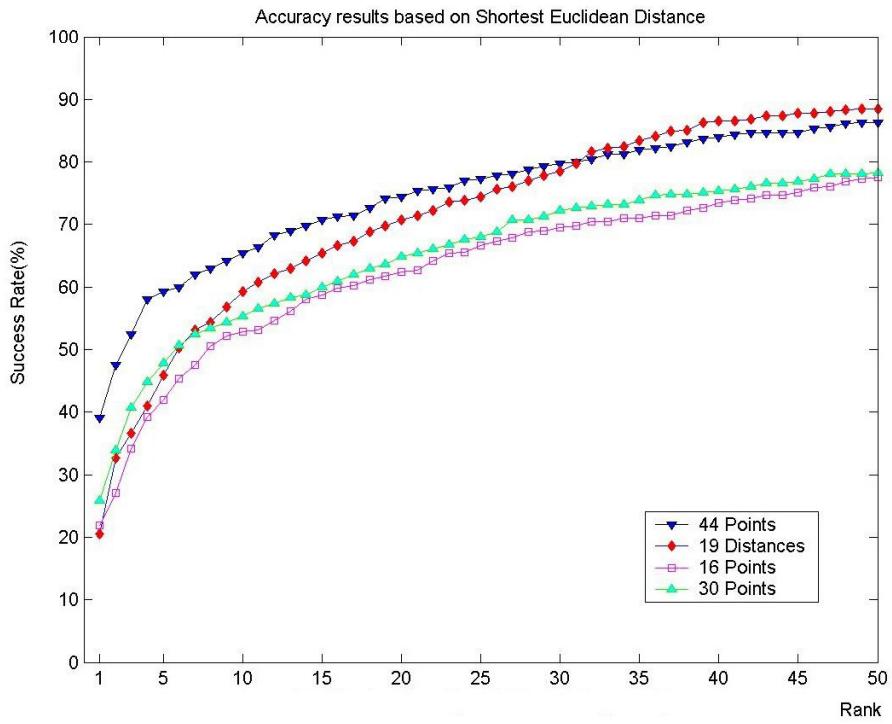


Figure 24: Accuracy results based on shortest Euclidean distance (distance, points)

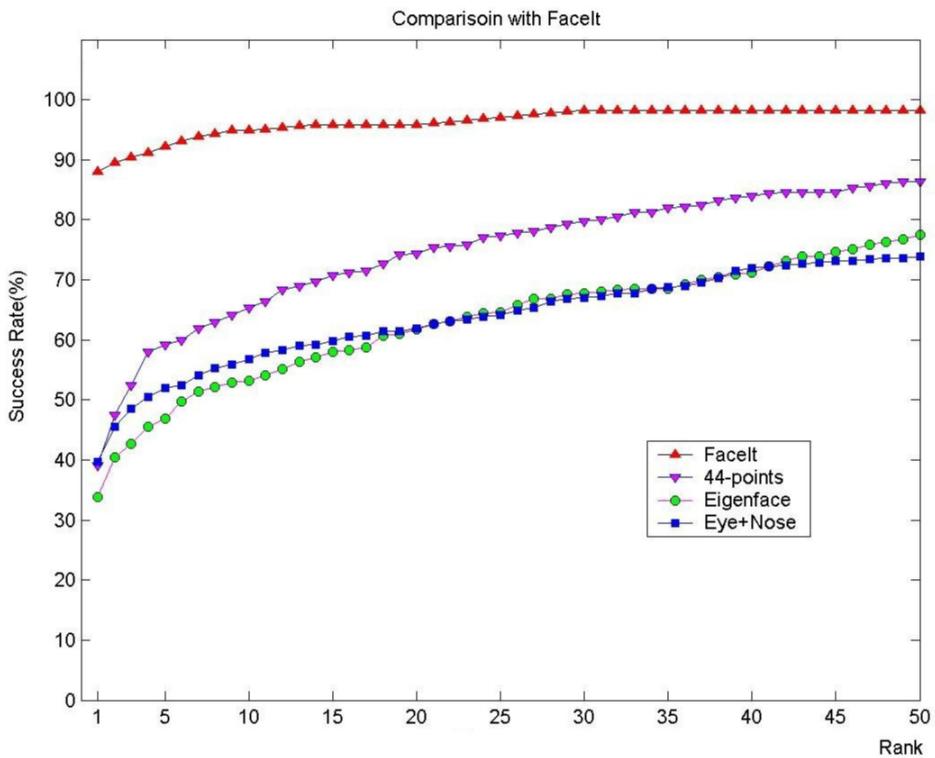


Figure 25: Comparison of our experiments with FaceIt®

4.2 Appearance Based Face Recognition

Before doing a comparison of visual and thermal images, we evaluated FaceIt® identification software, which uses still images, and FaceIt® surveillance, which uses real time video images, to check the robustness of current, best face recognition system [57][58]. FaceIt® is based on the local feature analysis (LFA) algorithm [14] and represents facial images in terms of local features derived statistically from a representative ensemble of faces. Figure 26 shows the FaceIt® software templates which utilizing the following steps. First, the gallery database, which includes faces images, should be created (a). Then, subject images, which we want to identify, should be chosen (b). Finally, after aligning the subject image and then clicking the search button, the results can be seen according to the confidence rate (c).

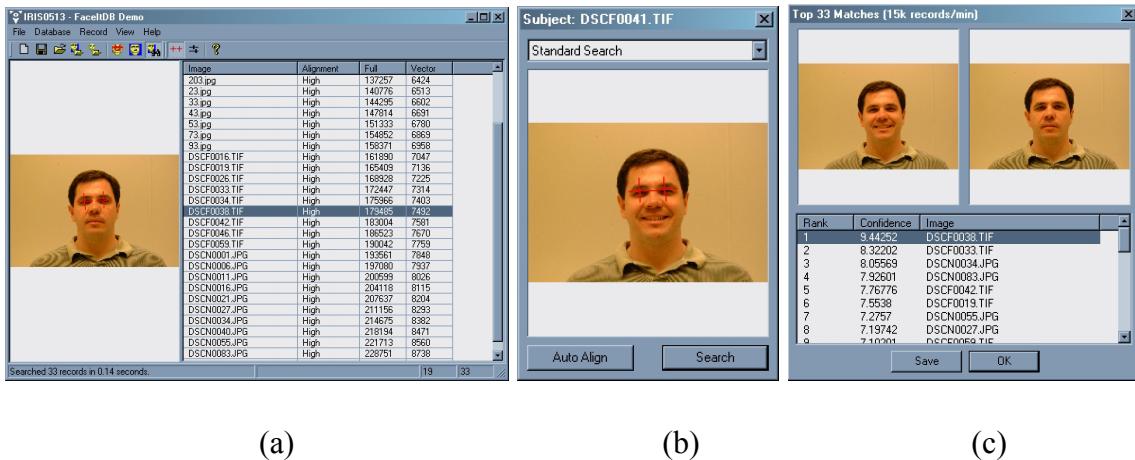


Figure 26: FaceIt® software templates; (a) gallery (visual or thermal) images, (b) subject image, (c) matched ranks sorted by confidence rate

In the identification experiment, we mainly focused on several facial image variations such as expression, illumination, age, pose, and face size which are the main concerns for facial recognition technology. According to the FERET evaluation report,

other factors such as compression and media type do not affect the performance and are not included in this experiment. We divided the evaluation into two main sections with an overall test and a detailed test. In the overall test, we can see the overall accuracy rates of FaceIt®. In the detailed test, we can see what variations affect the system's performance. For lack of databases with mixed variations, we only considered one variation at a time in the facial images for the detailed test. Figure 27 and

Figure 28 show example images of the same individuals under different conditions such as expression, illumination, age, and poses. Table 4 shows a summary and description of the tests included in this section. The overall performance of FaceIt® Identification for 1st match is about 88%. FaceIt® also works well under expression, illumination and face size variations in cases where these types of variations are not mixed. Age variation has proven to be a challenging problem for FaceIt®.

