

Chapter 6

Face Recognition in the Thermal Infrared *

Lawrence B. Wolff, Diego A. Socolinsky, and Christopher K. Eveland

Equinox Corporation, 9 West 57th Street, New York, New York 10019
{wolff,diego,eveland}@equinoxsensors.com

Summary. Recent research has demonstrated distinct advantages of using thermal infrared imaging for improving face recognition performance. While conventional video cameras sense reflected light, thermal infrared cameras primarily measure emitted radiation from objects such as faces. Visible and thermal infrared image data collections of frontal faces have been on-going at NIST for over two years, producing the most comprehensive face database known to involve thermal infrared imagery. Rigorous experimentation with this database has revealed consistently superior recognition performance of algorithms when applied to thermal infrared, particularly under variable illumination conditions. Physical phenomenology responsible for this observation is analyzed. An end-to-end face recognition system incorporating simultaneous coregistered thermal infrared and visible has been developed and tested indoors with good performance.

6.1 Introduction

Accelerated developments in camera technology over the last decade have given computer vision researchers a whole new diversity of imaging options, particularly in the infrared spectrum. Conventional video cameras use photosensitive silicon that is typically able to measure energy at electromagnetic wavelengths from $0.4\ \mu\text{m}$ to just over $1.0\ \mu\text{m}$. Multiple technologies are currently available, with dwindling cost and increasing performance, which are capable of image measurement in different regions of the infrared spectrum, as shown in Figure 6.1. Figure 6.2 shows the different appearances of a human face in the visible, shortwave infrared (SWIR) midwave infrared (MWIR), and longwave infrared (LWIR) spectra. Although in the infrared, the near-infrared (NIR) and SWIR spectra are still reflective and differences in appearance between the visible, NIR and SWIR are due to reflective material properties. Both NIR and SWIR have been found to have advantages over imaging in the visible for face detection [1] and detecting disguise [2].

* This research was supported by the DARPA Human Identification at a Distance (HID) program under contract #DARPA/AFOSR F49620-01-C-0008.

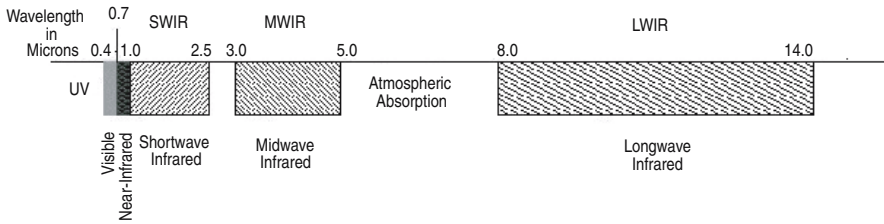


Figure 6.1. Nomenclature for various parts of the electromagnetic spectrum.



Figure 6.2. A face simultaneously imaged in the (a) visible spectrum, $0.4\text{--}0.7\ \mu\text{m}$, (b) shortwave infrared, $0.9\text{--}1.7\ \mu\text{m}$, (c) midwave infrared, $3.0\text{--}5.0\ \mu\text{m}$, and (d) longwave infrared, $8.0\text{--}14.0\ \mu\text{m}$.

At wavelengths of $3\ \mu\text{m}$ and longer imaged radiation from objects becomes significantly emissive due to temperature, and is hence generally termed the thermal infrared. The thermal infrared spectrum is divided into two primary spectra, the MWIR and LWIR. Between these spectra lies a strong atmospheric absorption band between approximately 5 and $8\ \mu\text{m}$ wavelength, where imaging becomes extremely difficult due to nearly complete opaqueness of air. The range beyond $14\ \mu\text{m}$ is termed the very longwave infrared (VLWIR) and although in recent years it has received increased attention, it remains beyond the scope of this chapter. The amount of emitted radiation depends on both the temperature and the *emissivity* of the material. Emissivity in the thermal infrared is conversely analogous to the notion of reflective albedo used in the computer vision literature [3, 4]. For instance, a Lambertian reflector can appear white or grey depending on its efficiency for reflecting light energy. The more efficient it is in reflecting energy (more reflectance albedo) the less efficient it is in thermally emitting energy respective to its temperature (less emissivity). Objects with perfect emissivity of 1.0 are completely black. Many materials that are poor absorbers transmit most light energy while reflecting only a small portion. This applies to a variety of different types of glass and plastics in the visible spectrum.

As detailed in the following section, the spectral distribution of energy emitted by an object is simply the product of the Planck distribution for a given temperature, with the emissivity of the object as function of wavelength [5]. In the vicinity of human body temperature (37°C), the Planck

distribution has a maximum in the LWIR around $9\text{ }\mu\text{m}$, and is approximately one-sixth of this maximum in the MWIR. As we will show through empirical measurement, the emissivity of human skin in the MWIR is at least 0.91, and at least 0.97 in the LWIR. Therefore, face recognition in the thermal infrared favors the LWIR, since LWIR emission is much higher than that in the MWIR. Thermal infrared imaging for face recognition first used MWIR platinum silicide detectors in the early 1990s [6]. At that time, cooled LWIR technology was very expensive. By the late 1990s, uncooled microbolometer imaging technology in the LWIR became more accessible and affordable, enabling wider experimental applications in this regime. At that time, cooled MWIR technology was about ten times more sensitive than uncooled microbolometer LWIR technology, and even though faces are more emissive in the LWIR, in the late 1990s MWIR could still discern more image detail of the human face. At present, uncooled microbolometer LWIR technology coming off the assembly lines is rapidly approaching one-half of the sensitivity of cooled MWIR. For face recognition in the thermal infrared, this is a turning point as for the first time the most appropriate thermal infrared imaging technology (i.e. LWIR) for studying human faces is also the most affordable.

For over two years, data collections of both visible and thermal infrared imagery of faces have been taken and continue to take place at regular intervals of 6 months at the National Institute of Science and Technology (NIST). This effort is supported by the DARPA HID program [7]. Section 6.2 describes the comprehensive database resulting from these collections, consisting of over 100,000 images of over 300 individuals so far. This database has provided the empirical foundation with which to rigorously compare the performance of various face recognition algorithms between visible and thermal infrared imagery. Some of these results are summarized in Section 6.6. Also described in Section 6.6 is a recently completed full-working prototype of the equinox access control environment (ACE) face recognition system, which uses fused coregistered visible and LWIR imagery from a novel sensor system.

The main advantage of thermal infrared imaging for boosting face recognition performance is its apparent invariance to changing illumination. Section 6.4 attempts to characterize how well thermal infrared images of the human face are invariant to illumination changes. Section 6.5 delves deeper into explaining the physical phenomenology responsible for this invariance by computing the emissivity of human skin in the MWIR and the LWIR. This culminates in a preliminary thermal model for human skin. Finally, Section 6.7 briefly overviews some of the remaining challenges that thermal infrared imaging does not immediately remedy for face recognition.

6.2 The Equinox Visible/Infrared Face Database

Figure 6.3 shows the experimental set-up for imagery being collected on a regular basis at NIST, simultaneously in the visible, SWIR, MWIR, and LWIR spectra. The objectives for these ongoing data collections are as follows:

1. To be able to directly and rigorously compare the performance of face recognition algorithms between visible imagery and imagery in the various modalities of the infrared spectrum.
2. To produce face imagery simultaneously in these modalities under variable illumination conditions.
3. To produce face imagery with significant intrapersonal variation for each imaging condition and modality.

