

On Ear-based Human Identification in the Mid-Wave Infrared Spectrum

Ayman Abaza^a, and Thirimachos Bourlai^b

^a*West Virginia High Tech Foundation, 1000 Technology Drive, Fairmont, USA*

^b*West Virginia University, PO Box 6201, Morgantown, USA*

Abstract

In this paper the problem of human ear recognition in the Mid-Wave Infrared (MWIR) spectrum is studied in order to illustrate the advantages and limitations of the ear-based biometrics that can operate in day and night time environments. The main contributions of this work are two-fold: First, a dual-band database is assembled that consists of visible (baseline) and mid-wave IR left and right profile face images. Profile face images were collected using a high definition mid-wave IR camera that is capable of acquiring thermal imprints of human skin. Second, a fully automated, thermal imaging based, ear recognition system is proposed that is designed and developed to perform real-time human identification.

The proposed system tests several feature extraction methods, namely: (i) intensity-based such as Independent Component Analysis (ICA), Principal Component Analysis (PCA), and Linear Discriminant Analysis (LDA); (ii) shape-based such as Scale Invariant Feature Transform (SIFT); as well as (iii) texture-based such as Local Binary Patterns (LBP), and Local Ternary Patterns (LTP). Experimental results suggest that LTP (followed by LBP) yields the best performance ($Rank1 = 80.68\%$) on manually segmented ears and ($Rank1 = 68.18\%$) on ear images that are automatically detected and segmented. By fusing the matching scores obtained by LBP and LTP, the identification performance increases by about 5%. Although these results are promising, the outcomes of our study suggest that the design and development of automated ear-based recognition systems that can operate efficiently in the lower part of the passive IR spectrum is a very challenging task.

Keywords: Thermal Infrared Imaging, Thermal Ear Features, Score Level Fusion and Multi-Biometrics.

1. Introduction

Automated methods of recognizing an individual are based on measurable human body characteristics. However, different biological characteristics exhibit both strengths and weaknesses. For example, fingerprints and irises can result in high recognition rates but in order to be measured accurately they require controlled conditions and the subjects need to interact cooperatively with a device. In military and law enforcement applications where remote recognition is necessary, face and ear biometrics can be used for establishing human identity, because they have several advantages over other biometric traits: they are non-intrusive, understandable, and facial images can be captured in either cooperative or non-cooperative manner at variable standoff distances. Unlike human faces, ears are not affected by facial expression or by differences in background [1].

The ear biometric has certain advantages over other biometrics: ears are relatively static in size and structure over each individual's life [2]. One of the concerns when using the ear biometric is that there is no longitudinal study (to our knowledge at this point in time) that shows that ear shape is static. The forensic science literature though reports that ear growth, after the first four months of birth, is highly linear [2]. The rate of stretching is approximately five times greater than normal during the period from four months to the age of eight; after which it is constant until around the age of seventy when it again increases. To the best of our knowledge, there is only one biometric experiment on the effect of aging on ear recognition by Ibrahim et al. [3]. This experiment shows that ear recognition rate is marginally affected over the eleven months period of data collection and the authors concluded that "ear can be used in various applications as a time invariant biometric".

A typical 2-D ear biometric recognition system has three modules:

1. *Detection (segmentation)*: The ear region is localized in a given 2D image and then segmented for further processing.
2. *Feature Extraction*: Ear-based features are extracted using certain attributes. Some of these features required ear alignment such as Iannarelli's geometric distances; while others do not require alignment such as SIFT features. Iannarelli [2] manually aligned the ear images; Other methods extracted the external ear contour (edges) and used the two extreme, having maximum in-between distance, points to align the ear image [4, 5].

3. *Matching*: In this module the features extracted by a query ear are compared against those extracted by multiple ears in the enrollment (gallery) database to associate the identity of the query ear to one of those in the database. The features extracted from each ear sample of the gallery dataset are matched against those extracted from the samples of the probe dataset.

While most of the current ear-based recognition approaches are mainly focused in performing feature extraction on data acquired in the visible band [6], in this work we perform a set of experiments and investigate the efficiency of several feature extraction approaches on profile mid-wave infrared face images. Figure 1 illustrates why designing mid-wave IR based ear recognition techniques is important.

The MWIR band has both reflective and emissive properties and it can operate in both day and night time environments. Other advantages of MWIR over the active IR band, i.e. Near IR or Short-wave IR (see fig.2) are: (i) MWIR imagery can be acquired without any external illumination in day or night environments (regions in the active IR band might require an external light source), (ii) anatomical features not observable in the active IR spectrum can be observable in MWIR [7], and (iii) background clutter in MWIR images is not always visible.

The above observations as well as the potential future applications in law enforcement and the military, were the main motivation for our work. To the best of our knowledge, this paper represents the first attempt in the open literature to investigate the problem of ear recognition in a specific range of the passive infrared (thermal) band.

