A Study on Using Mid-Wave Infrared Images for Face Recognition

Thirimachos Bourlai, Arun Ross, Cunjian Chen, Lawrence Hornak

West Virginia University
PO Box 6201, Morgantown, West Virginia 26506

ABSTRACT

The problem of face identification in the Mid-Wave InfraRed (MWIR) spectrum is studied in order to understand the performance of intra-spectral (MWIR to MWIR) and cross-spectral (visible to MWIR) matching. The contributions of this work are two-fold. First, a database of 50 subjects is assembled and used to illustrate the challenges associated with the problem. Second, a set of experiments is performed in order to demonstrate the possibility of MWIR intra-spectral and cross-spectral matching. Experiments show that images captured in the MWIR band can be efficiently matched to MWIR images using existing techniques (originally not designed to address such a problem). These results are comparable to the baseline results, i.e., when comparing visible to visible face images. Experiments also show that cross-spectral matching (the heterogeneous problem, where gallery and probe sets have face images acquired in different spectral bands) is a very challenging problem. In order to perform cross-spectral matching, we use multiple texture descriptors and demonstrate that fusing these descriptors improves recognition performance. Experiments on a small database, suggests that the problem of cross-spectral matching requires further investigation.

Keywords: Face Recognition, Mid-Wave InfraRed, Intra-spectral, Cross-spectral

1. INTRODUCTION

Most face recognition (FR) systems are based on images captured in the visible range of the electromagnetic spectrum (380-750 *nm*). However, in harsh environmental conditions characterized by unfavorable lighting and pronounced shadows (such as a nighttime environment[1]), human recognition based solely on visible spectral images may not be feasible.[2] Thus, recognition of faces in the infrared spectrum has become an area of growing interest[3].[4]

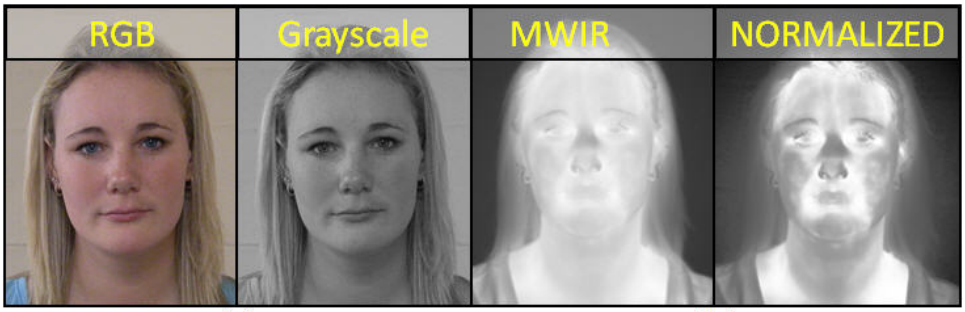


Figure 1. Frontal face images captured in the (a) visible (*Canon EOS 5D MII*) and (b) MWIR spectrum (*FLIR SC8000*).

Further author information: (Send correspondence to T.B.)
T.B.: E-mail: ThBourlai@mail.wvu.edu, Telephone: 1 304 293 4326
A.R.: E-mail: Arun.Ross@mail.wvu.edu, Telephone: 1 304 293 9135
C.C.: E-mail: cchen10@mix.wvu.edu, Telephone: 1 304 293 9135
L.H.: E-mail: Lawrence.Hornak@mail.wvu.edu, Telephone: 1 304 293 9694

Sensing Technologies for Global Health, Military Medicine, Disaster Response, and Environmental Monitoring II; and Biometric Technology for Human Identification IX, edited by Sárka O. Southern, et al., Proc. of SPIE Vol. 8371, 83711K

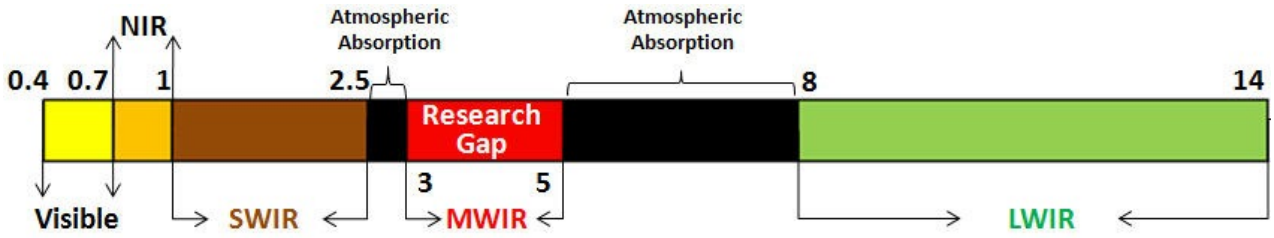


Figure 2. Electromagnetic Spectrum Map and the MWIR face recognition research gap. The illustrated wavelengths are in μm.

The infrared (IR) spectrum can be further divided into different spectral bands. The boundaries between these bands can vary depending on the scientific field involved (e.g., optical radiation, astrophysics, or sensor technology[5]). The IR bands discussed in this work are based on the response of various detectors (see Fig. 2). The IR spectrum is comprised of the active IR band (Near-IR or Short-Wave IR), and the thermal (passive) IR band. The passive IR band is further divided into the Mid-Wave (MWIR) and the Long-Wave InfraRed (LWIR) band. MWIR ranges from 3-5μm, while LWIR ranges from 7-14μm. Both MWIR and LWIR cameras can sense temperature variations across the face at a distance, and produce thermograms in the form of 2D images. However, while both pertain to the thermal spectrum, they reveal different image characteristics of the facial skin. Note that the word "thermal" does not precisely describe the spectral range considered in this work - "MWIR" is more precise (in fact, FLIR cameras are categorized as MWIR or LWIR).

There are differences between active and passive infrared bands. The difference between MWIR and LWIR is that MWIR has both reflective and emissive properties, whereas LWIR consists primarily of emitted radiation. The main benefits of MWIR over the active IR band are: (1) MWIR imagery can be acquired without any external illumination in day or night environments; however, regions in the active IR band may require an external light source; (2) anatomical features (e.g., vein patterns) that may not be observable in the active IR spectrum, may be observable in MWIR; (3) in MWIR images, background clutter is not visible and the tasks of face detection, skin localization, and segmentation can be easier and more reliable than in active IR and visible bands.