Towards the first objective a configuration of four different sensors of respective modalities and interface software has been set-up for simultaneous acquisition of visible, SWIR, MWIR, and LWIR imagery. All infrared cameras are of the Indigo Merlin Series with 320×240 resolution. The visible camera is a Pulnix 6710 with 640×480 resolution. A special optical design insures precise pixel coregistration of visible and LWIR imaging modalities [8, 9]. Although the SWIR and MWIR imaging modalities are boresighted², impairing precise coregistration at close distances, the physical separation between these cameras has been minimized beyond what is shown in Figure 6.3 so that views are nearly identical. Optically coregistered sensors in the NIR were in use [2, 1] for face detection. The complexity of coregistering visible and LWIR wavelengths, however, is much greater due to the larger disparity between them.

Towards the second and third objectives, collection of image data was repeated for three different illumination conditions: (i) Frontal, (ii) frontal-Left, (iii) frontal-Right, with lamps shown in Figure 6.3 using standard 3200 K color temperature photographic bulbs. Figure 6.4 shows the emission curve for these bulbs in the wavelengths of interest. Forty image frame sequences of visible, SWIR, MWIR, and LWIR were digitized simultaneously at 10 frames/second (i.e., 4-seconds duration), while a human subject was reciting the vowels “a,” “e,” “i,” “o,” “u.” This creates a continuous image sequence with changes in expression throughout providing significant intrapersonal variation over the course of multiple frames. At the same time there is little facial movement between consecutive image frames 1/10 second apart, allowing for analysis of image variations due to temporal sensor noise. Figure 6.2 is an example of one such multimodal frame within this 40-frame sequence. After the acquisition of each 40-frame contiguous image sequence, for each illumination three more static images are taken of individuals told to make extreme expressions of “smile,” “frown,” and “surprise.”

² The term *boresighted* typically refers to cameras that have been placed alongside each other and aimed in the same direction. Due to the separation between the cameras, it is impossible to obtain the same view of a 3D object from both sensors. This is often exploited in stereo vision to compute depth maps.

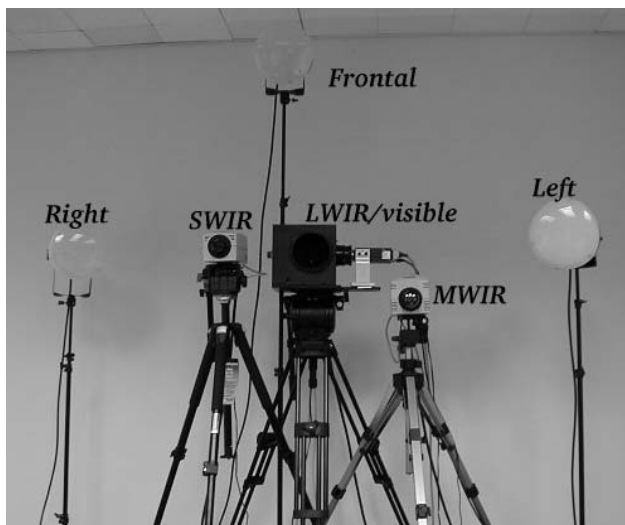


Figure 6.3. Camera and illumination equipment set-up used for simultaneous data collection of visible, SWIR, MWIR, and LWIR imagery.

Prior to data collection, the radiometric calibration procedure described in Section 6.3 was performed for the Indigo Merlin series MWIR and LWIR cameras using a Model 350 Mikron blackbody source. Software was developed to convert raw MWIR and LWIR image grayscale values directly into respective thermal emission values from ground-truth blackbody images. Raw image gray values for the MWIR and LWIR cameras are 12-bit integers from which floating point thermal emission values were computed and then rounded back to 12-bit values with appropriate dynamic range.

At present, the Equinox visible/infrared database consists of over 300 individuals imaged over five separate data collections at NIST. At least 60 individuals have participated in two or more of these data collections so that intrapersonal variations over 6 months or more can be analyzed. To summarize, for each individual a 40-frame sequence plus three static images were taken for three different illuminations and four spectral image modalities. Not including duplicate individuals, at least $340 \text{ individuals} * 43 \text{ images/modality} * 3 \text{ illuminations/individual} * 4 \text{ modalities} = 175,440$ images are contained in the Equinox database. Almost all of this database was collected indoors, with outdoor imagery beginning to be collected during the last data collection at NIST in April 2002. A portion of this database is available on the Internet at <http://www.equinoxsensors.com/hid>.

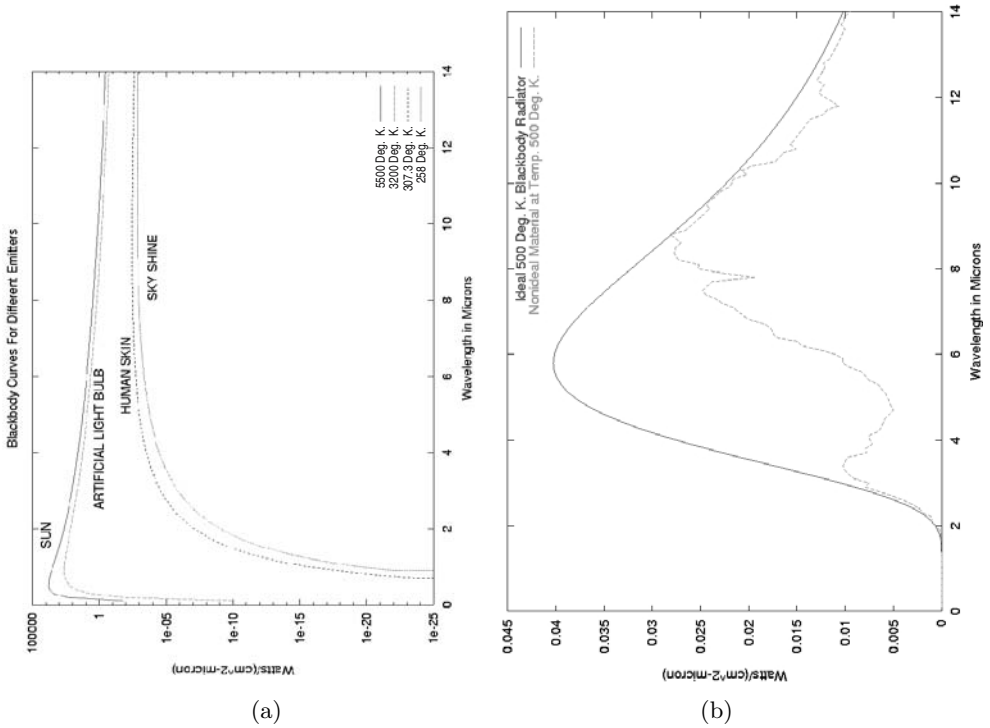


Figure 6.4. (a) Blackbody Planck curves comparing thermal IR emission from common natural and artificial illumination sources to thermal IR emission from human skin. (b) Comparison of an ideal blackbody Planck curve with a nonideal emitter at the same temperature.

6.3 Calibration Of Thermal IR Sensors

All objects above absolute zero temperature emit electromagnetic radiation. In the early 1900s Planck was the first to characterize the spectral distribution of this radiation for a *blackbody*, which is an object that completely absorbs electromagnetic radiation at all wavelengths [5]. According to Planck's law, the spectral distribution of emission from a blackbody at temperature T , is given by

$$W(\lambda, T) = \frac{2\pi hc^2}{\lambda^5 (e^{\frac{hc}{\lambda kT}} - 1.0)} \quad [\text{Watts/cm}^2] \mu\text{m}^{-1}, \quad (6.1)$$

$$Q(\lambda, T) = \frac{2\pi c}{\lambda^4 (e^{\frac{hc}{\lambda kT}} - 1.0)} \quad [\text{Photons/cm}^2 - \text{sec}] \mu\text{m}^{-1}, \quad (6.2)$$

expressed in two different units of energy flux which are commonly used. In the above formulas, h is Planck's constant, k is Boltzmann's constant, c is the speed of light, and λ is wavelength. Figure 6.4(a) shows a comparison of

blackbody spectral distributions corresponding to the various temperatures for the Sun, artificial lightbulb illumination at 3200 K color temperature [10], human skin, and average temperature for the atmosphere.