1.1. Most Related Work in the Thermal Band

The importance of using the Middle-Wave Infrared (MWIR) band for human recognition was first discussed in [8, 9]. In [8], the authors studied the problem of eye detection in the MWIR spectrum and discuss the effect of eye detection, one of the fundamental step in most face recognition systems, on face recognition. The authors showed that MWIR face images can efficiently be matched to MWIR face images (same session). They also suggested that cross-session (i.e. outdoors vs. indoors) matching in the MWIR band can significantly degrade recognition performance (in terms of rank-1 scores). In [9], the authors also discussed the challenges associated when designing face identification systems that operate in the MWIR band. The

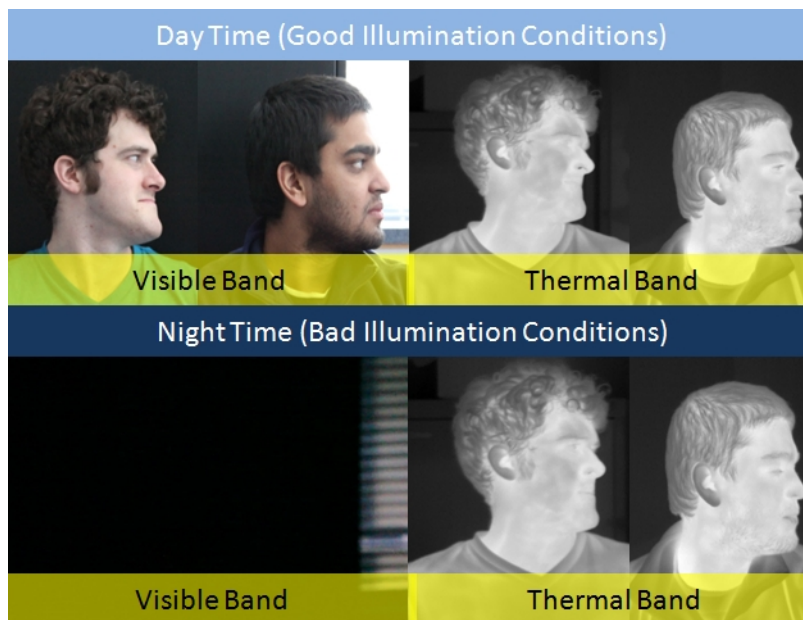


Figure 1: Sample images of two subjects acquired in day time and night time conditions, using a visible and a thermal camera.

authors illustrated the advantages and limitations of intra-spectral (MWIR to MWIR) matching, while they also revealed how challenging the problem of cross-spectral (MWIR to visible or vice versa) matching is.

In [10], the authors proposed a statistically-based algorithm for physiological feature extraction (namely wrinkles, veins, edges, and perimeters of facial characteristics). Fiducial (that is, reference) points are subsequently detected either manually or automatically. For that purpose, they used three different extractors, i.e., a fingerprint-based minutiae detector, the scale-invariant feature transform (SIFT), and the speeded up robust features (SURF) detector. Finally, they matched faces using an alignment-based matching algorithm with the ability of finding the correspondences between a stored set of gallery points and an input set of probe points. A summary of MWIR-based algorithms, methods and experiments can be found in [7].

While all the aforementioned MWIR-based approaches focused in the area of face recognition, there is no reported work in the open literature regarding ear recognition in MWIR band in particular, or thermal IR band in general.

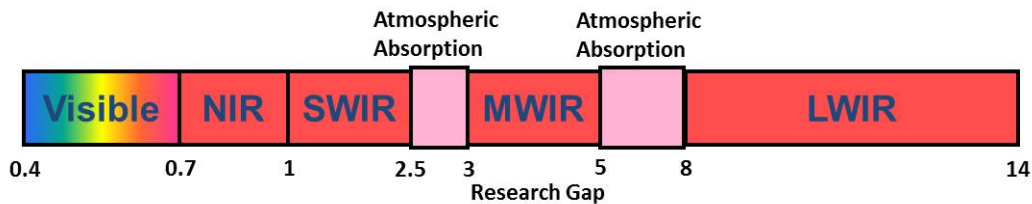


Figure 2: The electromagnetic spectrum illustrates various sub-bands from Visible to InfraRed. The wavelengths are in μm .

1.2. Drawbacks of thermal images

Though thermal images have found to be useful in face recognition [8, 9], there are some limitations for thermal band [11].