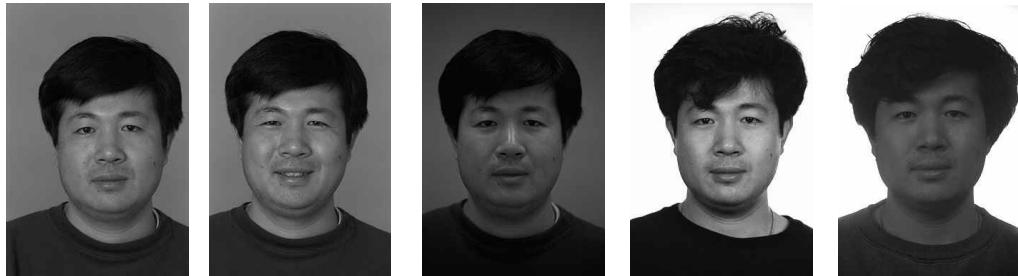


Figure 27: Example images of same individual under different conditions tested with FaceIt® Identification [images from FERET]



Figure 28: Example images of same individual with different poses

Table 4: Experimental results for FaceIt® Identification

Tests	Gallery	Subject	1 st Match Success Rate (%)	1 st 10 Matches Success Rate (%)
Overall Test	700(fa)	1,676(fa, fb)	1,475 (88.0%)	1,577 (94.1%)
Detailed Test				
Expression	200(ba)	200(bj)	197 (98.5%)	200 (100 %)
Illumination	200(ba)	200(bk)	188 (94.0%)	197 (98.5%)
Age	80(fa)	104(fa)	83 (79.8%)	99 (95.2%)
Pose	200(ba)	200(bb ~bh) /pose	Frontal image gives the best result.	
Face Size	200(ba)	200(ba)	No affect as long as the distance between the eyes is more than 20 pixels.	

In the pose test, we still achieve good accuracy rates within $\pm 25^\circ$ poses. Table 5 and Figure 29 show a summary of the pose tests (R-right rotation, L-left rotation). The greater the pose deviations from the frontal view, the less accuracy FaceIt® achieved and the more manual aligning required. Table 6 shows a description of the tests that were not included in this report and the reasons why they were not included.

Table 5: Summary of pose test

Pose(R, L)	First Match (%)	Within 10(%)	Manual Aligning Required (%)
90°L	N/A	N/A	100.0
60°L	34.5	71.0	13.5
40°L	65.0	91.0	4.5
25°L	95.0	99.5	2.5
15°L	97.5	100.0	0.5
0	100.0	100.0	0.0
15°R	99.0	99.5	0.0
25°R	90.5	99.5	2.0
40°R	61.5	87.5	4.5
60°R	27.5	65.0	11.0
90°R	N/A	N/A	100.0

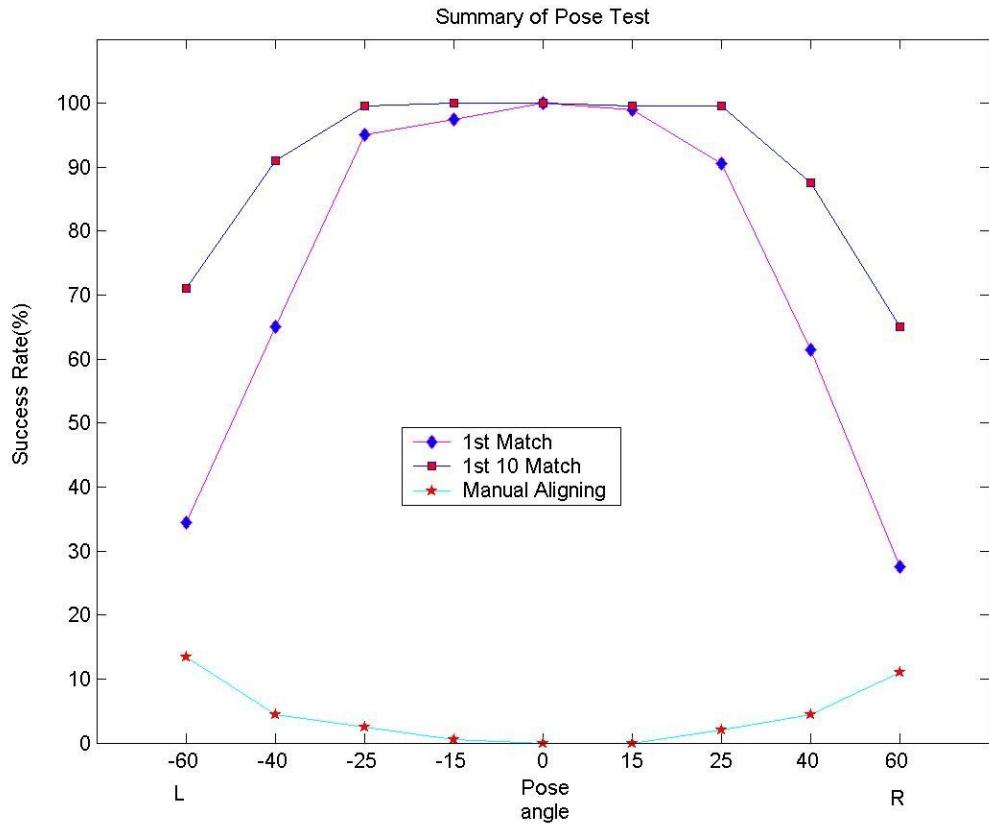


Figure 29: Pose test summary

Table 6: Test items not included in FaceIt® identification experiment

Not included	Description	Reason
Compression	Different compression ratios by JPEG	Does not affect performance
Media	Images stored on different media CCD or 35 film	Does not affect performance
Image type	BMP, JPG, TIFF, etc	Does not affect performance
Temporal	Time delay of a photo taken	Covered by overall and age test
Resolution	Image resolution	Features should be seen clearly

In the surveillance experiment, we used a small PC camera from Logitech which was attached to a PC using a USB port to acquire live face images from real scenes. Since we do not have standard test procedures for the surveillance test, we used randomly captured face images and matched these against databases which were used previously in the

FaceIt® identification test. In order to see the effects of variations, we applied different database sizes (the small DB was the IRIS database which contains 34 faces while the large DB was 700 faces from FERET plus the IRIS DB) and lighting conditions to face images. Since face variations are hard to measure, we divided variations such as pose, expressions and age into small and large variations. Figure 30 shows an example of captured faces used in the experiment. When we captured the faces, any person with significant variations such as quick head rotation or continuous or notable expression changes was considered as a large variation, while the others were considered as small variations.



Figure 30: Example images of captured faces for FaceIt® Surveillance experiment

Table 7 provides a result summary for this experiment. The elapsed time between preparation of the IRIS database and the captured faces was approximately 3 months. The basic distance between the camera and the faces was 2~3 ft. The detailed test only used a person who seemed to be moderately well recognized in the overall test. From detailed tests 1 to 4, we can see how the database size and facial variations affect performance. From detailed tests 3 to 8, we can see how lighting can affect the performance. We can also observe how distance affects the performance from detailed tests 8 and 9. For the lighting conditions, we set ‘High’ as indoor ambient illumination conditions and

‘Medium’ as not ambient but still recognizable by human eyes. ‘Front’, ‘Medium’, ‘Side’, and ‘Back’ tell the direction of additional lighting.