In reality, only very few objects are near perfect energy absorbers, particularly at all wavelengths. The proportional amount of energy emission with respect to a perfect absorber is called the *emissivity* $\epsilon(T, \lambda, \psi)$, which takes values in the range $[0, 1]$. In addition to temperature T and wavelength λ , this can also be a function of emission angle ψ . Kirchoff's law states that the emissivity at a point on an object is equal to the *absorption* $\alpha(T, \lambda, \psi)$, namely:

$$\epsilon(T, \lambda, \psi) = \alpha(T, \lambda, \psi) .$$

This is a fundamental law that effectively asserts the conservation of energy. Blackbody objects are therefore the most efficient radiators, and for a given temperature T emit the most energy possible at any given wavelength. Expressed in the same units as equations (6.1) and (6.2) above, the spectral distribution of emission from an object with emissivity $\epsilon(T, \lambda, \psi)$, is given by:

$$\epsilon(T, \lambda, \psi) \times W(\lambda, T), \quad \epsilon(T, \lambda, \psi) \times Q(\lambda, T) .$$

For illustrative purposes, Figure 6.4(b) compares the spectral distribution of the emission of an ideal blackbody at 500 K (227°C) with that of a nonideal emitter (e.g., could be a piece of bare metal) also at the same temperature. In this case the nonideal emitter has low emissivity at wavelengths in the MWIR spectral region (3-5 μm) and generally high emissivity in the LWIR spectral region (8-14 μm).

Under most practical conditions, 2D imaging array thermal IR sensors (i.e., what are termed *staring arrays*) measure simultaneously over broadband wavelength spectra, as opposed to making measurements at narrow, almost monochromatic, wavelengths (e.g., an IR spectrophotometer which measures only one point in a scene). With a staring array sensor it is possible to measure *average emissivity* over a broadband spectrum (e.g., 3-5 μm , 8-14 μm), which in Figure 6.4(b) is simply the ratio of the area under the nonideal curve to the area under the Planck curve over the respective wavelength spectrum.

Some of these principles can be observed in Figure 6.5 (in Section 6.4). Plastic materials transparent in the visible spectrum that compose glasses are opaque in the LWIR and appear dark. Emissivity of this material is small in the visible spectrum while being significantly above 0.80 in the MWIR and LWIR spectral regions. The dark appearance of glasses in the LWIR and the MWIR relative to thermal emission from human facial skin is mostly due to the glasses being close to room temperature, about 15°C cooler than body temperature. We performed simple experiments whereby these same pair of glasses were heated close to body temperature. Sure enough, the glasses appeared thermally much brighter, but did not show as much thermal emission as facial skin at the same temperature. Also, from Figure 6.5 the influence of reflection of external illumination from glasses is far more prominent than

that from facial skin. All of this initially suggests that facial skin has very high emissivity, significantly higher than that of the material comprising glasses. A quantitative estimate of the average emissivity of facial skin in the MWIR and LWIR is developed in Section 6.5, supporting this assertion.

Just like visible video cameras, thermal IR cameras measure energy of electromagnetic radiation, the main difference being that because thermal IR cameras sense at such long wavelengths, they measure radiation that has been typically thermally emitted. Of course, visible cameras see radiation emitted from very hot sources (e.g., the sun or artificial lightbulbs which are thousands of degrees Kelvin) but the primary scene elements of interest in the visible are objects from which such light is reflected. Sometimes there is the misconception that thermal IR cameras directly measure temperature, which would be true if all objects were blackbodies. Temperature can be determined indirectly from a thermal IR camera by measurement of energy of emitted radiation, using precise knowledge of emissivity of the object, which is dependent upon a number of parameters.



Figure 6.5. A qualitative demonstration of the illumination invariance for LWIR imagery of a face under different illuminations. Top row: Visible imagery of a face under three illumination conditions respectively front, left, and right. Bottom row: Co-registered thermal IR imagery simultaneously acquired for each of the three images in top row respectively.

Thermal IR cameras can be radiometrically calibrated using a blackbody ground-truth source. Radiometric calibration achieves a direct relationship between the gray value response at a pixel and the absolute amount of thermal emission from the corresponding scene element. This relationship is called *responsivity*. Depending on the type of thermal IR camera being used, thermal emission flux is measured in terms of watts/cm² or photons/(cm² – second) [11]. The gray value response of pixels for a MWIR camera with an in-

dium antimonide (InSb) focal plane array is linear with respect to photons/(cm²–second). The gray value response of pixels for an LWIR camera using a microbolometer focal plane array is linear with respect to watts/cm². Two-point radiometric calibration uses a blackbody plate filling the field of view of the thermal IR camera and capturing images for the blackbody at two different temperatures. Given that human body temperature is 37°C, two good temperatures to use for calibrating the imaging of humans in a room temperature scene would be 20°C and 40°C (293 K and 313 K), as these are relatively evenly spread about the temperature of skin. A relatively large difference between the calibration temperatures will insure numerical stability of the linear regression, while a choice of temperatures nearby the temperature of interest minimizes possible effects from a secondary nonlinear response of the focal plane array.

Since absolute thermal emission is known by computing the area under the Planck curve for the corresponding temperature and wavelength spectrum, a responsivity line is generated at each pixel by two (greyvalue, thermal emission) coordinate values. The slope of this responsivity line is called the “gain” and the vertical translation of the line is “offset.” The gain and offset for each pixel on a thermal IR focal plane array can be significantly variable across the array. Radiometric calibration standardizes thermal emission measurement by generating a responsivity line for each pixel.

Figure 6.6 shows responsivity lines respective to different integration times, for a single pixel near the center of a MWIR InSb focal plane array that was used to collect face imagery. Eight different temperatures of a blackbody were used to generate multiple data points demonstrating the highly linear response. It is clearly important to record all thermal IR camera parameters for a given radiometric calibration. Note that the responsivity lines for different integration times intersect at the same point, related to various DC bias control settings on the camera. Beyond camera parameters, if an MWIR or LWIR camera is originally radiometrically calibrated in an indoors environment, taking it outdoors where there is a significant ambient temperature difference, the gain and offset of linear responsivity of focal plane array pixels will change as the optical lens temperature in front of the focal plane array changes. Radiometric calibration standardizes all thermal IR data collections, whether they are taken under different environmental factors or with different thermal IR cameras or at different times.

6.4 Measuring Illumination Invariance

Variation in illumination is one of the biggest factors that confounds face recognition algorithms in the visible spectrum [12, 13]. It has been recognized in the past [2, 14, 6, 15] that changes in illumination appear to play less of a role in the thermal infrared, but how does one quantify this invariance in terms that are meaningful to face recognition? One way is to quantitatively compare

the effect that variation in illumination has on face images in the thermal infrared with other factors that contribute to changes in face imagery, such as variations in facial expression and more subtle variations due to camera noise.

Illumination invariance of the human face in the thermal infrared can be qualitatively observed in Figure 6.5 for a coregistered LWIR and visible video camera sequence of a face under three different illumination conditions. For this sequence a single 60-W light bulb mounted in a desk lamp illuminates a face in an otherwise completely dark room and is moved into different positions. The top row of visible video imagery shows dramatic changes in the appearance of the face. The bottom row shows LWIR imagery which, unlike its coregistered visible counterpart, appears to be remarkably invariant across different illuminations, except in the image area corresponding to the glasses. As we will see, illumination invariance in the thermal infrared, while not being completely ideal, is nonetheless strongly approximate.