- **Health or Physical Related Conditions:** There are various conditions under which the thermal characteristics for human skin (e.g. face regions) can be altered. These can be generally categorized into health variations (e.g. high fever, consumption of alcohol etc.), or physical variations related to stress, emotional state or physical activity. Since human ears (as well as the nose and eyebrows) are more exposed to external air flow from the environment, they are generally cooler than other parts of the human face and less affected by changes of human physiology [12]. Figure 3 shows examples of ear images that appear darker than areas of visible (exposed) blood vessels (e.g. superficial temporal artery [13]).
- **Occlusion:** The usage of glasses or human hair blocks most of the thermal energy emitted for human skin (including faces or ears).
- **Thermal Sensitivity:** thermal sensitivity of the MWIR cameras is important 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.
- **Cost:** High-end sensors are very expensive. 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 US dollars, making them more affordable and thus researchers can utilize them in several innovative ways.

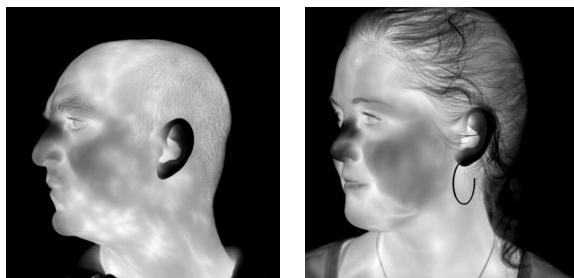


Figure 3: Examples of ear images that appear darker in MWIR band.

1.3. Our Approach

In this work, we designed and developed an ear-based recognition system that works in the MWIR band. The main modules of the system are ear detection, feature extraction and matching. Our ear detection approach is based on a new cascaded AdaBoost framework that is more computationally efficient when compared to the baseline one [14]. Then we investigate several feature extraction methods using MWIR left and right profile face images. Finally, a set of experiments is conducted to determine the effect of the aforementioned design steps in recognition performance:

- The *first* set of experiments investigate the recognition accuracy of intensity-based and shape-based feature extraction approaches;
- The *second* set of experiments use automated ear detection [14], then investigate the same feature extraction approaches;
- The *third* set of experiments evaluate the effect of combining several features on recognition performance (see Fig. 4)

The methods studied in this work are evaluated using the WVU Visible-Thermal Profile Face (VTPF) database that consists of face images of more than 80 subjects and that was captured on two different sessions. The database was assembled indoors spanning over a time period of about 3 months, with at least 2 weeks between data re-acquisition for all participants.

1.4. Paper Organization

The rest of the paper is organized as follows: Section 2 describes the hardware used in data collection and the acquired database. Section 3 describes the ear detection technique. Section 4 presents several methods for

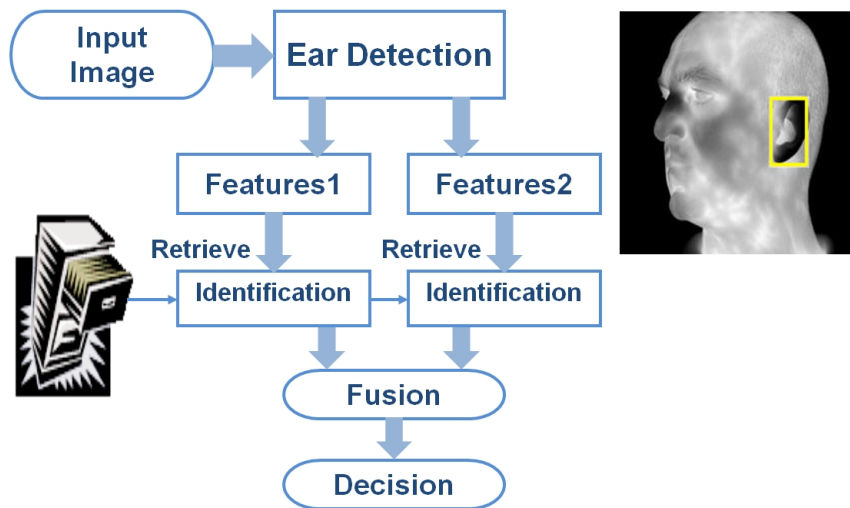


Figure 4: Block diagram of the proposed thermal ear recognition system.

feature extraction, namely: intensity-based such as ICA, PCA, and LDA; shape-based such as SIFT; and texture-based such as LBP, and LTP. Section 5 describes various experiments to evaluate the proposed technique, and Section 6 provides concluding remarks, and sketches our plans for future work.

2. Thermal Imagery

The thermal camera used in this work is a high definition Middle-Wave Infrared (MWIR) camera produced by FLIR Systems¹. It is capable of acquiring thermal imprints of human skin and analyzing the thermal distributions and temporal variations. 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 $3\text{-}5\mu\text{m}$, 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.

¹“FLIR Systems,” <http://www.flir.com>, 2011.