Table 7: Summary of experimental results (basic distance 2~3ft, time elapsed 3 months,
Sub: Subjects)

Test No.	Description	Gallery DB size	Lighting	Captured faces /individuals	1 st match (Num/Sub)	1 st 10 matches (Num/Sub)
Overall 1	Small DB & Large Variations	34	High & Front	758/13	55.8 % (423/758)	96.6 % (732/758)
Detail 1	Small DB & Large Variations	34	High & Front	200/1	55.0 % (110/200)	99.0% (198/200)
Detail 2	Large DB & Large Variations	734	High & Front	200/1	47.5 % (95/200)	78.5 % (157/200)
Detail 3	Small DB & Small Variations	34	High & Front	200/1	67.0 % (134/200)	99.0% (198/200)
Detail 4	Large DB & Small Variations	734	High & Front	200/1	60.5 % (121/200)	93.0 % (186/200)
Detail 5	Small DB & Small Variations	34	Medium	200/1	34.0 % (68/200)	96.5.0% (193/200)
Detail 6	Small DB & Small Variations	34	Medium & Side	200/1	60.5 % (121/200)	98.5% (197/200)
Detail 7	Small DB & Small Variations	34	Medium & Back	200/1	32.0 % (64/200)	80.5 % (161/200)
Detail 8	Small DB & Small Variations Dist: 9~12ft	34	Medium	200/1	0.0 % (0/100)	16.0% (16/100)
Detail 9	Small DB & Small Variations Dist: 9~12ft	34	Medium & Front	200/1	5.0 % (5/100)	78.0 % (78/100)

Figure 31 shows the effects of database size and variations while Figure 32 addresses lighting and distance. A small DB, small variations, close distance, high lighting and additional frontal lighting result in best performance. This is why controlling the environment is important to achieve better performance. The main factors that degrade the performance will be feature changes due to low resolution in captured facial images which are caused by distance, lighting and expressions.

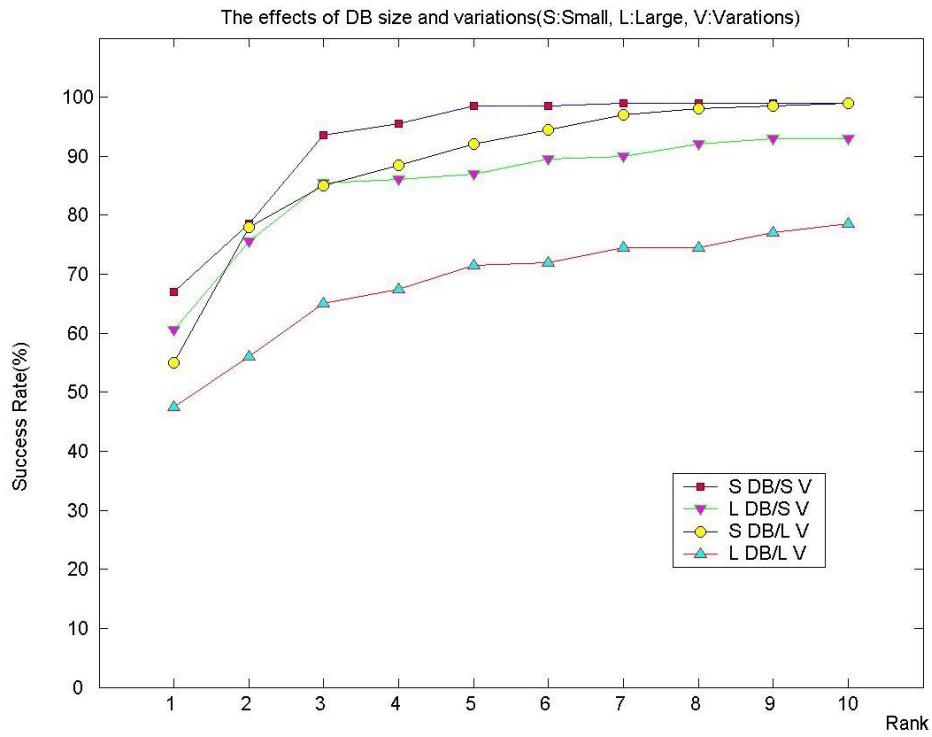


Figure 31: The effects of DB size and variations

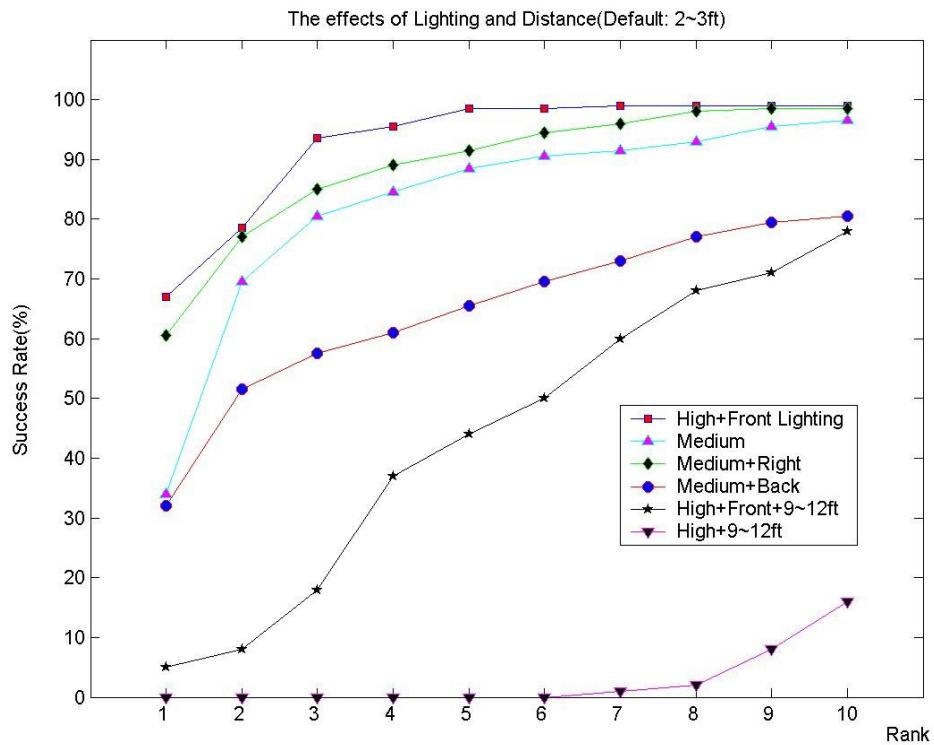


Figure 32: The effects of lighting and distance

Based on the identification experiment, lighting, expression and other variations seem to not much affect the performance independently, we can predict that mixed-variations seem to significantly affect the performance as shown in the surveillance experiment. Overall performance which has a 1st matching accuracy rate of 88% has significantly dropped to 60~70 %. There is a large discrepancy in the performance of visual based face recognition when we use existing database and images taken in the real world. Without solving the problem of affect of variation around the face, the future of face recognition technology will be unpredictable.