In general, MWIR cameras (when compared to LWIR) are also better suited to see at longer ranges.[6] Although the use of LWIR infrared images for biometric recognition has been the subject of research for several years,[7,8] MWIR has not received much attention in the open literature. In terms of face recognition, MWIR imaging was used in the DARPA Human ID Project, but did not yield promising results due to the immaturity of the sensor technology available at that time (2001-2005) relative to today's devices.

Given the significant improvement in MWIR technology in recent years, viz., higher resolution and thermal sensitivity, this paper will focus on answering the following questions: (1) Is face recognition in the MWIR band feasible, and can we effectively match MWIR face images against each other ? (2) Can we match MWIR face images against visible images? (3) Can system's cross-spectral matching performance improve with fusion and to what extent?

1.1 Goals and Contributions

In this work we investigate face recognition on the Visible-Thermal Database (VTD) that consists of visible and MWIR frontal face images of 50 subjects. Four different experiments have been performed. The *first experiment* investigates the matching of high quality face images obtained in the visible spectrum for the purpose of establishing baseline performance. We know that while low-light MWIR images would not affect recognition performance (MWIR is invariant to ambient illumination), low-light visible images will be affected (this is outside the scope of the paper). The *second experiment* compares MWIR against MWIR images. The *third experiment* compares facial images captured in the MWIR spectrum with those in the visible spectrum (heterogeneous problem), and also involves a feature level fusion study aimed at improving the recognition performance.

1.2 Paper Organization

The rest of this paper is organized as follows. Section 2 describes the MWIR imagery used in this work. Sections 3 and 4 provide a summary of the face recognition algorithms employed and the experiments conducted. Section 5 presents the results while conclusions are drawn in Section 6.

2. MID-WAVE INFRARED IMAGERY

- **Canon EOS 5D Mark II**: This digital SLR camera has a 21.1-megapixel full-frame CMOS sensor with DIGIC 4 Image Processor and a vast ISO Range of 100-6400. It also has *Auto Lighting Optimizer* and *Peripheral Illumination Correction* that enhances its capability. In this work, Canon (Fig. 1(a)) is used to obtain standard RGB, ultra-high resolution frontal pose face images in the visible spectrum.

- **FLIR SC8000**: The camera used in this work is a high definition MWIR camera produced by FLIR Systems*. It is capable of acquiring thermal-based imprints of human skin and analyzing the thermal distributions and temporal variations, corresponding to emission of 3-5 micrometer wavelength. The camera is capable of generating high definition thermal images and operating in diverse testing environments. It features a high resolution 1024×1024 Indium Antimonide (InSb) Focal Plane Array (FPA) achieving mega-pixel image resolution in a single thermal image. The spectral range of the camera is 3-5μm (corresponding to a temperature range of -20 to 500 degrees Celsius), and it has a 14 bit dynamic range and a Noise Equivalent Temperature Difference (NEDT) of less than 25mK. The camera was outfitted with a 50 mm MWIR lens also provided by FLIR Systems.

The aforementioned cameras were employed for data collection in order to assemble the visible and MWIR (VTB) Face database.

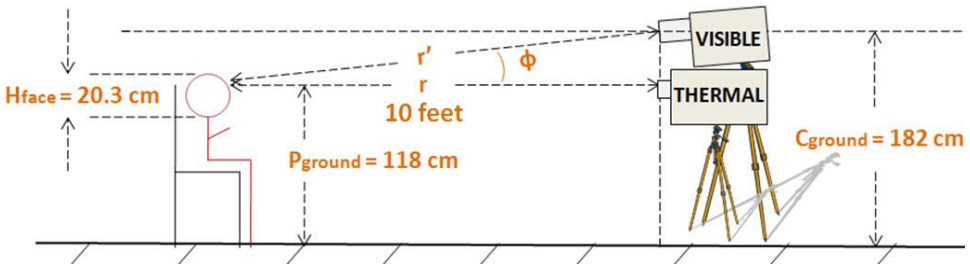


Figure 3. The live subject-capture setup using the visible and MWIR cameras.

The live face capture configuration we used is illustrated in Fig. 3. The standoff distance was set to 6.5 feet. The database was assembled indoors spanning over a time period of 20 days. In the beginning of the session, the subjects were briefed about the data collection process after which they signed a consent document. In total, 50 subjects (33 male + 17 female) participated in this experiment, and the database has 15 full frontal face MWIR and visible images of each subject, resulting in a total of 1,500 images (750 MWIR and 750 visible).

An important issue to discuss is the thermal sensitivity of the MWIR camera since it impacts image quality, i.e., the greater the sensitivity, the more accurate the camera can be in order to produce higher quality images [†]. The MWIR camera that we used has the highest sensitivity in the market. The associated camera software was used not only to perform data collection, but also regular calibration (before acquiring each set of thermal full frontal face images). Calibration ensures that the camera operates at its optimum performance and guarantees measurement accuracy and reliability. The camera operated in a controlled, low temperature environment (room

*"FLIR Systems," http://www.flir.com, 2011.

[†]The thermal sensitivity for an infrared camera is measured in milli-Kelvins (mK). Current cameras are able to detect temperature differences as small as 0.03^oC. And this equates to a thermal sensitivity of 30mK. The greater the sensitivity the more accurate the camera can be as well as producing more detailed images, i.e., 33 mK is nearly 3 times as sensitive as a 100 mK.

temperature) where better thermal sensitivity can be exploited since the thermal contrast (temperature delta within an image) is very low. The software tool provided by FLIR performs also non-uniformity correction, linearity correction and bad pixel replacement (less than 0.5% of the pixels are bad/dead) as post processing operations on the imagery. The software was set to control the temperature scale limits (e.g., setting the temperature range from 28^o to 41^o Celsius that is the typical range of human body temperature) during data collection.