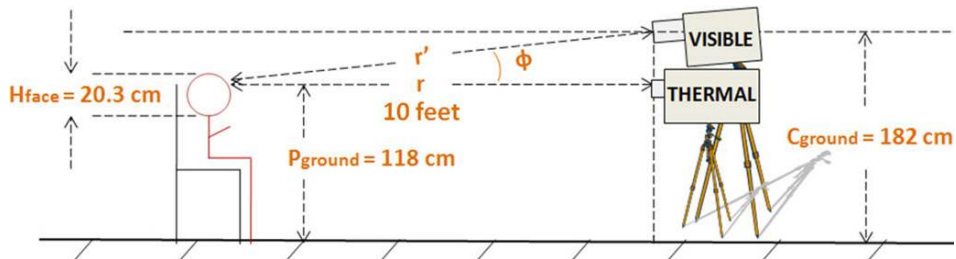


Figure 5: Block diagram of the live subject-capture setup using the visible and MWIR cameras.

The live face capture configuration we used is illustrated in Fig. 5. The camera features a high resolution 1024x1024 Indium Antimonide (InSb) Focal Plane Array (FPA) achieving mega-pixel image resolution in a single thermal image. The distance between the subject’s ear and the camera was set to about 6.5 feet and thus, the size of an average ear region was 73x48 pixels. 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. We had access for a top of the line FLIR MWIR camera for a very short period of time, and that is why we only managed to collect about 88 subjects, out of which about 57 subjects participated in two sessions. One of the issues though is that the data collection was not originally designed to address the problem of ear detection (but face recognition) and thus, the available profile images were not always usable for the problems of ear detection and recognition.

2.1. Issues with Acquiring MWIR Face Images

While the target was to collect only full profile images ($+/- 90^\circ$ head yaw) some subjects did not follow the exact data collection protocol. This resulted in having some sample images with head yaws (y-axis) that range from $+/- 10^\circ$ to $+/- 90^\circ$. Thus, we had to deal with a more challenging problem (i.e., detecting ears under variable poses), since having only full profile data would result in better detection rates.

Another issue is the thermal sensitivity of the camera that is an important factor since it impacts image quality, i.e., the lower the sensitivity, the more accurate the camera can be in order to produce higher quality images. The camera that we are using has the highest sensitivity in the market. The

associated software of the camera was used to perform regular calibration (before acquired each set of thermal profile images), which ensures that the camera operates to its optimum performance, and as a result it guarantees measurement accuracy and reliability. The camera operated in a controlled, low temperature environment (room temperature) where better thermal sensitivity can be exploited since the thermal contrast (temperature delta within an image) is very low. An additional software tool (provided by FLIR) was used to remove noise (e.g, dead pixels) from face images and control the temperature scale limits (e.g., setting the temperature range from 28° to 41° Celsius that is the typical range of human body temperature) during data collection.

3. Ear Detection System

The ear detection procedure classifies images based on the value of rectangular features. These features encode ad-hoc domain knowledge, and can work faster than pixel-based ones [15]. The original approach to detection (baseline) has been widely used to solve the problem of face detection (e.g. in the work of Viola-Jones). A modified version of the baseline approach was used to the ear modality in 2008 [16].

For our study and in order to have accurate ear detection results using MWIR profile face images, we re-designed the methodological steps of the original ear detection procedure, and thus, managed to reduce the learning time. Hence, the required training time was reduced from several weeks (using the original Viola-Jones method) to several hours (using our proposed approach) [14].

In the ear detection method [14], each rectangle feature f represents the main component of the weak classifier $h(x, f, p, \phi)$, where ϕ is a threshold, and p is the polarity indicating the direction of the inequality:

$$h(x, f, p, \phi) = \begin{cases} 1 & \text{if } p \cdot f(x) < p \cdot \phi \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The learner is called a weak classifier due to its low performance. Each ensemble classifier consists of a set of T weak classifiers, where θ is the threshold of the ensemble (strong) classifier:

$$H(x) = \begin{cases} \text{Continue} & \sum_{t=1}^T \alpha_t \cdot h_t(x) > \theta \\ \text{Reject} & \text{otherwise} \end{cases} \quad (2)$$