4.3 Decision Fusion

Decision fusion can be accomplished by ranked-list combination [39][41]. From each ranked-list of visual and thermal recognition results, new ranked lists can be produced. Instead of seeking ranked-lists, we used confidence rates (similarity scores) that determine ranked-lists. Confidence rates tell how the two images (gallery, probe) are similar. Current face recognition algorithms produce a numeric similarity score between an unknown face and each image in the gallery. The best similarity score between the probe and gallery images then determines the identity. The relationship between gallery and probe images is important and contributes to overall face recognition performance. Although FaceIt® software is not developed for thermal images; the decision of similarity between gallery and probe images is the same. That is why we can apply thermal images for the evaluation of visual and thermal images for face recognition.

FaceIt® has a maximum confidence score of 10.0. This maximum confidence can only appear if both images (gallery and probe) are exactly the same. In real applications

where mixed variations over faces may occur, we do not expect to achieve maximum confidence. We lose some confidence if variations occur. The more variation occurs between gallery and probe images, the more score will be lost. This is one of the reasons why controlling the environment where faces are captured is important. We saved each visual and thermal recognition scores in order to test fusion methods based on average, and higher confidence rates in terms of

$$Score_F = (W_V * Score_V + W_T * Score_T) \quad (8)$$

$$Score_F = Score_V \quad (Score_V > Score_T), \quad Score_F = Score_T \quad (Score_V < Score_T) \quad (9)$$

where the new fused score, $Score_F$, can be derived by utilizing the scores of thermal and visual recognition results, $Score_T$, $Score_V$, respectively. W_V , and W_T can be denoted as weighting factors from both visual and thermal face recognition experts. These weighting factors can be applied based on the overall performance of visual and thermal face recognition. These weighting factors are considered as equal after comparison of the performance of visual and thermal face recognition which we will discuss later in section 4.4.

Figure 33 shows an example of average confidence based fusion methods. The first row shows input images of visual, thermal, and expected output, which are in the gallery. The second and third rows show the ranked-based visual and thermal recognition results. Finally the fourth row shows the final results using average or higher confidence rate of visual and thermal fused recognition results. 7th rank in the visual recognition performance and 1st rank in the thermal recognition performance yield 1st rank in the average confidence based fused process.

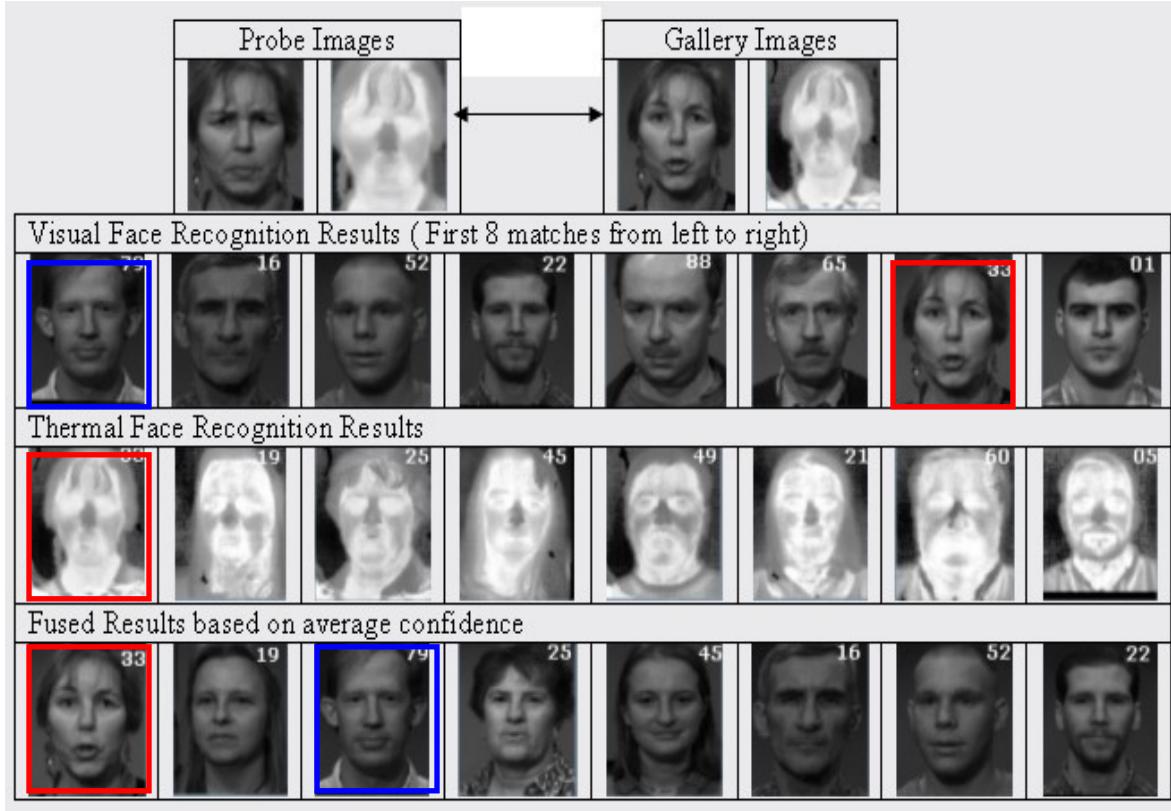


Figure 33: Decision fusion example

4.4 Visual and Thermal Comparison Results

In an effort to verify that thermal face recognition is an alternative for visual face recognition, we compared the performance of visual and thermal images. Automatic eye localization via FaceIt® succeeded in 95% of the images and the remaining 5% of the images from Table 1 had to be aligned manually to achieve the best recognition rates for visual images. In fully automatic verification applications, the inexact eye location affects the overall performance of face recognition system when the operator cannot assist.

Thermal image alignment was done in two ways and the evaluation conducted separately for each case; (1) manually with the user clicking on the eye locations and then, (2) automatically by using the eye coordinates found in the registered visual images.

Figure 34 shows the comparison of results for visual and thermal face recognition. In cases where no eyeglasses were worn (probe 1, 2, 3), thermal face recognition achieved higher performance ranks and confidence rates than visual face recognition under different lighting conditions and expressions. Thermal images also easily converged within 10 ranks while visual images do not in most cases. Manual eye clicking in the thermal images yields slightly better recognition result than using eye locations from the visual images.