3. INTRA-SPECTRAL MATCHING

In our intra-spectral (visible to visible, i.e., baseline, and MWIR to MWIR) experiments, we used both commercial and academic software. While the pre-processing routines employed by the commercial software are not known, the academic software employed the following pre-processing method.

3.1 Face Detection

First, the Viola & Jones *face detection* algorithm was applied to the baseline (visible) dataset of the VTB database. It was used to localize the spatial extent of the face and determine its boundary. In the case of the MWIR dataset, the face area was localized by first performing binirization of the image and then blob analysis to localize the face region.

3.2 Geometric Normalization

A *geometric normalization* scheme was applied to images acquired after face detection. The normalization scheme compensated for slight perturbations in the frontal pose, and consisted of eye detection and affine transformation. When using the visible dataset, automated eye detection was performed. The method used was based on a template matching algorithm where the coordinates of the eye were automatically obtained.[9] Traditional eye detection techniques did not work when employing the face image dataset acquired in the MWIR band. Thus, eyes centers were automatically located using the method proposed by Bourlai and Jafri,[10] designed specifically to operate with still frontal face images acquired in the MWIR band. In either dataset, after the eye centers were found, the canonical faces were automatically constructed by applying an affine transformation. Eventually, all faces were warped to the same dimension of 150×130 pixels.

3.3 Photometric Normalization

Although the MIWR images are robust to lighting variations, their representation changes after applying different normalization techniques (see Fig. 1). For *photometric normalization* and in order to facilitate *intra-spectral (MWIR to MWIR) matching*, we employed the contrast limited adaptive histogram equalization (CLAHE) technique.[11] CLAHE operates on local regions (8x8 for our experiments) in the image and applies histogram equalization, illustrated in equation (1), to each sub-region.

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

Here, M and N are the number of pixels and gray level bins in each sub-region, and h is the histogram of each sub-region. To increase contrast without amplifying noise, CLAHE redistributes each histogram such that the height falls below the clip limit threshold (.01 in our experiments). More specifically, gray level counts beyond the clip limit are uniformly redistributed among the gray levels below the clip threshold. Finally each sub-region is combined using bilinear interpolation.

3.4 Face Recognition Methods

Both commercial and academic software were employed to perform the face recognition experiments: (1) Commercial software *Identity Tools G8®* provided by L1 Systems[‡]; (2) standard face recognition methods provided by the CSU Face Identification Evaluation System,[12] including *Principle Components Analysis* (PCA),[13–15] *a*

[‡]www.l1id.com

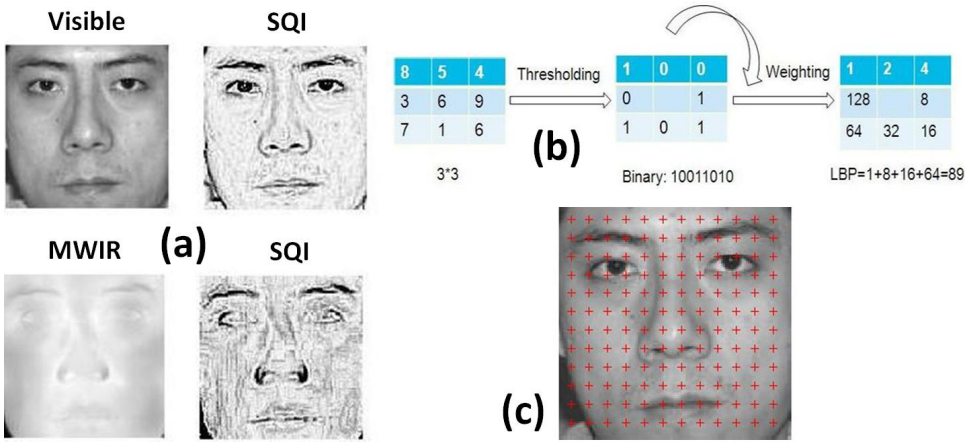


Figure 4. (a) Visible and MWIR images before and after applying the SQI normalization technique. Images are from the VTD database; (b) Definition of LBP operator; (c) Example of the dense sampling method. This is an example visible spectrum image taken from the VTD database.

combined Principle Components Analysis and Linear Discriminant Analysis algorithm (PCA+LDA),[16] and the *Bayesian Intra-personal/Extra-personal Classifier* (BIC) using either the Maximum likelihood (ML) or the Maximum a posteriori (MAP) hypothesis;[17] (3) the *Local Binary Pattern* (LBP) method;[18] and (4) the *Local Ternary Pattern* (LTP) method.[19]

4. CROSS-SPECTRAL MATCHING

4.1 Pre-Processing

The methods for face detection and geometric normalization were the same as in the intra-spectral matching experiments. Due to the complexity of the *cross-spectral matching* experiments (gallery and probe sets have face images pertaining to different spectral bands), various photometric normalization techniques were tested. For this purpose, we employed the Difference-of-Gaussian (DoG) Filtering method,[19] as well as a collection of the following photometric normalization techniques (employed from the INFace toolbox[20] implemented in Matlab): single-scale-retinex; multi-scale retinex; single-scale self quotient image; multi-scale self quotient image; and homomorphic-filtering-based normalization.

For the *cross-spectral matching* experiments, among all the illumination normalization approaches evaluated, the single-scale self quotient image (SQI) and the Difference-of-Gaussian (DoG) methods were observed to be the most effective ones:

(a) Self Quotient Image (Single-Scale SQI): According to the Lambertian model, the image formation process is described as follows:

$$I(x,y) = \rho_w(x,y)n(x,y)s, \tag{2}$$

where $\rho_w(x,y)$ is the albedo of the facial surface, n is the surface normal, and s is the lighting reflection. To reduce the impact of illumination, it is necessary to separate the extrinsic factor s from ρ and n. The self-quotient image, Q, of I is defined as,[20]:[21]

$$Q = \frac{I(x,y)}{\hat{I(x,y)}} = \frac{\rho_w(x,y)n(x,y)s}{G * [\rho_w(x,y)n(x,y)s]}, \tag{3}$$

where $\hat{I}$ is the smoothed version of I and G is the smoothing kernel. Samples of the normalization results are shown in Figure 4 (a).