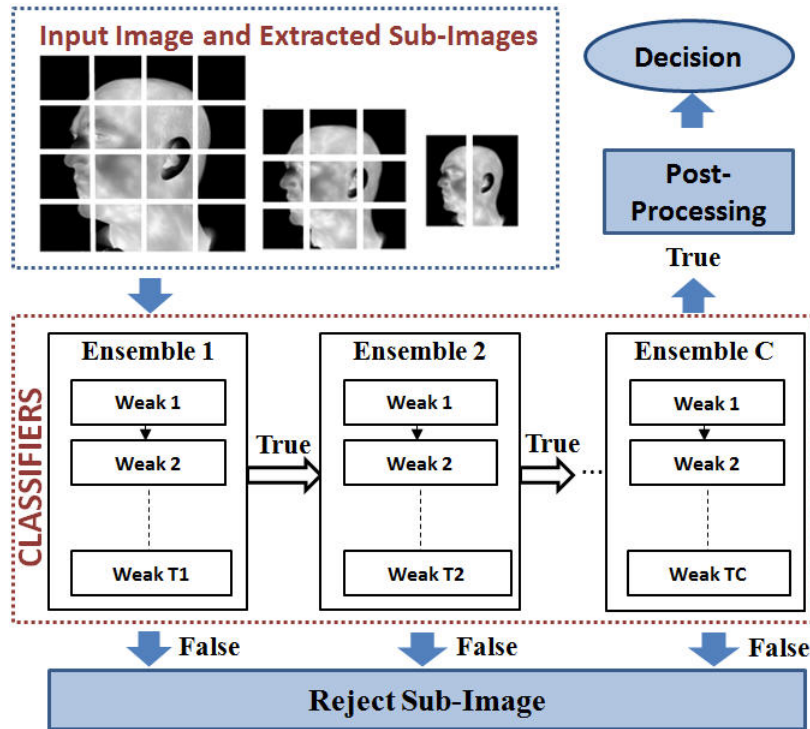


Figure 6: Overview of the proposed ear detection approach: An input thermal image is scaled multiple times. For each scale a set of all possible 24×16 pixel sub-images (including overlapping ones) is extracted. Then each sub-image is evaluated through a cascaded AdaBoost classifier, after which a decision about the location of ears (when detected) is made.

These ensemble classifiers have a cascade arrangement forming the ear detection system (see Fig. 6).

In order to detect an ear in an input image, the image is scanned using the proposed cascaded AdaBoost system. This means that the input image is divided into overlapped sub-images of 24×16 ², and each region is evaluated. Then, the image is scaled down by a factor $s = 0.8$ and the process is repeated. At this post-processing stage, all the detected regions at various levels of the scale pyramid are scaled back to the original image resolution. Finally, the overlapped detected regions are combined into one rectangle

²The average ear size is 73×48 pixels.

region.

4. Feature Extraction

In this section, we examine various ear recognition algorithms proposed in the visible domain and attempt to apply these techniques in the MWIR domain. These feature extraction methods are: intensity-based such as (ICA)[17], (PCA) [18], and (LDA)[19]; shape-based such as (SIFT); and texture-based such as (LBP), and (LTP). Since ICA, PCA, and LDA are well known algorithms, we do not provide more information on them. What follows is a description of SIFT, LBP and LTP feature extraction methods.

4.1. Extraction of Ear-based Features using SIFT

The scale invariant feature transform (SIFT) is a shape based feature extraction method for extracting highly distinctive invariant features. SIFT algorithm [20], consists of four major stages:

scale-space extreme detection: the Difference-Of-Gaussian (DOG) function is applied to an ear image to identify candidate points that are invariant to scale and orientation.

key point localization: this step rejects points having low contrast (sensitive to noise) or are poorly localized along an edge.

orientation assignment: a gradient orientation histogram is computed in the neighborhood of each key-point, where histogram peaks correspond to dominant orientations.

key point descriptor: for each selected key-point orientation, the feature descriptor is computed as a set of orientation histograms.

4.2. Extraction of Ear-based Features using LBP

LBP operators are one of the best performing texture descriptors [21]. The basic LBP operator assigns a decimal value to each pixel in the image by thresholding (P) neighbor pixels at distance (R), as shown in fig. 7. The histogram (H) of these decimal values represents the feature vector.

Ojala et al.[21] called a local binary pattern uniform, if it contains at most two bitwise transitions from 0 to 1 or vice versa when the bit pattern is considered circular. This is considered as a feature selection method that

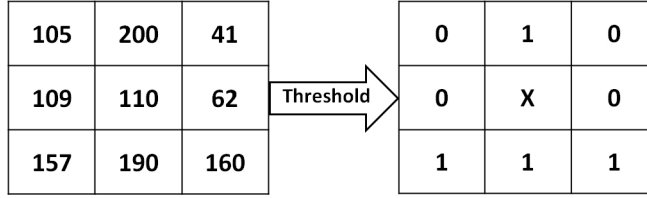


Figure 7: Basic LBP operator: (1) For a given input image pixel and its 8 neighbors, (2) Each neighbor pixel greater than or equal to the center pixel is assigned 1 otherwise it is assigned 0, (3) These binary values are arranged to form a binary number (01110010), which is transferred to a decimal equivalent (114).

reduced the number of features, and hence reduce the processing time. Experimental results, using 180 pre-processed MWIR ear images, showed that:

- 93.55% of the patterns in the 8,1 neighborhood are uniform pixels,
- 89.57% of the patterns in the 8,2 neighborhood are uniform pixels,
- 84.16% of the patterns in the 16,2 neighborhood are uniform pixels,

Hence, we decided to use the uniform pattern only.