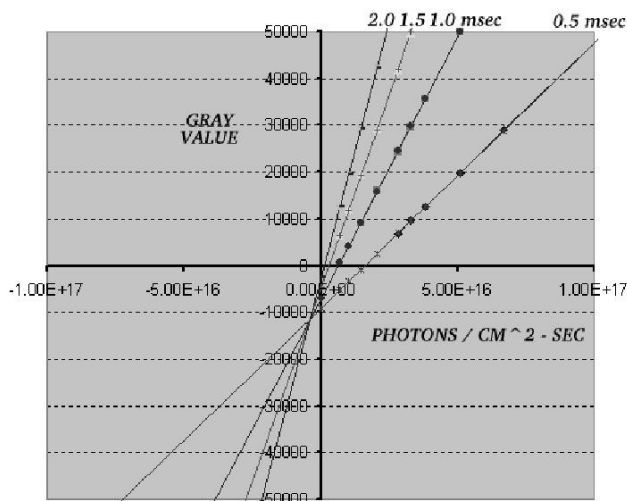


Figure 6.6. Responsivity curves for different integration times for the Indigo Merlin Series MWIR camera used for collecting face images.

Figure 6.7 shows simultaneously acquired MWIR and LWIR images of a subject from the Equinox database, together with corresponding gray value histograms of an individual under the three illumination conditions previously described. These images are the third image frame out of each respective 40-image frame sequence. Gray values in the histograms are represented as 16-bit integers with the high 12-bits being the actual image gray value. The gray level histograms are remarkably stable across different illuminations for both the MWIR and the LWIR images. Of the variations that are present in the respective histograms, which are due to change in illumination and which are

due to other factors? For instance, note the darker mouth region in the MWIR image for right illumination as compared to the mouth region in the MWIR images for other illuminations. The darker mouth region is due to the subject breathing-in room temperature air at the moment, thereby cooling down the mouth. This has nothing to do with any illumination condition.

The histograms in Figure 6.7 can be compared with those in Figure 6.8, which shows gray value histograms corresponding to the fourth and twentieth image frame out of the 40-image frame sequence respective to the frontal illumination condition. In this case, illumination is the same but the fourth frame being consecutive with the third frame isolates changes due to camera noise, and the twentieth frame occurring just under two seconds later means the subject has changed facial expression. The variations in the gray level histogram due to camera noise and to different facial expression under the same illumination are of similar magnitude to variations occurring under different illumination.

A quantitative analysis of invariance in the framework of hypothesis testing was also performed. The following analysis is repeated for two different distance measures between images. Firstly we consider the L^2 distance between normalized images taken as vectors. Secondly, we use the Kullback–Leibler divergence³ between the histograms of the normalized faces, given by

$$I(P, Q) = \int P \log \frac{P}{Q},$$

where P and Q are the respective normalized histograms.

For each video sequence of $40 + 3$ frames⁴, we compute the $43 \cdot 42/2 = 903$ distances between normalized faces for distinct pairs of frames. Also, we compute the $43 \cdot 43 = 1849$ distances between normalized faces for sequences of the same subject and modality, one sequence with frontal illumination and the other with lateral illumination. From these computations we estimate (nonparametrically) the distribution of distances for images with the same illumination condition and with different illumination conditions. Figures 6.9, 6.10, 6.11, and 6.12 show the estimated distributions for the L^2 distance and KL -divergence for two subjects in our database. With an infinite supply of images, we would expect the distances to behave according to a χ distribution with the number of degrees of freedom matching the number of pixels in the normalized faces, and indeed the experimental estimates approximate χ distributions.

It is clear from Figures 6.9, 6.10, 6.11, and 6.12 that the distances between normalized visible faces with different illumination conditions are much larger

³ The Kullback–Leibler divergence does not satisfy the triangle inequality, and thus is not strictly a distance. However, it provides an information-theoretic measure of similarity between probability distributions.

⁴ Recall that 40 consecutive video frames were collected while subjects recited the vowels, and then three additional static frames were acquired while the subjects were asked to act out the expressions “smile,” “frown,” and “surprise.”

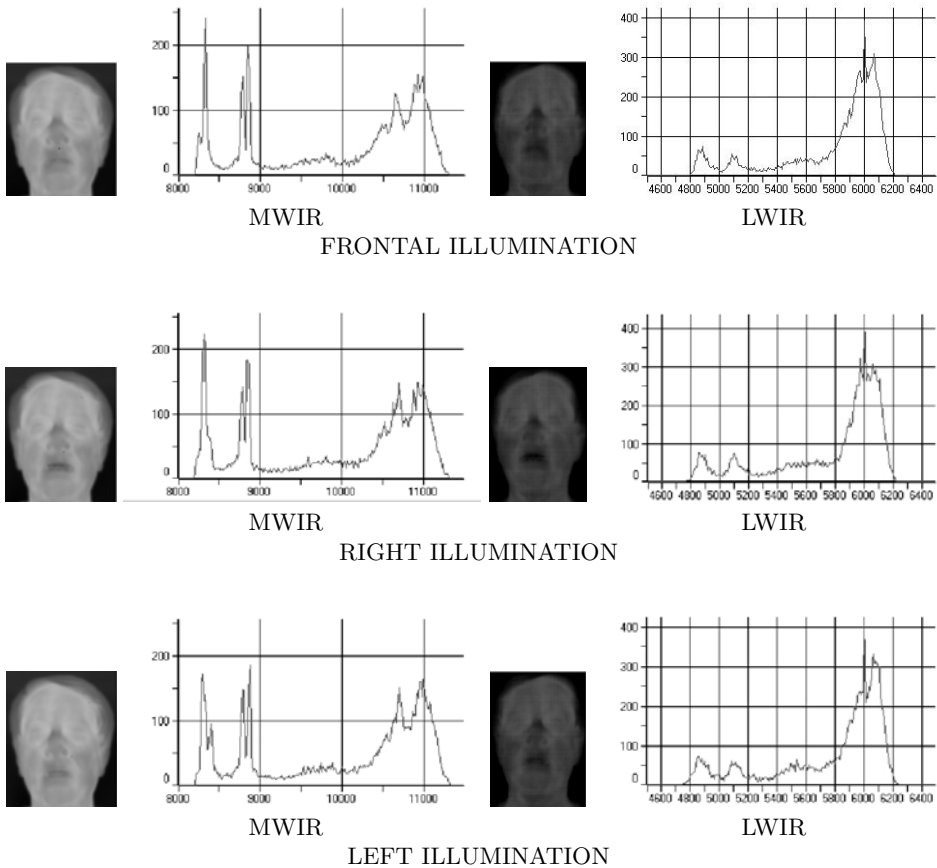


Figure 6.7. MWIR and LWIR imagery of a face for three illumination conditions and respective histograms of the third frame out of a sequence of 40 images.

than those for visible faces with the same illumination condition. This indicates that the variation in appearance due to change in illumination is much larger than that due to change in facial expression. The corresponding statement for LWIR imagery does not hold. That is, looking once again at Figures 6.9, 6.10, 6.11, and 6.12, one can see that the distribution of distances between normalized faces with different illumination conditions is comparable (but not equal; see below) to the distribution obtained by using images acquired with the same illumination condition. In other words, the variation in appearance introduced by changes in illumination and expression is comparable to that induced by changes in facial expression alone. Phrasing these statements as formal hypothesis, we can reject the null-hypothesis of illumination invari-