(b) Difference-of-Gaussian (DoG) Filtering: It is another type of normalization that is utilized to increase the visibility of edges and other details present in the MWIR images.[19] It involves the subtraction of one blurred version of an original gray scale image from another, less blurred version of the original. The blurred

images are obtained by convolving the original gray scale image with two Gaussian kernels having differing standard deviations, e.g., σ_0 and σ_1:

$$D(x, y|\sigma_0, \sigma_1) = [G(x, y, \sigma_0) - G(x, y, \sigma_1)] * I(x, y). \tag{4}$$

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

$$G(x, y, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x^2+y^2)/2\sigma^2}. \tag{5}$$

4.2 Feature Extraction

Four different descriptors were used: LBP, LTP, Pyramid Histogram of Oriented Gradients (PHOG),[22] and Scale-Invariant Feature Transform (SIFT).[23] The LBP operator was introduced as a texture descriptor.[19] Patterns in an image are computed by thresholding 3×3 neighborhoods based on the value of the center pixel. Then, the resulting binary pattern is converted to a decimal value. In,[24] the authors proposed to extend it to include neighborhoods of different sizes to account for textures at different scales.

The local neighborhood is defined as a set of sampling points evenly spaced on a circle (Figure 4 (b)). The LBP operator used in our experiments is described as $LBP^{u^2}_{P,R}$, where P refers to the number of sampling points placed on a circle with radius R. The symbol u^2 represents the uniform pattern, which accounts for the most frequently occurring pattern in our experiments. The pattern is important because it is capable of characterizing patches that contain edges and corners. In our experiments, when MWIR images were employed, it was observed that 58 out of 256 8-bit patterns are uniform, which means the binary pattern contains at most two bit transitions. The other non-uniform patterns are summarized in one bin. The binary pattern for pixels, lying in a circle f_p, $p = 0, 1, \ldots, P-1$ with the center pixel f_c, is computed as follows,

$$S(f_p - f_c) = \left\{ \begin{array}{ll} 1 & \text{if } f_p - f_c \geq 0; \\ 0 & \text{if } f_p - f_c \leq 0. \end{array} \right\} \tag{6}$$

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

$$LBP_{P,R} = \sum_{p=0}^{P-1} S(f_p - f_c) 2^P. \tag{7}$$

LBP is invariant to monotonic gray-level transformations. However, one of the drawbacks is that LBP tends to be sensitive to noise in near-uniform image regions. This is because the binary code is computed by thresholding the exact center of the pixel region.

To overcome such a limitation, the LTP technique was proposed[19] where the quantization is performed as follows:

$$S(f_p - f_c) = \left\{ \begin{array}{ll} 1 & \text{if } f_p - f_c \geq t; \\ 0 & \text{if } |f_p - f_c| \leq t. \\ -1 & \text{if } f_p - f_c \leq -t. \end{array} \right\} \tag{8}$$

The output from LTP is a 3-valued pattern, instead of a binary pattern. The difference is that the threshold t can now be adjusted to produce different patterns. The user-specific threshold also makes the LTP code more resistant to noise.

PHOG is another spatial shape descriptor used in this work. It is generally used for image classification. It represents the spatial distribution of edges. This descriptor is formulated as a vector representation and is mainly inspired by two sources, i.e., the use of the pyramid representation[25] and the Histogram of Orientation Gradients (HOG).[26] Finally, we used the SIFT descriptor that is relatively fast to compute. SIFT has been used successfully for face authentication and face-specific SIFT features have been proposed in.[27]

4.2.1 Dense Sampling

The aforementioned feature extraction methods usually divide the patches into non-overlapping blocks. In this paper we considered a more flexible feature extraction approach by applying a dense sampling method. For instance, an image size of 139×139 pixels can be divided into patches of size 16×16 with 12 pixels patch displacement (Figure 4 (c)). In our experiments, there are totally 144 patches sampled from each input image.

Then, for each patch we extracted LBP and SIFT features. The PHOG descriptor was applied to the entire image. Therefore, both local and global information were captured. The way each feature descriptor was used is listed below:

- LBP: each patch produces a 59 dimensional feature vector. There are a total of $59 \times 144 = 8496$ features for a single image.
- PHOG: there are 680 features for three levels.
- SIFT: each patch produces a 128 dimensional feature vector. There are a total of $128 \times 144 = 18432$ features for a single image.

After feature normalization all the individual feature vectors were concatenated.[28]

4.2.2 TPLBP and FPLBP

Apart from the feature extraction methods that have been described, more advanced descriptors can be utilized to cope with the challenge cross-matching problems. LBP-based coding method exploits pixel-to-pixel relationship between center and its neighborhoods. This representation often characterizes the structure information well for the same modality. However, such assumption might not be valid for different modalities where variations caused by sensing is much more evident than intra-variations. Three-Patch LBP (TPLBP) and Four-Patch LBP (FPLBP),[29] on the other hand, encode the similarities between neighboring patches of pixels. This self-similarity based descriptor can effectively address the challenge nature of cross-spectral matching, exhibiting robustness across sensing difference. The TPLBP centering on the pixel x_c can be described as:

$$TPLBP(x_c) = \sum_{i}^{S} f(d(C_i, C_o) - d(C_{mod(i+\alpha,S)}, C_o))2^i, \tag{9}$$

where C_i and $C_{mod(i+\alpha,S)}$ are the two patches defined along the circle, and C_o is the center patch lying in the middle of the circle. Here, C_i, $C_{mod(i+\alpha,S)}$ and C_o are the three patches being utilized to calculate the descriptor, hence called Three-Patch LBP. S is the total number of patches sampled on the circle and w is the size of patch. The distance between the center patch and neighboring patch is determined by the radius r. The parameter α is to decide which pairs of patches to be considered along the ring. The similarity function d is defined by: $d(C_i, C_o) = \sum_{i,o}^{w}(x_i - x_o)^2$. Here, x_i and x_o are the pixels defined within the patches. Such similarity function can be replaced with any other functions that define the patch similarity. The function f is the threshold function that is similar to the one used in LBP. We found the Euclidean norm function sufficient in our case. The definition of FPLBP utilizes four patches to compute the binary code and is detailed described in the following:

$$FPLBP(x_c) = \sum_{i}^{S/2} f(d(C_{1i}, C_{2,mod(i+\alpha,S)}) - d(C_{1,i+S/2}, C_{2,mod(i+S/2+\alpha,S)}))2^i. \tag{10}$$