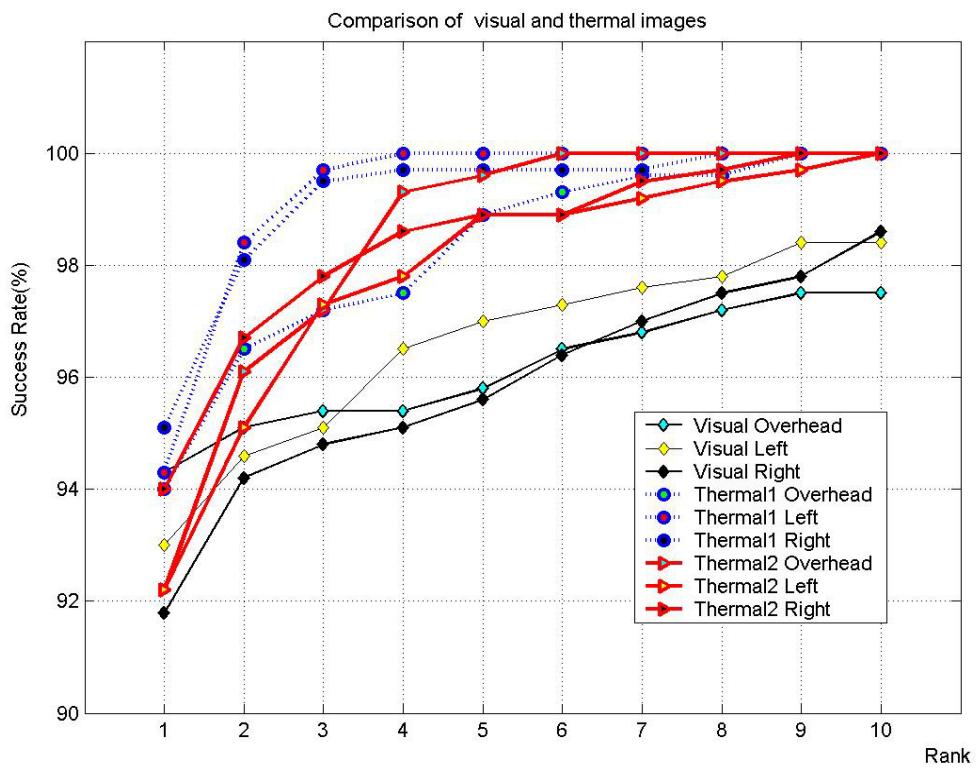


Figure 34: Performance evaluation when individuals are wearing no eyeglasses; Thermal1- manual eye positioning, Thermal2- same eye coordinates from the visual images

Thermal images of individuals wearing eyeglasses result in poor performance since the eyeglasses block the infrared emissions while visual images are only slightly affected

in performance in this case. This is because FaceIt®, which is developed for visual images, is more weighted to consider the eyes resulting in a loss of a small score in visual images but losing significant score in the thermal images. FaceIt® will yield better results if it provides less score around the eyes when eyeglasses are detected.

4.5 Fusion of Visual and Thermal Face Recognition

We compared visual and thermal recognition results with different fusion methods described in section 2.3 and section 4.3. Instead of dividing different probes, we combine all probes (1, 2, and 3) with a total of 1039 images. Eye locations from the visual images commonly was used for thermal and fused (data fusion) images. Average confidence between gallery and probe images of same person was also calculated. It is reasonable to consider the weights of each classifier (expert) to fuse results properly. For example, if we want to weigh each classifier differently based on the first 10 matches success rates, we need to apply more weight on thermal face recognition than visual face recognition. Here, we apply the weights based on the performance of 1st match success rates. The performance of visual and thermal face recognition are almost similar; therefore, we assume the weights of visual and thermal face recognition are equal.

Figure 35 shows the summary of performance comparison. All fusion methods yield better performance over a single modality, higher confidence-based decision fusion give us especially satisfactory results with the highest average confidence rate. Although average confidence based decision fusion gives the highest performance rate in this experiment, it is too early to say average confidence-based decision fusion is the best solution for the fusion of visual and thermal face recognition.

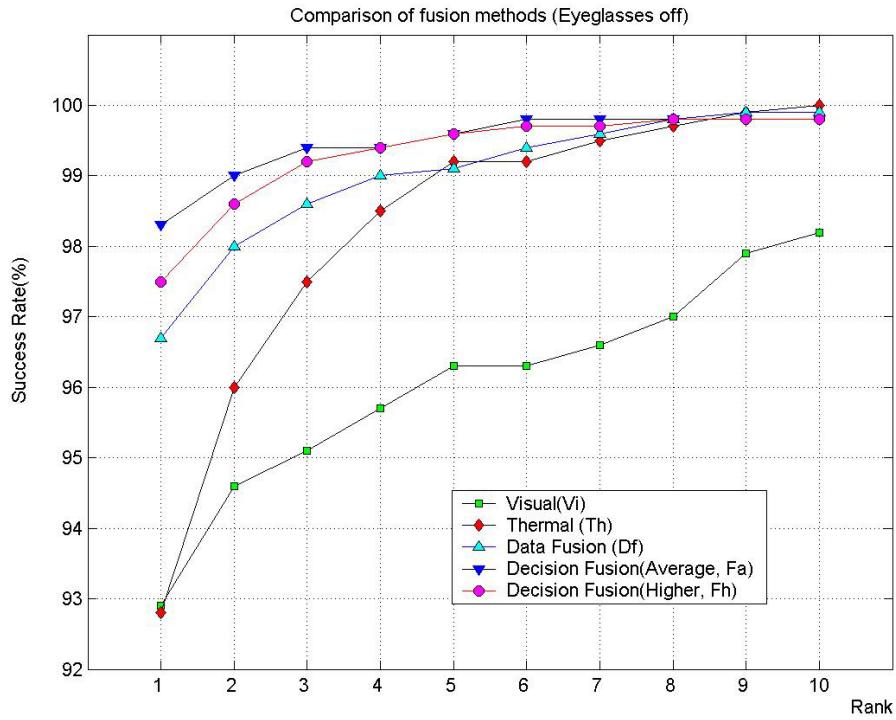


Figure 35: Fused performance evaluation with eyeglasses off (Probe 1, 2, 3)

Figure 36 and Figure 37 show the detail comparison results of visual, thermal, data fusion (average), and decision fusion (average, higher confidence rate) when individuals are wearing eyeglasses and when eyeglasses are replaced with eye templates respectively. The total images used in this experiment were 514 (Probe 4, 5, 6). Eyeglasses slightly affect the performance of visual face recognition system while significantly affecting the performance of thermal face recognition. In this case, the fused images also result in poor performance. On the other hand, decision fusion gives us almost similar performance to visual face recognition. This is because the confidence of visual images is usually higher than that of thermal images when individuals are wearing eyeglasses; therefore, visual images play an important role in decision fusion. After eyeglass removal and replacement, the performance of thermal face recognition increased dramatically which resulted in improved performance of both the data and decision fusion schemes.

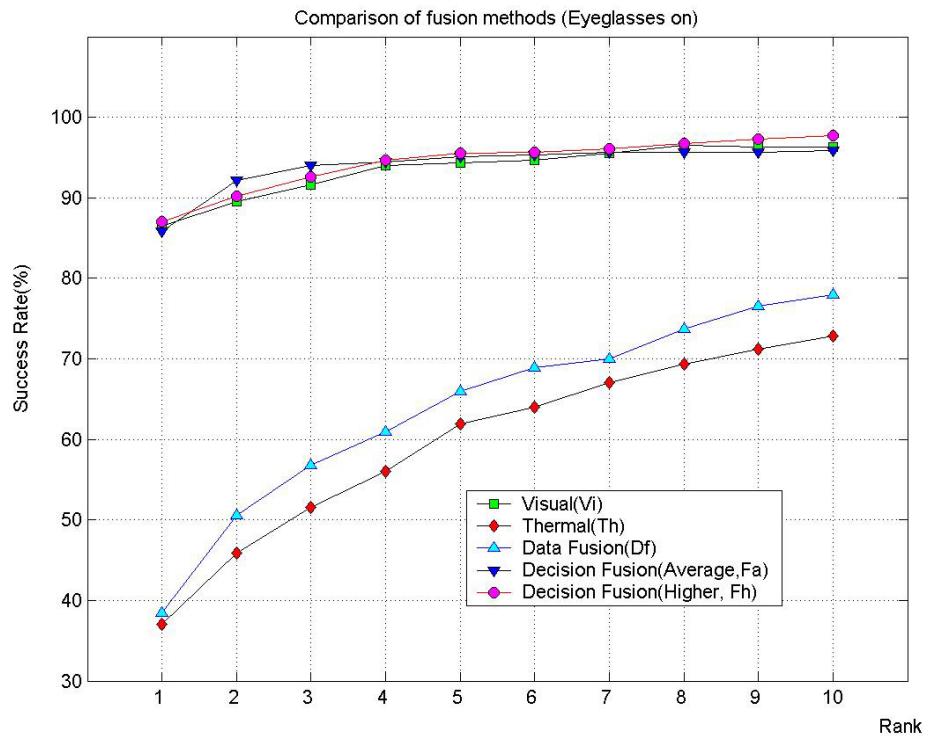


Figure 36: Performance comparison when eyeglasses are present (Probe 4, 5, 6)