For the block based division, the image is divided into N blocks. These blocks can be of arbitrary size and can overlap. The LBP operator is applied to each block separately, and their corresponding histograms $H = [h_1 h_2 \dots h_N]$ are calculated. Integration of these blocks can be at the feature level, by concatenating the histograms extracted from various blocks, then the overall histogram H is used for matching. One of the main advantage of dividing the image into sub-blocks is that these blocks are expected to be more discriminative than using the whole image.

4.3. Extraction of Ear-based Features using LTP

Local Ternary Patterns (LTP) operators extends LBP to 3-valued codes [22], in which gray levels in a zone of width $\pm t$ around centered value are quantized to zero, ones above this are quantized to +1 and ones below it to -1, i.e. the indicator $s(u)$ is replaced by a 3-valued, as shown in fig. 8. This trinary number is transferred into two binary numbers. In other words, two separate channels of LBP descriptors are formed. Two histograms of these two decimal values are concatenated to form the feature vector.

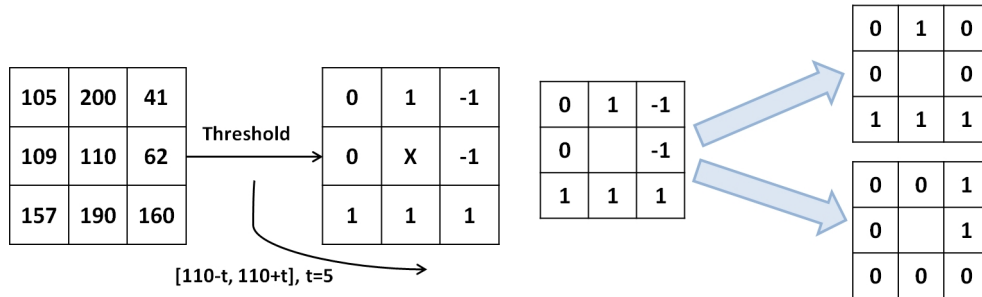


Figure 8: Basic LTP operator: (1) For a given input image pixel and its 8 neighbors, (2) Each neighbor pixel greater than the center pixel plus t is assigned 1, or less than the center pixel minus t is assigned to -1, otherwise it is assigned 0, (3) These trinary values are arranged to form a trinary number (01“-1”“-1”1110), (4) This trinary number is transferred into two binary numbers (01001110, 00110000), which are then transferred into two decimal equivalents (92, 48).

5. Experimental Results

In this section, we present various experiments to evaluate:

- The effect on ear recognition performance when using several feature extraction techniques in the MWIR band.
- The effect of score-level fusion on ear recognition performance.

We start this section by describing the database, the training phase, and then, we demonstrate various experiments to prove the proposed concept of ear recognition of ear recognition in the MWIR band.

5.1. Database

The aforementioned thermal system (see Section 2) was employed for data collection in order to assemble, as part of a pilot study, the Thermal Profile Face (TPF) database (see Fig. 9). The standoff distance was set to 6.5 feet. The thermal-based database used was assembled indoors spanning over a time period of about 3 months, with at least 2 weeks between data re-acquisition for all participants.

For testing, 57 subjects (40 males + 17 females) were used in this experiment. Due to some technical reasons like complete hair occlusion and severe head profile angle, 12 subjects were excluded from one or more session; hence

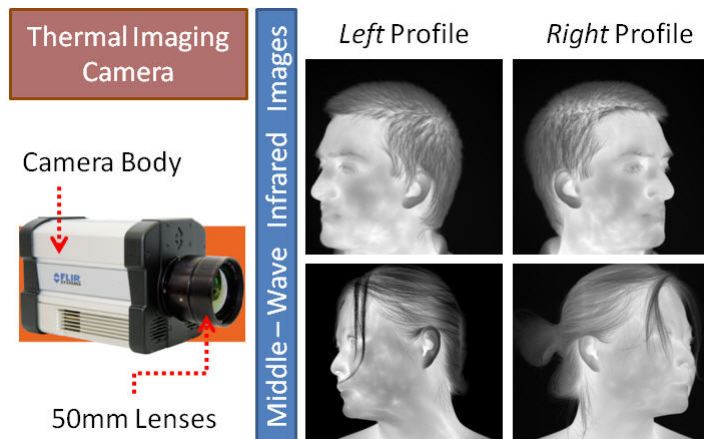


Figure 9: MWIR camera setup and acquired sample profile face images. Note that originally the camera acquires video frames where each pixel has a temperature value ranging from ~ 28 -41 degrees Celsius. Then, these frames are converted to gray-scale images for display purposes.