Here, C_{1i}, $C_{2,mod(i+\alpha,S)}$, $C_{1,i+S/2}$ and $C_{2,mod(i+S/2+\alpha,S)}$ are the four patches being utilized to calculate the descriptor. The definition of the parameters are consistent with TPLBP, though the layout of patches are different now. To better understand the descriptors, the readers are encouraged to refer to the original publication.[29] To extract image signature or features from TPLBP and FPLBP encoded images, we simply divide the image into non-overlapping patches as LBP did and spatial histogram features are computed within each patch. However, to make the histogram features more robust to illumination changes, they are normalized first to unit length

and truncated at 0.2 before being normalized again. As a consequence, the extracted image signature is a concatenation of normalized histograms. Such histogram-based signature considers both the geometry and texture information in image representation. For TPLBP, the default parameters are region size: 16×16, $alpha = 3$, $r = 1$,and $w = 3$. For FPLBP, the parameters are region size: 8×8, $alpha = 6$, $r1 = 2$, $r2 = 4$ and $w = 3$. The same parameters are used across both visible (Probe)-MWIR (Gallery) matching and MWIR (Probe)-visible (Gallery) matching.

4.3 Similarity Measures for Feature Matching

To compute the similarity between histogram features, four measures were used, i.e., chi-squared distance,[19] distance transform,[19] Euclidean distance (L2) and City-block distance (L1).

The chi-squared distance is defined as follows:

$$\chi^2(n,m) = \frac{1}{2}\sum_{1}^{l}\frac{h_n(k) - h_m(k)}{h_n(k) + h_m(k)}, \tag{11}$$

where h_n and h_m are the two histogram feature vectors, l is the length of the feature vector, and n and m are two sample vectors extracted from an image of the gallery and probe sets respectively.

The distance transform (defined as the distance or similarity metric from image X to image Y) is defined as follows:

$$D(X,Y) = \sum_{Y(i,j)} w(d_X^{k_{Y(i,j)}}(i,j)), \tag{12}$$

where $k_{Y(i,j)}$ is the code value of pixel (i,j) of image Y, and w is a user-controlled penalty function.[19]

The Euclidean distance is defined as follows:

$$d(h_n,h_m) = \sqrt{\sum_{k=1}^{l}(h_n(k) - h_m(k))}, \tag{13}$$

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

$$d(h_n,h_m) = \sum_{k=1}^{l}|h_n(k) - h_m(k)|. \tag{14}$$

5. EXPERIMENTAL SCENARIOS

By using the visible and MWIR images in the assembled database, four different types of face identification experiments were performed: (i) Visible vs. Visible (baseline), (ii) MWIR vs. MWIR, (iii) cross-spectral (Visible vs. MWIR) using individual feature vectors, and (iv) cross-spectral after feature-level fusion of the selected descriptors. Scenarios (iii) and (iv) represent a heterogeneous FR problem. The *first* experimental scenario matches high quality images in the visible spectrum for the purpose of establishing a baseline for comparison. In the *second* experiment (i.e., intra-spectral matching) we compare MWIR against MWIR face images. In the *third* experiment we analyze cross-spectral matching, i.e., we compare MWIR against visible spectrum images. Our *fourth* experiment involves fusion of multiple algorithms for improving the recognition performance when comparing photometrically normalized MWIR against photometrically normalized visible images. The experiment utilizes fusion algorithms on MWIR images whose matching performance was experimentally observed to be inferior.

The identification performance of the system is evaluated through the cumulative match characteristic (CMC) curve. The CMC curve measures the $1:m$ identification system performance, and judges the ranking capability of the identification system.

5.1 Intra-Spectral Matching

The results of the baseline experiments are illustrated in Fig. 5. We can see that when G8 was used, the identification rate was 100% at Rank-1. When texture based techniques were used (e.g., LTB, LTP), the rank-1 identification rate was 98.8%. Other appearance based techniques were also used (e.g., PCA, LDA) but achieved lower performance (e.g., PCA resulted in 71% identification rate at Rank-1).

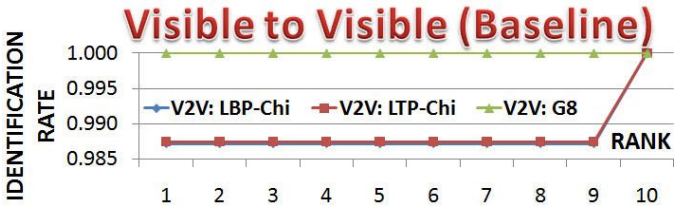


Figure 5. Baseline identification rates (Rank-1 to Rank-10): comparing visible to visible face images.

In the second set of experiments, we compared MWIR to MWIR face images. For this purpose we employed both commercial (Identity Tools G8, provided by L1 Systems) and academic software (CSU Face Identification Evaluation System[12]). The results are summarized in Fig. 6. There are three main conclusions here: (a) FR in MWIR spectrum is comparable to traditional FR when operating in the visible spectrum (e.g., for MWIR we achieved 100% identification rate at Rank-2). However, MWIR FR can be considered superior to visible FR in theory due to the fact that similar performance can be obtained when using MWIR face images acquired in complete darkness; (b) while texture based approaches would be expected to operate much better in the visible spectrum, results prove that there can be significant textural information in MWIR imagery that can result in comparable identification rates (i.e., 97% in MWIR vs. 98.75% in visible at Rank-1 when using LBP); (c) appearance based approaches (e.g., PCA) perform much better in the MWIR spectrum. This was expected due the fact that MWIR imaging is illumination invariant. A similar conclusion was reported in[30] where PCA was employed when using visible and Long-Wave Infrared face images.