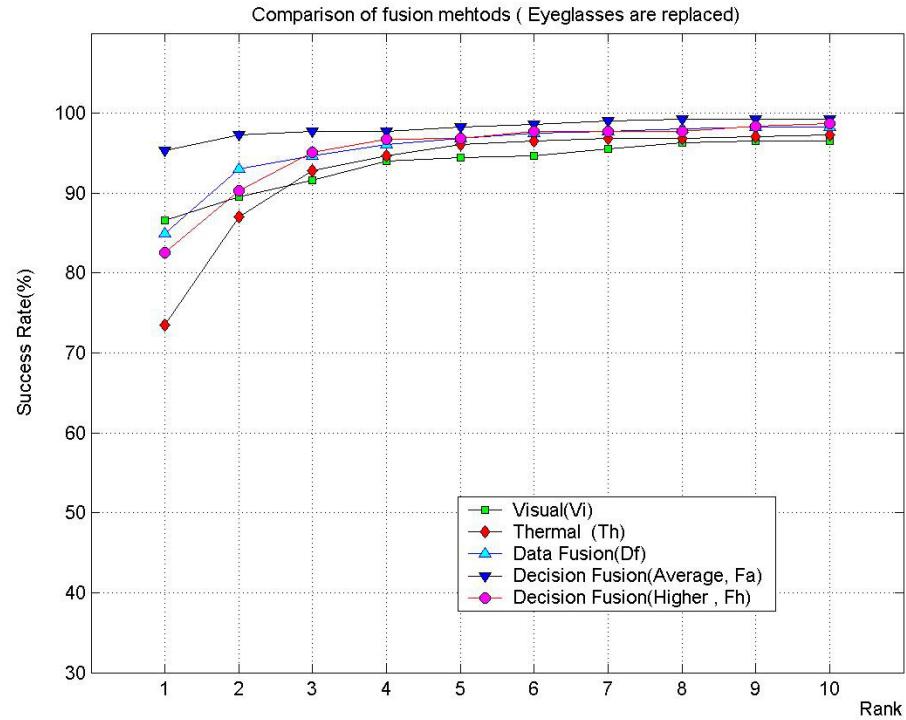


Figure 37: Performance comparison when eyeglasses are replaced

From Figures 35 to 37, we compared our experiment only with cumulative success matching rates. We also compared the 1st match success rate from the previous experimental results as shown in Figure 38. 1st match is important for the evaluation of the verification test which can be a function of 1st match and thresholds (confidence rates). Therefore, we added a comparison of average confidence rates of visual, thermal and fusion methods as shown in Figure 39. From the results of Figure 38 and 39, we can predict that the similarity between two images of the same individual can be degraded by eyeglasses in thermal images, but can be upgraded by replacement of the eyeglasses. Individuals who are not wearing eyeglasses give us the best confidence rates.

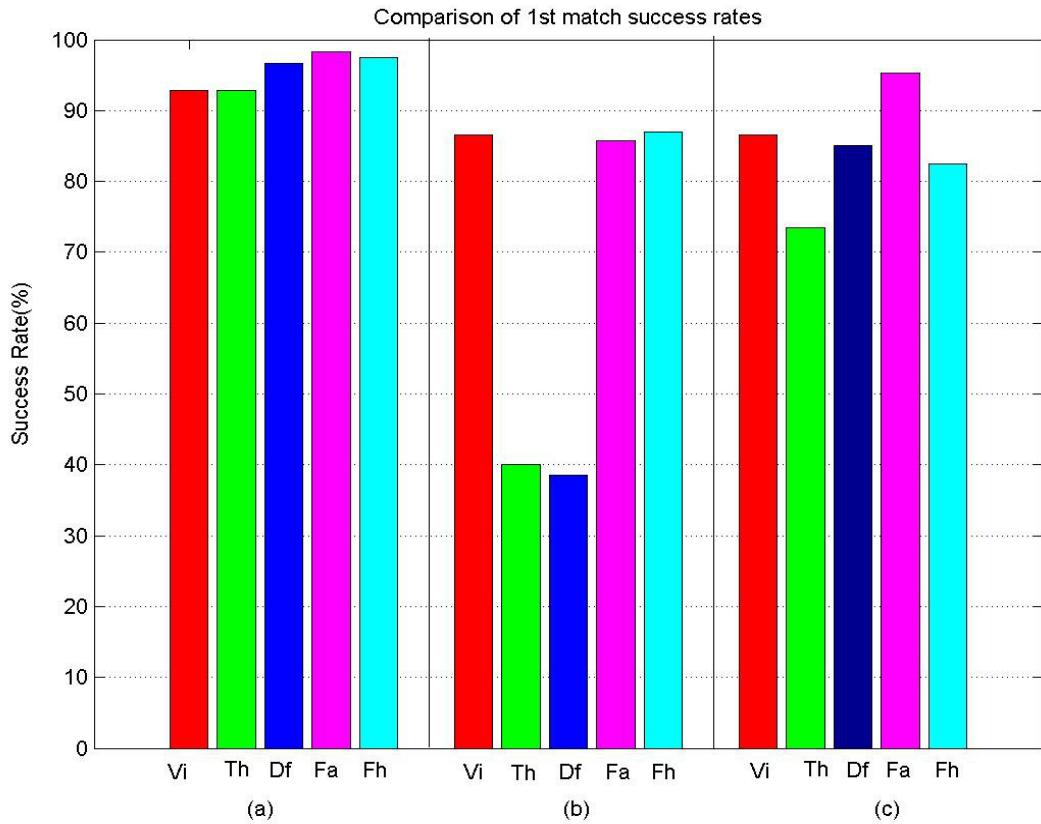


Figure 38: Performance comparison of 1st match success rates; (a) No eyeglasses, (b) eyeglasses, and (c) eyeglasses removed