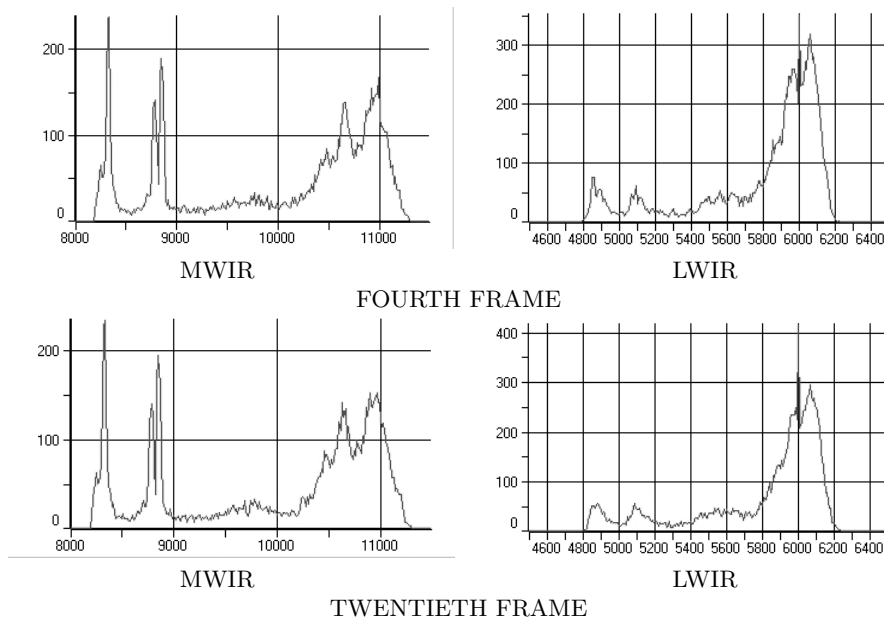


Figure 6.8. MWIR and LWIR imagery of the same face as Figure 6.7 respective to frontal illumination for the fourth frame (top row) and twentieth frame (bottom row) out of a sequence of 40 images.

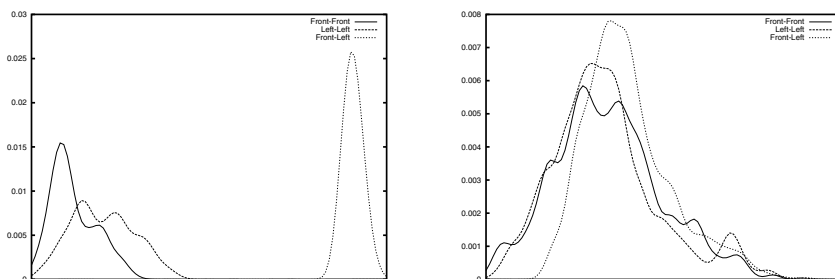


Figure 6.9. Distribution of L^2 distances for visible (left) and LWIR (right) images of subject 2344.

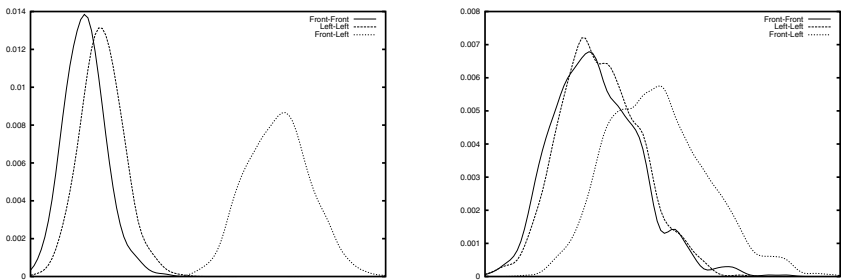


Figure 6.10. Distribution of Kullback–Leibler divergences for visible (left) and LWIR (right) images of subject 2344.

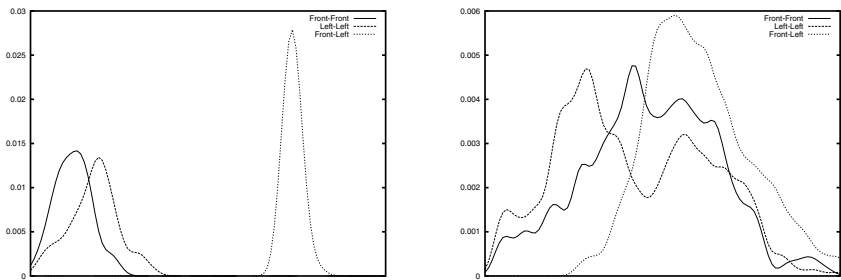


Figure 6.11. Distribution of L^2 distances for visible (left) and LWIR (right) images of subject 2413.

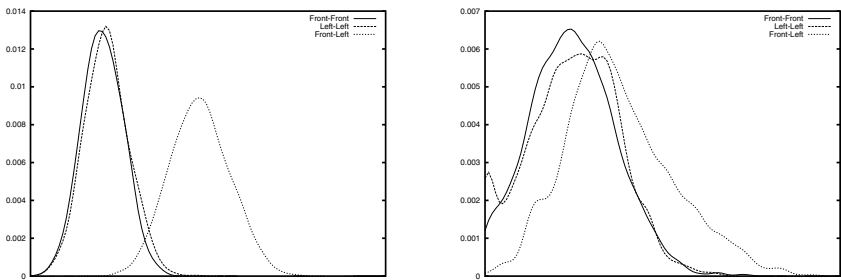


Figure 6.12. Distribution of Kullback–Leibler divergences for visible (left) and LWIR (right) images of subject 2413.

ance for visible imagery with a p -value smaller than 0.01⁵, whereas we are unable to reject the null-hypothesis for LWIR imagery with any significant confidence. The slight shift in the distributions to the right for variable illumination suggests that illumination invariance in the LWIR is not completely ideal.

6.5 Emissivity Of Human Facial Skin

Figure 6.4(a) shows that the amount of thermal emission from a common light-bulb is three to four orders of magnitude greater than the thermal emission from skin in both the 3–5 μm MWIR region and the 8–14 μm LWIR region. Empirical observation with our own MWIR and LWIR cameras showed that direct illumination from an incandescent filament through lightbulb glass and plastic diffuser is at least 300 times greater than thermal emission from human facial skin. This is a rather striking fact given that thermal IR imagery of faces is highly illumination invariant. Human skin must absorb a large quantity of radiation in both the MWIR and the LWIR implying that skin has very high emissivity.

Figure 6.13 shows a human subject in the same scene with a 6in * 6in square blackbody (Mikron Model 345) imaged in the MWIR and LWIR spectra. Separate images are taken for the blackbody at two different temperatures: 32°C and 35° C. The corresponding histograms show gray value modes for the facial skin image region and for the blackbody image region. Prior to imaging, an Anritsu thermocouple was used to make contact temperature measurements on the forehead, on both cheeks and on the chin of the human subject. An average skin surface temperature of 32°C was observed. Note, however that the face thermally emits more energy than does a 32°C blackbody. Recall that a blackbody is, by definition, a perfect emitter at all temperatures and wavelengths. Therefore, we have a physical contradiction unless we can account for the extra radiation. Since the path self-emission from the atmosphere between the subject and the sensor is negligible compared to the emission from the subject, we conclude that the extra radiation must be originating below the skin surface (where body temperature is around 37°C) and shining through the translucent skin layer and onto the sensor. This may reveal an important aspect of how thermal emission arises from human anatomy and perhaps even a physical mechanism for why skin has such high absorption in the thermal IR.

Figure 6.14 illustrates a preliminary high-level model of human skin in terms of optical and thermal properties. Evidently, skin layers must be significantly transmissive to thermal emission from underlying internal anatomy

⁵ This means that the likelihood of our rejecting the hypothesis of illumination invariance for visible imagery while at the same time the hypothesis being true is lower than 1%[16].