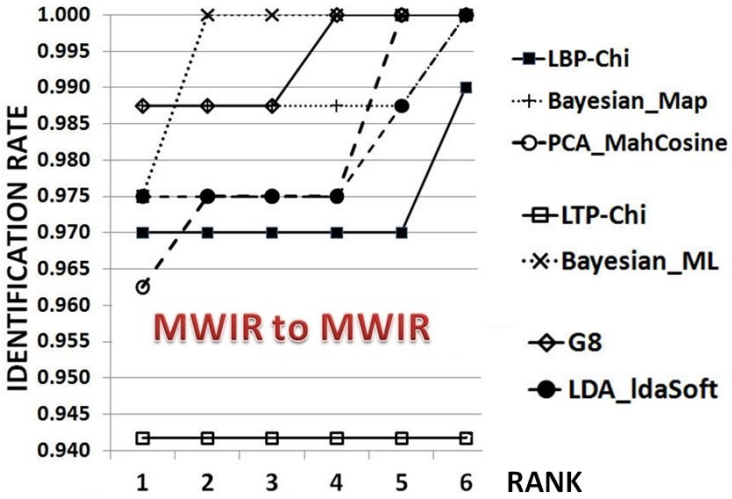


Figure 6. Intra-Spectral identification rates (Rank-1 to Rank-6): comparing MWIR to MWIR face images.

5.2 Cross-Spectral Matching

For this set of experiments we used 39 subjects (out of 50 in VTD)[§], with 4 samples per subject for the visible and 3 for the MWIR dataset. Face images were geometrically normalized based on eye coordinates and a mask

[§]We did not include subjects where their face images were not fully frontal

was applied to exclude background information (see Fig. 8). Six different descriptors have been used (LBP, LTP, PHOG, SIFT, TPLBP and FPLBP), two different illumination normalization techniques (SQI and DoG), and four similarity measures (L1, L2, CHI and DT as described in Section 4.3) were used. Feature level fusion was also employed (before and after applying normalization), when using the L1 and L2 measures. While all combinations of descriptors, illumination normalization and distance measures were tested, here we report the best results only. The scenarios investigated are the use of MWIR images as gallery and visible images as probes, and vice versa (Fig. 7). Both scenarios represent realistic practical scenarios as described in Section 1.1. All

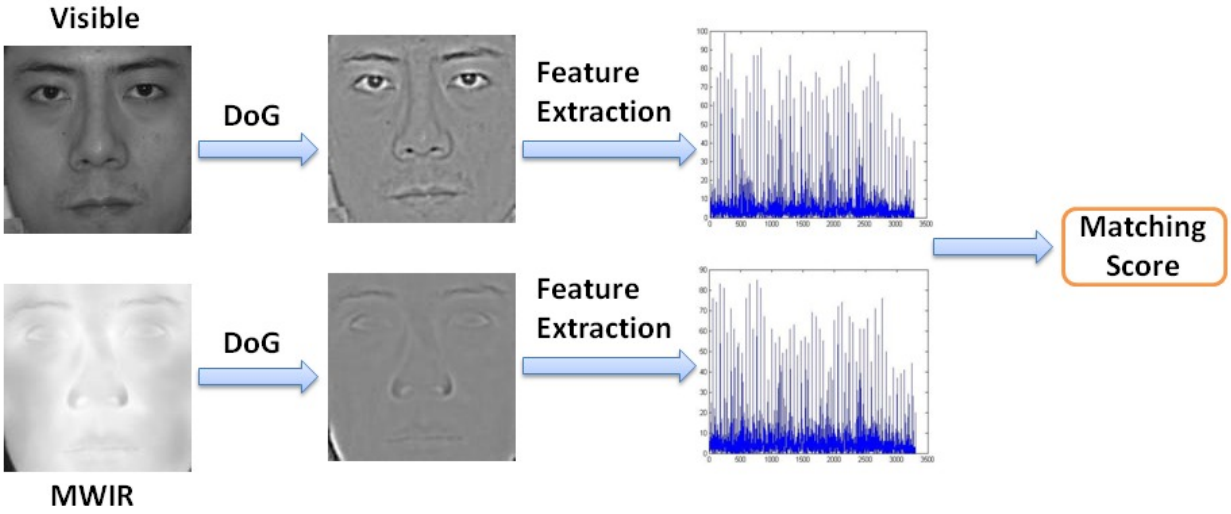


Figure 7. Flow chart of the cross-spectral matching system. Sample images are from the VTD database. The feature extraction part utilizes LBP histogram features. The matching score(5806.34) is computed from chi-square distance.

cross-spectral matching results are summarized in Table 1. TPLBP obtains the best Rank-1 identification rate (53.9%) in the scenario of MWIR vs. V matching. Such a single descriptor is even better than multiple descriptors used here. For the other scenario of V vs. MWIR matching, it has attained a second best accuracy of 32.7%. Experimental results indicate that cross-spectral matching is rather challenging when MWIR images are used. Similar observations have been reported in[31] where images acquired in the LWIR domain were used. Further, our experiments suggest the importance of illumination normalization for cross-spectral matching. Specifically, the SQI method was observed to positively impact recognition performance. Furthermore, the cross-spectral matching of MWIR and visible results with four main descriptors are illustrated in Fig. 9. As noted, the evaluation metric of CMC is employed here. Though the reported Rank-1 identification rate is a bit low, the Rank-5 identification rate is much higher, indicating the feasibility of such cross-spectral matching in practical applications. The results illustrated in Table 2 is same to Table 1, except for the TPLBP method. Here, we use a different parameter set for TPLBP, which produces higher accuracy in V vs. MWIR matching, up by around 10%. The fused result is calculated from the score level fusion of the four listed methods.

6. CONCLUSIONS

We presented a systematic performance analysis of various standard face recognition algorithms on visible and MWIR imagery. To facilitate our analysis, we first conducted a comprehensive data collection with a novel sensor system capable of acquiring MWIR images at a video frame-rate of 30 frames per second. The data collection effort (face images of subjects were acquired with intra-personal variability) was designed to investigate the hypothesis that MWIR imagery would yield high recognition performance.