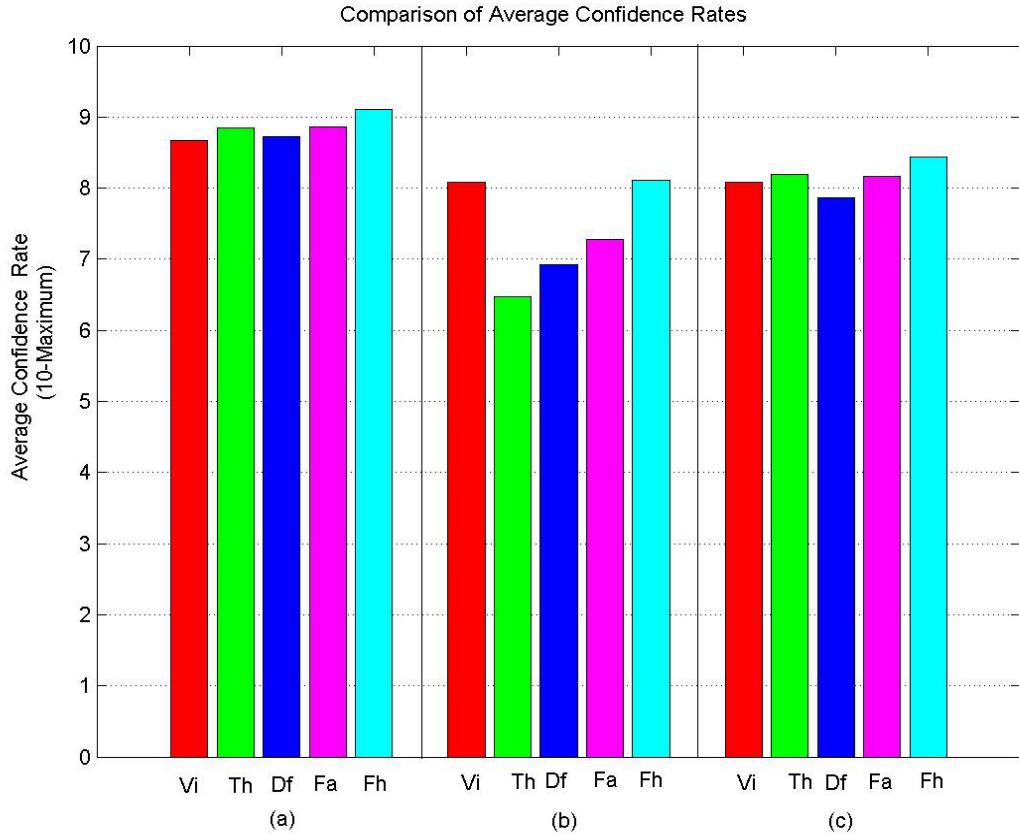


Figure 39: Performance comparison of average confidence rates; (a) No eyeglasses, (b) eyeglasses, and (c) eyeglasses removed

5. Conclusion

We reviewed basic algorithms of information fusion in order to compare visual and thermal images with different fusion methods. Since fusion of information is a relatively new research area, it is necessary to show, using the basic fusion algorithms, that fused methods lead to better performance over a single modality before discussing complex fusion algorithms, especially feature level fusion. With taking advantages of each visual and thermal image, the new fused system can be implemented in collaborating low-level data fusion and high-level decision fusion. In order to find which variations can affect the performance of a visual face recognition system, we compared the performance of face

recognition algorithms. Without evaluation of the face recognition algorithms, the best features for face recognition cannot be simply determined. That is why the best feature sets for face recognition are still arguable. Extensive evaluation can help to achieve a more reliable set of face features and help to increase the performance of the overall system. FaceIt® is considered as an expert application for both visual and thermal face recognition after comparison of its performance using visual and thermal face images.

We demonstrated that thermal face recognition achieved higher performance ranks and average confidence rates than visual face recognition under different lighting conditions and expressions in cases where no eyeglasses were worn [59]. Thermal images also easily converged within 10 ranks while visual images do not in most cases. Thermal images of individuals wearing eyeglasses result in poor performance since the eyeglasses block the infrared emissions. An automatic eyeglass detection scheme is proposed by the analysis of blobs and ellipse fitting schemes. Since here we considered only 2 blobs, the extension of the number of blobs can be helpful in finding the facial features in the thermal images. If robust facial features are found without the assistance of a co-registered camera, those can be used for normalization of the thermal faces which is a challenging problem in pattern recognition technology. Only thermal-based face recognition systems should be developed if we want to recognize people under dark illumination conditions where eye detection and face detection in the visual images fail. After detection of eyeglasses, the eyeglass regions can be replaced with average eye templates for better performance.

Decision fusion methods, if eyeglasses are present or not, give us satisfactory results. The performance of decision fusion also increased after replacing eyeglasses with an

average eye template in thermal images. The decision fusion stage can also prevent unexpected variations since the similarity may not be enough to claim an identity. These variations can include eyeglasses in thermal images, extreme facial changes, and lighting changes in visual images, and pose changes in both visual and thermal images. Decision fusion can be extended if we increase the number of classifiers or sensors, and will definitely provide a more robust system. Hopefully, the algorithms and comparisons presented in this paper do provide a step toward impacting on increasing the performance of face recognition systems.

References

- [1] J. L. Center, Jr., “Practical Application of Facial Recognition: Automated Facial Recognition Access Control System,” *Face Recognition: From Theory to Applications*, pp.402-408, 1998.
- [2] P. J. Phillips and P. Rauss, “The Face Recognition Technology (FERET) Program,” *Proc. Office of National Drug Control Policy, CTAC Int. Technology Symposium*, pp.8-11, 1997.
- [3] Y. Adini, Y. Moses, and S. Ullman, “Face Recognition: The Problem of Compensating for Changes in Illumination Direction,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp.721-732, 1997.
- [4] D. M. Blackburn, J. M. Bone, and P. J. Phillips, “Face Recognition Vendor Test 2000,” *Evaluation Report, National Institute of Standards and Technology*, pp.1-70, 2001.

- [5] M. Bone and D. Blackburn, "Face Recognition at a Chokepoint: Scenario Evaluation Results," *Evaluation Report, Department of Defense*, 2002.
- [6] P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone, "Face Recognition Vendor Test 2002," *Evaluation Report, National Institute of Standards and Technology*, pp.1-56, 2003.
- [7] R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 15, No. 10, pp.1042-1052, 1993.
- [8] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and Machine Recognition of Faces: A Survey," *Proceedings of the IEEE*, Vol. 83, No. 5, pp.705-740, 1995.
- [9] T. Kanade, "Picture Processing by Computer Complex and Recognition of Human Faces," *Technical Report, Kyoto University*, 1973.
- [10] I. J. Cox, J. Ghosn, and P. N. Yianilos, "Feature-Based Face Recognition Using Mixture-Distance," *Proc. Int. Conf. Computer Vision and Pattern Recognition*, pp.209-216, 1996.
- [11] B. S. Manjunath, R. Chellappa, and C. von der Malsburg, "A Feature Based Approach to Face Recognition," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition*, pp.373-378, 1992.
- [12] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 23, No. 2, pp.228-233, 2001.
- [13] M.Turk and Alex Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, Vol. 3, pp72-86, 1991.

- [14] P. S. Penev and J. J. Atick, “Local Feature Analysis: A General Statistical Theory for Object Representation,” *Network: Computation in Neural Systems*, Vol. 7, No. 3, pp.477-500, 1996.
- [15] P. S. Penev, “Local Feature Analysis: A Statistical Theory for Information Representation and Transmission,” *Ph.D. Thesis, The Rockefeller University*, 1998.
- [16] P. S. Penev, “Dimensionality Reduction by Sparsification in a Local-features Representation of Human Faces,” *Technical Report, The Rockefeller University*, 1999.
- [17] R. Cappelli, D. Maio, and D. Maltoni, “Multispace KL for Pattern Representation and Classification,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 23, No. 9, pp.977-996, 2001.
- [18] R. A. Fisher, “The Statistical Utilization of Multiple Measurements,” *Annals of Eugenics*, Vol. 8, pp.376-386, 1938.
- [19] R. Lotlikar and R. Kothari, “Fractional-step Dimensionality Reduction,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, No. 6, pp.623-627, 2000.
- [20] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp.711-720, 1997.
- [21] C. Liu and H. Wechsler, “A Shape- and Texture-based Enhanced Fisher Classifier for Face Recognition,” *IEEE Trans. Image Processing*, Vol. 10, No. 4, pp.598-608, 2001.