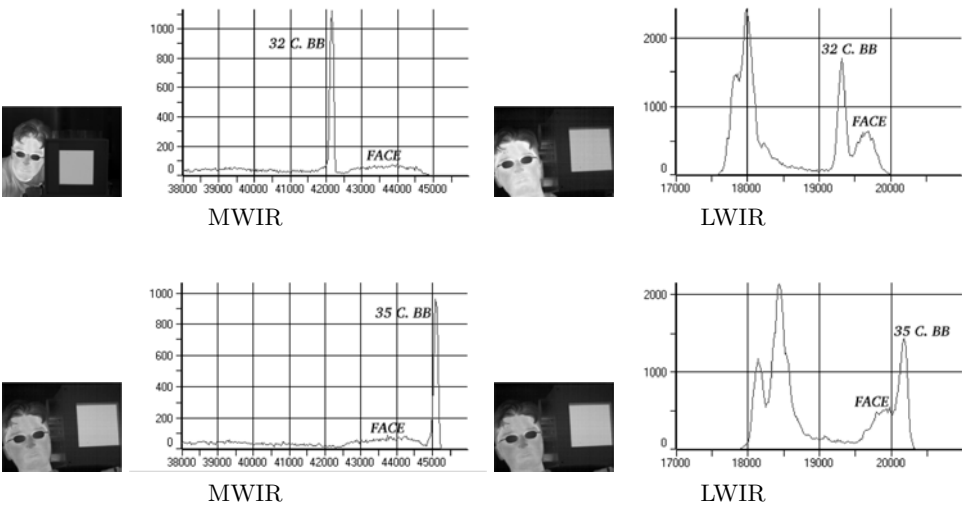


Figure 6.13. Direct comparison of MWIR and LWIR imagery of a face with a groundtruth blackbody at two different temperatures, 32°C and 35°C.

which is at a higher temperature. This is qualitatively evidenced from thermal observation of prominent vasculature beneath the skin particularly in the neck and forehead. Just how far below the skin surface thermal emission is transmitted is unclear and is an avenue for future research. If at least the outer layers of skin are transmissive, then incident thermal IR illumination must be first transmitted and then absorbed within deeper layers of skin or other anatomy. This may explain why the amount of thermal emission from skin seems to be independent of external skin color in the visible spectrum.

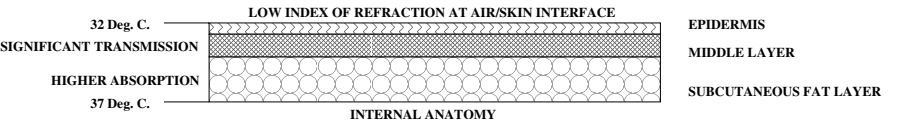


Figure 6.14. Preliminary thermal model for human skin.

We now proceed to compute a quantitative estimate of the average emissivity respective to the MWIR and the LWIR for human facial skin from the data in Figure 6.13. First we compute the mean thermally emitted energy of facial skin $\text{Skin}_{\text{energy}}^{\text{mean}}$. Since the thermal IR imagery used is radiometrically calibrated, we can compute the mean gray value in the histogram for the facial lobe and determine the corresponding energy by linearly interpolating

between the gray value peaks for the blackbody at 32°C (305 deg. K) and 35°C (308 K) and respective blackbody energies. For the MWIR this is

$$\text{Skin}_{\text{energy}}^{\text{mean}} = BB_{\text{energy}}^{305K} + [BB_{\text{energy}}^{308K} - BB_{\text{energy}}^{305K}] \frac{\text{Skin}_{\text{gray}}^{\text{mean}} - BB_{\text{gray}_{\text{max}}}^{305K}}{BB_{\text{gray}_{\text{max}}}^{308K} - BB_{\text{gray}_{\text{max}}}^{305K}}, \quad (6.3)$$

where

$$BB_{\text{energy}}^{308K} = \int_3^5 Q(\lambda, 308K) d\lambda,$$

$$BB_{\text{energy}}^{305K} = \int_3^5 Q(\lambda, 305K) d\lambda.$$

For the LWIR replace $Q(\lambda, T)$ with $W(\lambda, T)$ and integration occurs over wavelengths from 8 to 14 microns.

We then make a conservative estimate of the lower bound for average emissivity, ϵ , by comparing the mean thermally emitted energy of facial skin to a blackbody at internal body temperature 37°C. This yields:

$$\epsilon_{\text{mwir}}^{\text{skin}} > \frac{\text{Skin}_{\text{energy}}^{\text{mean}}}{\int_3^5 Q(\lambda, 310K) d\lambda} = 0.91,$$

$$\epsilon_{\text{lwir}}^{\text{skin}} > \frac{\text{Skin}_{\text{energy}}^{\text{mean}}}{\int_8^{14} W(\lambda, 310K) d\lambda} = 0.97.$$

These lower bounds are conservative as this effectively assumes that thermal emission is being sensed from a material that has a temperature of 37°C throughout. In reality there is a temperature gradient from the skin surface at 32°C through skin layers and blood vessels eventually to 37°C internal body temperature. The average temperature lies somewhere between 32 and 37°C. It is clear that skin at least has high emissivity in the MWIR and extremely high emissivity in the LWIR supporting a physical basis for excellent illumination invariance.

As the emissivity of skin is so close to 1.0, it is meaningful to quantify what is the average skin temperature due to the internal temperature gradient below the skin. This can be defined in terms of a blackbody equivalent temperature of skin, to be the temperature of a blackbody emitting equivalent energy as $\text{Skin}_{\text{energy}}^{\text{mean}}$. This temperature, $\text{Skin}BB^T$, can be computed by numerically solving the following integral equations:

$$\int_3^5 Q(\lambda, \text{Skin}BB_{\text{MWIR}}^T) d\lambda = \text{Skin}_{\text{MWIR}}^{\text{mean}} \text{ energy},$$

$$\int_8^{14} W(\lambda, \text{Skin}BB_{\text{LWIR}}^T) d\lambda = \text{Skin}_{\text{LWIR}}^{\text{mean}} \text{ energy}.$$

From the data presented in Figure 6.13 we compute:

$$\text{Skin}BB_{\text{MWIR}}^T = 34.3^\circ\text{C}, \quad \text{Skin}BB_{\text{LWIR}}^T = 34.7^\circ\text{C}.$$

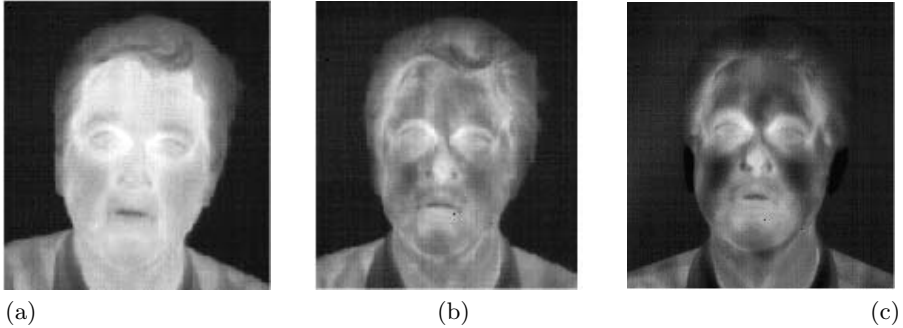


Figure 6.15. Face imaged in the LWIR (a) low activity, (b) after jogging, (c) after being outdoors at 0°C ambient temperature.