Different scenarios were tested allowing us to gain some understanding of the shortcomings of the MWIR spectrum with respect to the visible modality. As expected, cross-spectral experiments resulted in markedly reduced performance, before and after fusion. On the other hand, identification performance on MWIR imagery

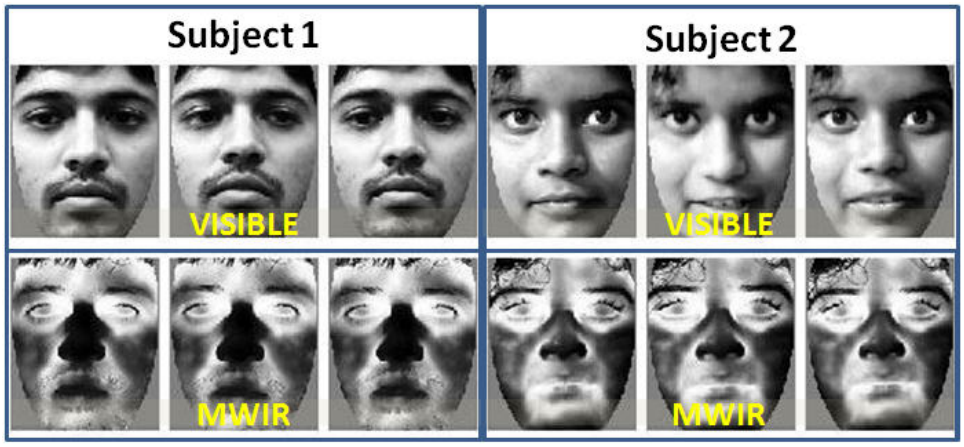
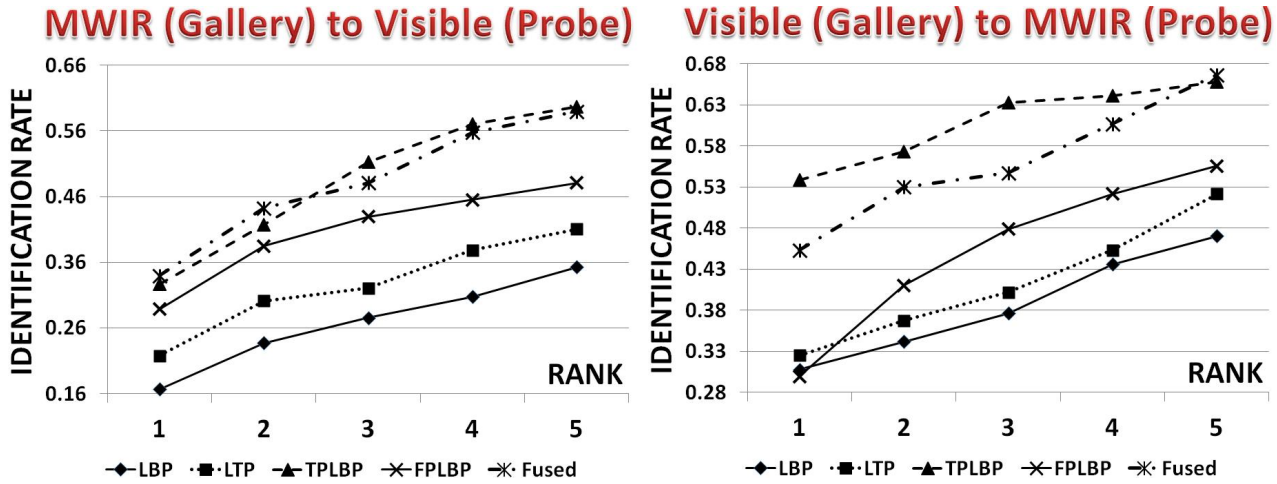


Figure 8. Sample images from the VTD database. Top row shows the visible images and the bottom row shows the corresponding MWIR images.



(a) MWIR (Gallery) vs. Visible (Probe)

(b) Visible (Gallery) vs. MWIR (Probe)

Figure 9. CMC curves when (a) MWIR images were used as gallery and visible images were used as probe, and (b) MWIR images were used as probe and MWIR images were used as gallery. The images indicate also the best performance that is achieved when using our proposed feature level fusion algorithm (TPLBP with DoG, and CHI).

appears to be comparable to that of visible imagery (baseline). Thus, it appears that MWIR modality holds great promise under reasonable operating conditions. Another benefit of using MWIR is the possibility of obtaining the same recognition rates when face images are acquired in complete darkness. However, further experiments and data collections will be necessary to investigate more challenging scenarios, e.g., when subjects are wearing glasses, or when images are acquired outdoors. We intend to expand our collection and analysis effort towards that direction.

The main disadvantage of thermal-based biometric identification at the present time is the high price of the majority of high end sensors. However, the cost of thermal security cameras has dropped considerably, and is now comparable to high end digital single-lens reflex (DSLR) cameras (visible band). For example, FLIR is now offering LWIR cameras starting at less than $3,000, making them more affordable and thus researchers can utilize them in several innovative ways. This scenario would have been inconceivable just two years ago.

ALGORITHM (NORM) [SM]	MWIR vs. V	V vs. MWIR
LBP (DoG)(CHI)	0.308	0.167
LBP (DoG)(DT)	0.197	0.180
LTP (DoG)(CHI)	0.325	0.218
LTP (DoG)(DT)	0.239	0.212
LBP+SIFT+PHOG (SQI) [L1]	0.376	0.186
LBP+SIFT+PHOG (SQI) [L2]	0.368	0.212
TPLBP (DoG)(CHI)	**0.539**	0.327
FPLBP (DoG)(CHI)	0.299	0.288
Fused (Score-level)	0.453	**0.340**

Table 1. Identification rates (Rank-1) for the cross-spectral scenarios: (i) MWIR vs. Visible, and (ii) Visible vs. MWIR. NORM=Normalization. SM=Similarity Measure.