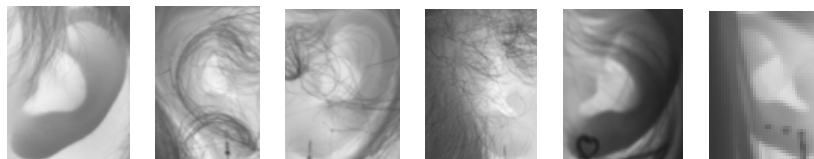


Figure 10: Examples of challenging ear occlusion that we had to use in testing.

excluded from the testing. In the testing phase, we end up using 45 subjects, 4 images per subjects (2 left, and 2 right images). An overview of the modes and numbers of subjects and images are provided in Table 1. More than 25% of these images are still having hair occlusion as shown in fig. 10.

For training, we used 31 (23 males + 8 females) subjects, 12 images per subjects (6 left, and 6 right images). Due to hair occlusion, 1 subject was excluded. Then for the remaining images, the ear regions were manually segmented.

5.2. Training Phase and Parameter Tuning

Using the training data set, we set up several identification experiments to choose some parameters so as to optimize the performance of several feature extraction methods. The first experiment was to choose the best LBP operator ($(LBP_{8,1}^U, LBP_{8,2}^U, \text{or } LBP_{16,2}^U)$), LBP optimum block size, as well as

Table 1: Details of the database used for the experiments

Mode	Subjects	Left (images)	Right (images)	Total (images)
Training	30	180	180	360
Testing Galleries	45	45	45	90
Testing Probes	45	45	45	90

Table 2: Identification rate (R1:Rank1) for several LBP and LTP operators, after dividing the image into various blocks

Operators	left Set (R1%)	Right Set (R1%)	Both (R1%)
$LBP_{8,1}$ 3x3	92.90	99.33	96.07
$LBP_{8,2}$ 3x3	96.77	99.33	98.03
$LBP_{16,2}$ 3x3	95.48	99.33	97.38
$LBP_{8,1}$ 5x5	92.90	99.33	96.07
$LBP_{8,2}$ 5x5	94.19	98.67	96.39
$LBP_{16,2}$ 5x5	94.19	98.00	96.07
LTP 3x3	95.70	98.89	97.27
LTP 5x5	93.55	98.89	96.17
LTP 7x7	94.62	98.89	96.72

LTP optimum block size. Details of this experiment is shown in table 2. For LBP, result show that the $LBP_{8,2}^U$ operator achieves the best performance, when dividing the images into $3x3$ blocks. For LTP, results consistently show that dividing the images into $3x3$ blocks yields the optimum performance. Please note the high performance return to the fact that the training data came from one session, so variation is only due to various poses. We trained the intensity based techniques via two methods. The first method is class specific, as left and right ears will have unique principle components. The second method is class general, as the left and right ears will have one general principle component (Both).

5.3. Performance of various ear features

In the *first experiment*, using the test dataset, we evaluated several feature extraction methods in terms of their recognition rate, and computational complexity, i.e. processing time to process (extract features) a query profile face image. This experiment shows LTP to yield the best performance, and LBP to yield the second best (details in table 3).

Table 3: Identification rate (R1:Rank1) for several ear features; Average processing time (ms); (ICA), (PCA), (LDA), (SIFT), (LBP), (LTP).

Operators	Left Set (R1%)	Right Set (R1%)	Both (R1%)	Time (ms)
SIFT	32.56	37.78	34.09	357
LBP	76.74	64.44	70.45	151
LTP	81.40	80.00	80.68	167
PCA	39.53	40.00	42.05	19
ICA	58.14	55.56	61.36	25
LDA	41.86	33.33	30.68	20
With automated detection [14]				
SIFT	30.23	35.56	32.95	
LBP	60.47	55.56	57.95	
LTP	72.09	64.44	68.18	
PCA	20.93	26.67	23.86	
ICA	30.23	42.22	35.23	
LDA	16.28	24.44	22.73	

In the *second experiment*, we use an automated ear detection technique [14], then we re-evaluated the above mentioned feature extraction methods (details in table 3, and fig. 11). This process required on average 260 ms. Experimentally LTP was proven to yield the best performance, while LBP to be the second best.

An interesting observation comes for Table 3. There we see that recognition of left ears has better accuracy than right ears. Based on our empirical evaluation, we determined that this error was due to pose (as shown in Fig. 13). We return this to the experimental setup, where the subjects first move their head to the right (exposing left ears more carefully), and then to the left. As the data collection was not designed originally for addressing the problem of ear recognition, the subjects and data collection operator were not careful in achieving perfect profile face images.