Modeling thermal emission from human faces is a good first step toward improving infrared face recognition performance. Much as understanding reflective phenomenology in the visible spectrum has led to development of algorithms that take explicit account of illumination variation [17, 18, 19], the same is true for understanding of underlying emissive phenomenology in the thermal context.

6.6 Comparison Of Face Recognition Performance In The Visible And Thermal Infrared

Over the course of the last two years, successively more comprehensive performance testing of existing appearance-based face recognition algorithms has occurred on the Equinox visible/infrared database [8, 9, 20]. The algorithms tested include principal component analysis (PCA) also known in the face recognition community as eigenfaces [12], local feature analysis (LFA) [21], linear discriminant analysis (LDA) — also known in the face recognition community as *Fisherfaces* [22] — and independent component analysis (ICA) [23]. Although we have available to us imagery from the visible spectrum and three different infrared spectra, we selected for direct comparison the visible and LWIR spectra since they are the most complementary respective to reflective versus emissive phenomenology. Also of key importance has turned out to be experimentation with precisely coregistered fusion of visible and LWIR imagery.

Prior to the mid-1990s, Wilder et al. [15] had directly compared performance on a smaller dataset of visible and thermal infrared imagery. In this case

the thermal infrared was MWIR imagery taken with a platinum silicide sensor. Their study concluded that both modalities yielded approximately equal performance. No image fusion of visible and MWIR was tested, although it was suggested in conclusion that such fusion might be beneficial.

Only the most basic features of testing conducted in [8, 9, 20] will be reviewed presently, and these references should be consulted for further details. Prior to testing, each face image is preprocessed using standard geometric image normalization techniques by manually locating eye features and frenulum. These images are then subsampled and subsequently cropped to remove all but the inner face. Figure 6.16 shows examples of normalized visible and LWIR image pairs from the Equinox database. As discussed in [24] face recognition performance is analyzed using pairs of sets called *gallery* and *probe* face image sets⁶. The gallery is an exemplar image set of known individuals, and the probe is an image set of unknown individuals that need to be classified. For testing correct identification performance of a face recognition algorithm, consecutive gallery images can be rank-ordered with respect to how well they match an unknown probe image with the closest gallery image match being the highest rank and consecutive lower rankings corresponding to consecutively worse matches with respect to a given metric. One way of quantifying correct identification by a given algorithm is by the percentage of probe images that correctly correspond to the matched individual who is highest ranked in the gallery. Table 6.1 shows a brief summary of performances for different algorithms on visible, LWIR, and fused imagery. This test set is particularly challenging for two reasons. First, the gallery and probe images were taken at different times, ranging from six months to two years apart. Secondly, while all the gallery images were acquired indoors, a portion of the probe images were acquired outdoors. We see that a PCA-based algorithm has very low performance on both visible and LWIR. Interestingly, in this case fused performance is actually lower than LWIR performance. This occurs only when recognition performance in one or both modalities is severely impaired, as is the case here. Performance for an LDA-based algorithm is much better, and exhibits improvement when visible and LWIR results are fused. Best performance on this set is obtained with Equinox's proprietary fused algorithm, which reduces the residual error by about 23% over the fused LDA-based result.

Figure 6.17 depicts distributions which compare the performance of an LDA-based algorithm with respect to a Monte Carlo simulation of 30,000 gallery-probe image set pairs for visible, LWIR and fused visible/LWIR modalities. Figure 6.17(a) shows the distribution of top-match recognition performance for an LDA-based algorithm when applied to visible, LWIR and fused imagery. It is easy to see that mean recognition rates are considerably higher for LWIR imagery than for visible imagery, and that when both modalities are fused, recognition performance climbs even higher. Not only is the mean

⁶ A third set, the *training* set, is used to determine algorithm parameters, and is disjoint from gallery and probe sets.



Figure 6.16. Example of visible (top) and LWIR (bottom) normalized face images.

Table 6.1. Top match recognition performance summary for different algorithms on visible, LWIR and fused imagery. Probe images are six months to two years older than corresponding gallery images. All gallery images are taken indoors, while some probes are taken outdoors.

	PCA Angle	LDA Angle	Equinox
Visible	19.355	55.323	
LWIR	30.968	61.452	
Fused	23.548	74.451	80.323

correct identification highest for fused visible/LWIR but note the smaller standard deviation, indicating more stability over variations in gallery and probe sets. In Figure 6.17(b), we see paired performance differences for the same set of experiments. In this case for each random experiment, the performance difference between LWIR and visible and fused and visible becomes one data point. The distribution of these differences, shown in the figure, indicates that LWIR affords an average performance gain of 6 percentage points over visible imagery, while fusion of visible and LWIR increases that gain to 9 percentage points. Note that this constitutes a reduction of the residual error by 75% when using fused imagery versus visible imagery alone.

Figure 6.18 shows a set of receiver operating characteristic (ROC) curves for an LDA-based facial verification algorithm applied to visible, LWIR and fused visible/LWIR imagery. These curves are obtained by averaging the results from thousands of experiments generated by randomly selecting different nonoverlapping gallery and probe sets. Recall that an ROC curve shows the trade-off between correct verification versus false acceptance, as the security setting of the system is varied from low to high. The equal-error-rate (EER) is the point on the curve at which false acceptance equals correct verification, and is often used as a scalar summary of the entire ROC curve, with lower values indicating higher performance. In this case, we see that the use of fused visible/LWIR imagery cuts the EER by more than 50% versus visible imagery alone.

An end-to-end face recognition system based upon coregistered visible and LWIR imagery has been developed. The bottom row of images in Figure 6.5 illustrates a visualization of fused visible/LWIR imagery. This system, called

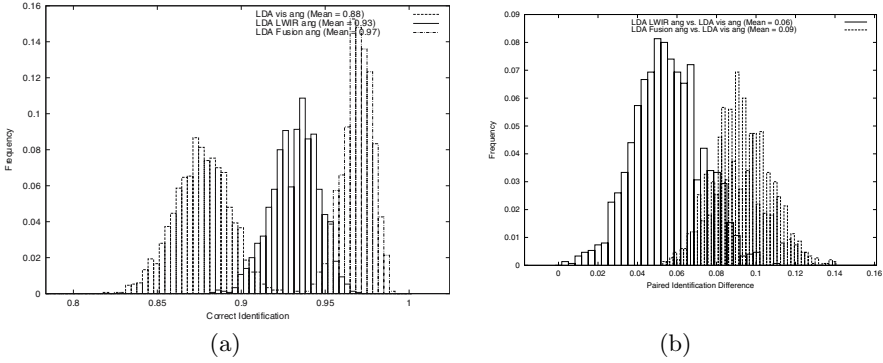


Figure 6.17. Performance comparisons of linear discriminant analysis (LDA) for visible, LWIR and fused visible/LWIR modalities. (a) Performance distributions, (b) paired performances. Taken from [20].

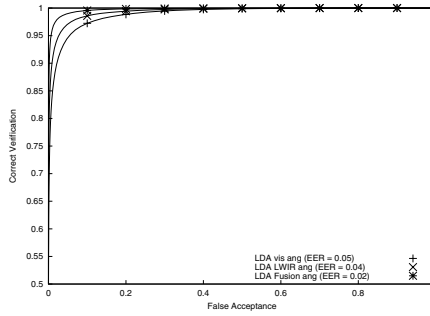


Figure 6.18. Receiver operator characteristic (ROC) curve comparing the same LDA-based algorithm for visible, LWIR, and fused visible/LWIR modalities. Taken from [20].

the Equinox access control environment (ACE), is capable of enrolling individuals into a database, and then for unknown individuals automatically detecting their faces in an image and recognizing whether they belong to the database. Face detection is another involved technical aspect separate from the recognition stage, and is beyond the scope of this article [25]. A recent demonstration of the Equinox ACE system over three days in July 2002 enrolled 105 individuals into this system at a trade show. Approximately two-thirds of these individuals returned to be recognized of which well over 90% were correctly identified as a top ranked match in the database. This included subjects that purposely attempted to fool the system by partially obstructing the view of their face, or attempting to mask their thermal appearance by

applying ice cubes to their face. Plate I shows one of the interfaces for this system explained in the caption. Note that while the system exploits fused visible and thermal imagery, the interface may be used in visible-only mode for ease of interpretation by the operator.