ALGORITHM (NORM) [SM]	MWIR vs. V	V vs. MWIR
LBP (DoG)(CHI)	0.308	0.167
LTP (DoG)(CHI)	0.325	0.218
TPLBP (DoG)(CHI)	**0.436**	**0.4231**
FPLBP (DoG)(CHI)	0.299	0.288
Fused (Score-level)	0.436	0.372

Table 2. Identification rates (Rank-1) for the cross-spectral scenarios: (i) MWIR vs. Visible, and (ii) Visible vs. MWIR. NORM=Normalization. SM=Similarity Measure. For TPLBP, we use $r = 2$ in this case.

Acknowledgments

This work is sponsored through CITeR award number 1003702CR. The authors are grateful to all WVU faculty and students, and FLIR for their contributions to this work.

REFERENCES

[1] Bourlai, T., Kalka, N., Cao, D., Decann, B., Jafri, Z., Nicolo, F., Whitelam, C., Zuo, J., Adjeroh, D., Cukic, B., Dawson, J., Hornak, L., Ross, A., and Schmid, N. A., [*Ascertaining Human Identity in Night Environments*], 471–478, Distributed Video Sensor Networks, Springer (2011).

[2] Selinger, A. and Socolinsky, D. A., "Face recognition in the dark," in [*Proc. Conference on Computer Vision and Pattern Recognition Workshop*], 129–134, IEEE, Washington, DC, USA (June 2004).

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

[4] Socolinsky, D., Selinger, A., and J.Neuheisel, "Face recognition with visible and thermal imagery," *Computer Vision and Image Understanding* **91**, 72–114 (2003).

[5] Miller, J. L., [*Principles Of Infrared Technology: A Practical Guide to the State of the Art*], Springer (1994).

[6] Müller, M., Schreer, O., and Sáenz, M. L., "Real-time image processing and fusion for a new high-speed dual-band infrared camera," in [*Proc. of the SPIE*], 654310 (2007).

[7] Socolinsky, D., [*Handbook of Biometrics*], ch. Multispectral Face Recognition, 293–313, Springer (2008).

[8] Socolinsky, D. and Selinger, A., "Thermal face recognition in an operational scenario," in [*IEEE Computer Society Conference on Computer Vision and Pattern Recognition*], **2**, 1012–1019 (2004).

[9] Bourlai, T., Whitelam, C., and Kakadiaris, I., "Pupil detection under lighting and pose variations in the visible and active infrared bands," in [*IEEE Workshop on Information Forensics and Security*], (2011).

[10] Bourlai, T. and Jafri, Z., "Eye Detection in the Middle-Wave Infrared Spectrum: Towards Recognition in the Dark," in [*Proc. on Intl. Workshop on Information Forensics and Security*], IEEE, Foz do Iguaçu, Brazil (November 2011).

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

[12] Bolme, D. S., Beveridge, J. R., Teixeira, M. L., and Draper, B. A., "The CSU face identification evaluation system: Its purpose, features and structure," in [*Proc. International Conference on Vision Systems*], 304 – 311 (April 2003).

[13] Sirovich, L. and Kirby, M., "Application of the karhunen-loeve procedure for the characterization of human faces," *Transactions on Pattern Analysis and Machine Intelligence* **12**(1), 103–108 (1990).

[14] Turk, M. and Pentland, A., "Eigenfaces for recognition," *Journal of Cognitive Neuroscience* **3**(1), 71–86 (1991).

[15] Devijver, A. P. and Kittler, J., [*Pattern recognition: A statistical approach*], Prentice-Hall, Englewood Cliffs, N. J. (1982).
[16] Belhumeur, P., Hespanha, J., and Kriegman, D. J., "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *Transactions on Pattern Analysis and Machine Intelligence* **19**, 45–58 (1996).
[17] Teixeira, M., *The Bayesian Intrapersonal/Extrapersonal Classifier*, Master's thesis, Colorado State University, Fort Collins, Colorado 80523 U.S.A. (2003).
[18] Pietikinen, M., "Image analysis with local binary patterns," in [*Scandinavian Conference on Image Analysis*], 115 – 118 (June 2005).
[19] Tan, X. and Triggs, B., "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *Trans. Img. Proc.* **19**, 1635–1650 (June 2010).
[20] Štruc, V., "Inface: A toolbox for illumination invariant face recognition," (2009).
[21] Wang, H., Li, S. Z., Wang, Y., and Zhang, J., "Self quotient image for face recognition," in [*ICIP*], 1397–1400 (2004).
[22] Bosch, A., Zisserman, A., and Munoz, X., "Representing shape with a spatial pyramid kernel," in [*Proceedings of the 6th ACM international conference on Image and video retrieval*], *CIVR '07*, 401–408 (2007).
[23] Lowe, D. G., "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision* **60**, 91–110 (November 2004).
[24] Ahonen, T., Hadid, A., and Pietikainen, M., "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.* **28**, 2037–2041 (December 2006).
[25] Lazebnik, S., Schmid, C., and Ponce, J., "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in [*Proc. Computer Vision and Pattern Recognition*], 2169–2178, IEEE Computer Society (2006).
[26] Dalal, N. and Triggs, B., "Histograms of oriented gradients for human detection," in [*Proc. Computer Vision and Pattern Recognition*], 886–893, IEEE Computer Society (2005).
[27] Luo, J., Ma, Y., Takikawa, E., Lao, S., Kawade, M., and Lu, B.-L., "Person-specific sift features for face recognition," in [*Proc. Acoustics, Speech and Signal Processing*], 593596 (2007).
[28] Ross, A. and Govindarajan, R., "Feature level fusion using hand and face biometrics," in [*Proc. of SPIE Conference on Biometric Technology for Human Identification II*], 196–204 (2005).
[29] Wolf, L., Hassner, T., and Taigman, Y., "Descriptor based methods in the wild," in [*ECCV Workshop*], (October 2008).
[30] Socolinsky, D. A., Wolff, L. B., Neuheisel, J. D., and Eveland, C. K., "Illumination invariant face recognition using thermal infrared imagery," in [*Proc. Conference on Computer Vision and Pattern Recognition*], I–527 – I–534, IEEE, Kauai, HI, USA (2001).
[31] Klare, B. and Jain, A. K., "Heterogeneous face recognition using kernel prototype similarities," *Transactions on Pattern Analysis and Machine Intelligence (Under Review)* (2011).