- [22] D. L. Swets and J. J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 18, No. 8, pp.831-836, 1996.
- [23] K. Etemad and R. Chellappa, "Discriminant Analysis for Recognition of Human Face Images," *Journal of Optical Society of America*, Vol. 14, No. 8, pp. 1724-1733, 1997.
- [24] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, Vol. 36, No. 3, pp.287-314, 1994.
- [25] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Trans. Neural Networks*, Vol. 13, No. 6, pp.1450-1464, 2002.
- [26] L. Wiskott, J. M. Fellous, N. Krüger, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp.775-779, 1997.
- [27] S. H. Lin, S. Y. Kung, and L. J. Lin, "Face Recognition/Detection by Probabilistic Decision-Based Neural Network," *IEEE Trans. Neural Networks*, Vol. 8, No. 1, pp.114-132, 1997.
- [28] V. N. Vapnik, *The Nature of Statistical Learning Theory*, New York: Springer-Verlag, 1995.
- [29] B. Schölkopf, *Support Vector Learning*, Munich, Germany: Oldenbourg-Verlag, 1997.

- [30] P. J. Phillips, "Support Vector Machines Applied to Face Recognition," *Advances in Neural Information Processing Systems 11*, M. S. Kearns, S. A. Solla, and D. A. Cohn, eds., 1998.
- [31] Y. Gao, and M. K. H. Leung, "Face Recognition using Line Edge Map," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 24, No. 6, pp.764-779, 2002.
- [32] S. Z. Li and J. Lu "Face recognition using the nearest feature line method," *IEEE Trans. Neural Networks*, Vol. 10, No. 2, pp.439-443, 1999.
- [33] Y. Adini, Y. Moses, and S. Ullman, "Face Recognition: The Problem of Compensating for Changes in Illumination Direction," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp.721-732, 1997.
- [34] Wolff, Socolinksy and Eveland, "Face Recognition in the Thermal Infrared", *Technical Report*, Equinox, 2002
- [35] Socolinsky, Wolff, Neuheisel and Eveland, "Illumination Invariant Face Recognition Using Thermal Infrared Imagery", *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Kauai, Hawaii, Dec 2001
- [36] A. Selinger and D. A. Socolinsky, "Appearance-Based Facial Recognition Using Visible and Thermal Imagery: A Comparative Study," *Technical Report*, Equinox Corporation, 2001.
- [37] F. Prokoski, "History, Current Status, and Future of Infrared Identification," *Proc. IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, Hilton Head, pp.5-14, 2000.

- [38] Y. Yoshitomi, T. Miyaura, S. Tomita, and S. Kimura, “Face identification using thermal image processing,” *Proc. IEEE Int. Workshop on Robot and Human Communication*, pp.374-379, 1997.
- [39] C. Sanderson, and K.K. Paliwal, “Information Fusion and Person Verification Using Speech & Face Information, IDIAP-RR 02-33, Martigny, Switzerland, 2002.
- [40] D. L. Hall and J. Llinas, “Multisensor Data Fusion,” *Handbook of Multisensor Data Fusion*, D. L. Hall and J. Llinas, ed., CRC Press, pp.1-10, 2001
- [41] Dasarthy, “Decision Fusion,” *IEEE Computer Society Press*, 1994
- [42] S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz, “Fusion of face and speech data for person identity verification,” *IEEE Trans. Neural Networks*, Vol. 10, No. 5, pp.1065-1074, 1999
- [43] S. Gutta, J. R. J. Huang, P. Jonathon, and H. Wechsler, “Mixture of experts for classification of gender, ethnic origin, and pose of human faces,” *IEEE Trans. Neural Networks*, Vol. 11, No. 4, pp. 948-960, 2000
- [44] Y. Fang, T. Tan, and Y. Wang, “Fusion of Global and Local Features for Face Verification,” *Proc. of Int. Conf. on Pattern Recognition*, 2002
- [45] K. Chang, K.W. Bowyer, S. Sarkar, B.Victor,”Comparison and Combination of Ear and Face Image in Appearance-Based Biometrics”, *IEEE Transactions on Pattern Analysis and machine Intelligence*, Vol. 25, No. 9, pp. 1160-1165, Sep 2003.
- [46] J. Wilder, P. J. Phillips, C. Jiang, and S. Wiener, “Comparison of Visible and Infrared Imagery for Face Recognition,” *Proc. Int. Conf. Automatic Face and Gesture Recognition*, pp.182-187, 199.

- [47] <http://www.equinoxsensors.com/products/HID.html>
- [48] J. Huang, C.Liu, and H. Wechsler, “Eye Detection and Face Recognition using Evolutionary Computation”, *Proceedings of NATO-ASI on Face Recognition: From Theory to Application*, 1998.
- [49] K.M. Lam and H. Yan, "An Analytical-to-holistic Approach for Face Recognition Based on a Single Frontal View, *IEEE Transactions on Pattern Analysis and machine Intelligence*, Vol. 20, No. 7, pp. 673-86, July 1998.
- [50] R. Mariani, “Subpixellic Eye Detection”, *Proceedings International conference on image analysis and processing*, 1999.
- [51] J. Huang, C.Liu, and H. Wechsler, “Eye Detection and Face Recognition using Evolutionary Computation”, Proceedings of NATO-ASI on Face Recognition: From Theory to Application, 1998.
- [52] J. Huang, and H. Wechsler, “Eye Detection and Optimal Wavelet Packets and Radial Basis Function(RBFs)”, *International Journal of Pattern Recognition and Artificial Intelligence*, Vol 13, No 7 , 1999.
- [53] F.L. Bookstein. Fitting conic sections to scattered data. *Computer Graphics and Image Processing*, (9):56--71, 1979.
- [54] Andrew W. Fitzgibbon, R.B.Fisher. A Buyer’s Guide to Conic Fitting. Proc.5th British Machine Vision Conference, Birmingham, pp. 513-522, 1995.
- [55] E. Hjelmas and H. K. Low, “Face Detection: A Survey,” *Computer Vision and Image Understanding*, Vol. 83, No. 3, pp.236-274, 2001.

- [56] M. H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting Faces in Images: A Survey," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 24, No. 1, pp.34-58, 2002.
- [57] J. Heo, B. Abidi, J. Paik, and M. A. Abidi, "Face Recognition: Evaluation Report For FaceIt®," *Proc. 6th International Conference on Quality Control by Artificial Vision QCAV03*, Gatlinburg, TN, USA, May 2003.
- [58] Y Kim, J. Paik, J. Heo, A. Koschan, B. Abidi, and M. Abidi, "Automatic face region tracking for highly accurate face recognition in unconstrained," *IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS2003*, Miami, FL, July 2003
- [59] J. Heo, B. Abidi, S. Kong, and M. Abidi, "Performance Comparison of Visual and Thermal Signatures for Face Recognition," *Biometric Consortium*, Arlington, VA, Sep 2003.

Vita

Jingu Heo was born in Daegu, Korea, on June 11, 1971, the son of Manyoung Heo and Malzo Han. After graduating in 1990 from Shimin High School, Daegu, Korea, he attended Kyungbook National University in Daegu where he received a Bachelor's degree in Electrical and Electronic Engineering. During his undergraduate studies, he served for the military for two and a half years. In the spring of 1997, he was employed by Samsung Electronics as a system engineer and worked for three and a half years. In 2001, he returned to the academic world as a master's student at the University of Tennessee in Electrical Engineering. He joined the Imaging, Robotics, and Intelligent Systems Laboratory as a graduate research assistant where he completed his Master's degree in 2003.