6.7 Conclusions, Challenges Ahead, and On-Going Work

This article provides a broad overview of research that has been proceeding at Equinox Corporation for the past two years on using thermal infrared imagery for enhancing face recognition. The following key aspects were described:

- Collection of a comprehensive database of thermal infrared imagery of human faces incorporating radiometric calibration, multiple illumination conditions, and imagery of duplicated individuals over time.
- Extensive experimental testing of the performance of appearance-based face recognition algorithms, directly comparing performance on visible, LWIR, and fused visible/LWIR modalities.
- Quantification of illumination invariance and examination of physical phenomenology responsible for illumination invariance of human faces in thermal infrared imagery.
- Development of a working end-to-end face recognition system using a novel sensor configuration that precisely coregisters visible and LWIR image modalities.

Statistically significant evidence was presented indicating that appearance-based face recognition algorithms applied to thermal infrared, particularly LWIR imaging, have consistently better performance than when applied to visible imagery. Application of these algorithms to fused visible/LWIR consistently showed even better improvement in performance.

To date, the largest issue not yet addressed by face recognition using thermal infrared is performance analysis under different extreme activity levels and extreme ambient temperature. It should be carefully noted that examples of extreme varying activity levels, which pose a potentially serious disadvantage to recognition using thermal infrared, have not yet been incorporated into the Equinox visible/infrared database. By virtue of the fact that the Equinox database is comprised of imagery taken at intervals of six months for over two years means that some range of normal activity level must be inherently incorporated for duplicate individuals. So far, this does not appear to have an adverse effect on recognition performance. More precisely, performance degradation over time is similar for visible and thermal imagery, with a possible advantage toward the thermal modality. However, it is important to note that data collections under different deliberate changes in activity level do need to be performed. Figure 6.15 shows the large difference in thermal infrared signatures for the same face at rest, after jogging and after coming in from outdoors at winter time. Additionally, the effect of other confusers such as

heavy makeup application must be evaluated. Existing data is not sufficient to perform a valid evaluation, but we intend to collect data specifically for this task. Evidence from small-scale experiments performed in-house indicates that recognition performance using fused visible/LWIR imagery remains high in the presence of differences in facial hair or glasses between gallery and probe images. These variations in appearance do not noticeably hinder our system's ability to recognize faces. We should, however, point out that these results may not be statistically significant, due to the small sample size. We do expect to see some performance degradation in a large-scale experiment. It has already been noted that thermal infrared imagery has the potential for being used to identify an individual's activity state and even state of inebriation [14, 26]. Unfortunately, this benefit may counterbalance to some degree the performance accuracy of unique face recognition capability. On-going work is incorporating more thermal infrared face imagery in outdoor environments, and will also shortly include varying activity and ambient temperature conditions as well.

References

- [1] Dowdall, J., Pavlidis, I., Bebis, G.: A face detection method based on multib and feature extraction in the near-IR spectrum. In: *Proceedings IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, Kauai, Hawaii (2002)
- [2] Pavlidis, I., Symosek: The imaging issue in an automatic face/disguise of detection system. In: *Proceedings IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, Hilton Head (2000)
- [3] Horn, B.: Understanding image intensities. *Artificial Intelligence* (1977) 1–31
- [4] Horn, B., Sjoberg, R.: Calculating the reflectance map. *Applied Optics* **18** (1979) 1770–1779
- [5] Siegal, R., Howell, J.: *Thermal Radiation Heat Transfer*. McGraw-Hill, New York (1981)
- [6] Prokoski, F.: Method for identifying individuals from analysis of elemental shapes derived from biosensor data. In: U.S. Patent 5,163,094, November 10 (1992)
- [7] DARPA Human Identification at a Distance (HID) Program, Equinox Corporation, P.W.: Using Visible and Thermal Infrared Imagery for Human Identification. DARPA/AFOSR Contract# F49620-01-C-0008 (2000-2003)
- [8] Socolinsky, D., Wolff, L., Neuheisel, J., Eveland, C.: Illumination invariant face recognition using thermal IR imagery. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Kauai, Hawaii (2001)

- [9] Socolinsky, D.A., Selinger, A.: A comparative analysis of face recognition performance with visible and thermal infrared imagery. In: Proceedings IAPR International Conference on Pattern Recognition, Quebec, Canada (2002)
- [10] Wyszecki, G., Stiles, W.S.: Color Science : Concepts and Methods, Quantitative Data and Formulae. Wiley Series in Pure and Applied Optics, John Wiley & Sons (1981)
- [11] Dereniak, E., Boreman, G.: Infrared Detectors and Systems. John Wiley & Sons (1996)
- [12] Turk, M., Pentland, A.: Eigenfaces for recognition. *J. Cognitive Neuroscience* **3** (1991) 71–86
- [13] Adini, Y., Moses, Y., Ullman, S.: Face rcognition: The problem of compensating for changes in illumination direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19** (1997) 721–732
- [14] Prokoski, F.: History, current status, and future of infrared identification. In: Proceedings IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications, Hilton Head (2000)
- [15] Wilder, J., Phillips, P., Jiang, C., Wiener, S.: Comparison of visible and infrared imagery for face recognition. In: Proceedings of 2nd International Conference on Automatic Face & Gesture Recognition, Killington, VT (1996) 182–187
- [16] P. J. Bickel, K. A. Doksum: Mathematical Statistics. Prentice-Hall, Englewood Cliffs, NJ (1977)
- [17] Shashua, A.: On photometric issues in 3D visual recognition from a single 2D image. *IJCV* **21** (1997) 99–122
- [18] Shashua, A., Raviv, T.R.: The quotient image: Class-based re-rendering and recognition with varying illuminations. *IEEE TPAMI* **23** (2001) 129–139
- [19] Zhao, W., Chellappa, R.: Robust Face Recognition using Symmetric Shape-from-Shading. Technical report, Center for Automation Research, University of Maryland, College Park, MD (1999) Available at <http://citeseer.nj.nec.com/zhao99robust.html>”.
- [20] Socolinsky, D.A., Selinger, A., Neuheisel, J.: Face recognition with visible and thermal infrared imagery. *Computer Vision and Image Understanding (CVIU) Special Issue on Face Recognition* (2003) Submitted.
- [21] Penev, P., Attick, J.: Local feature analysis: A general statistical theory for object representation. *Network: Computation in Neural Systems* **7** (1996) 477–500
- [22] Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions PAMI* **19** (1997) 711–720
- [23] Comon, P.: Independent component analysis: a new concept? *Signal Processing* **36** (1994) 287–314

- [24] Phillips, P., Moon, H., Rizvi, S., Rauss, P.: The FERET Evaluation Methodology for Face-Recognition Algorithms. Technical Report NISTIR 6264, National Institute of Standards and Technology (1999)
- [25] Eveland, C., Socolinsky, D., Wolff, L.: Tracking faces in infrared video. In: Proceedings IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications, Kauai, Hawaii (2002)
- [26] Prokoski, F.: Method and apparatus for recognizing and classifying individuals based on minutiae. In: U.S. Patent 6,173,068, January 9 (2001)