In the *third experiment*, we evaluate several fusion schemes of the best features extraction methods. We fused the best two techniques namely (LTP) and (LBP) at the score level, using several fusion rules. Details of this experiment are shown in Table 4. We found that: (i) this fusion does not help in case of using manually detected ears. We return this to the available data that carries hair occlusion for more than 25%; (ii) using a suitable fusion rule, namely the SUM or PROD, the performance, using automated detection, is

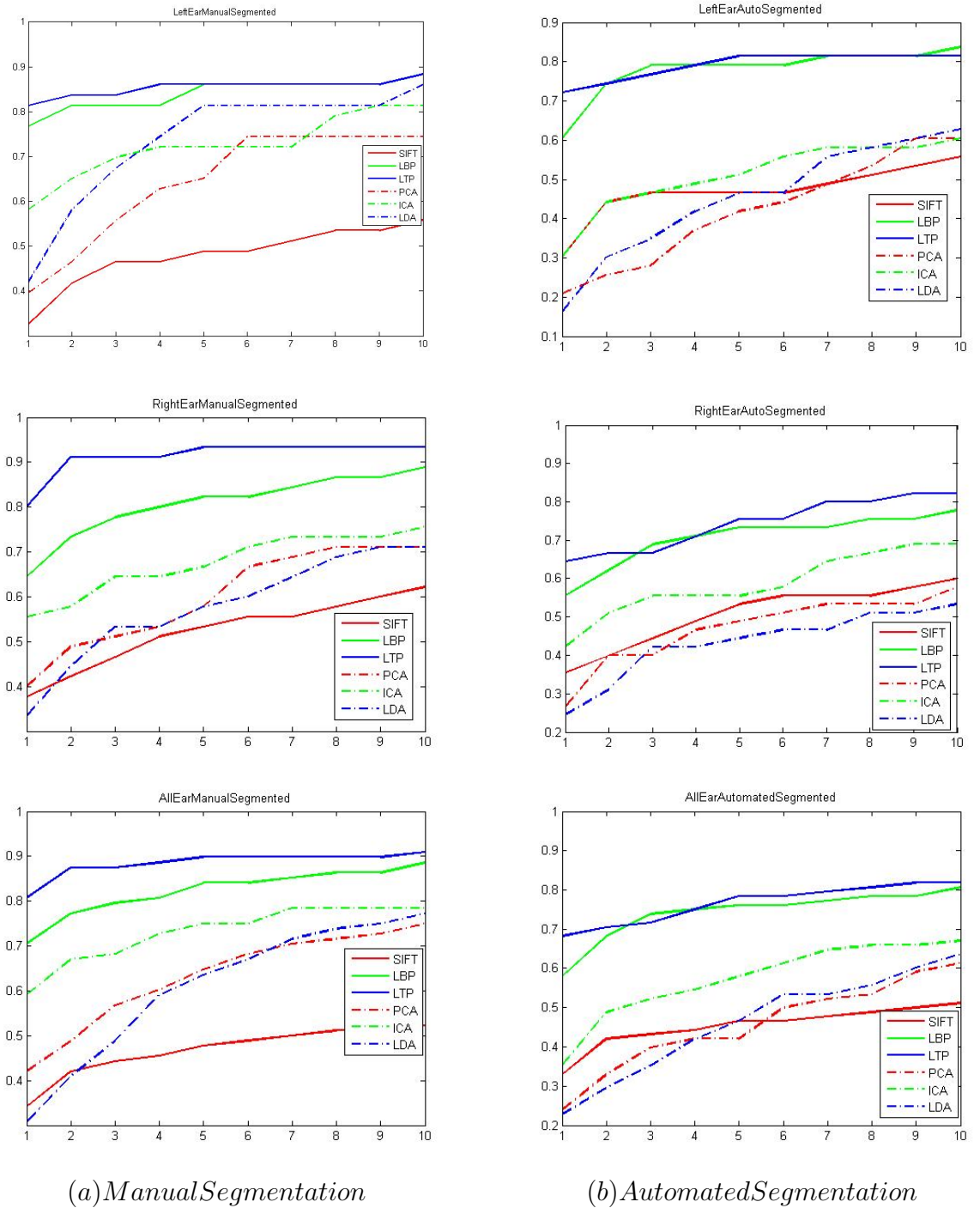
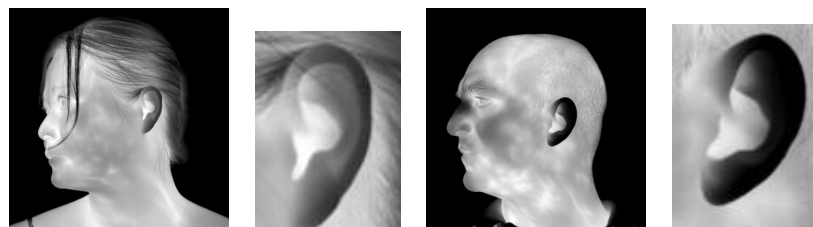
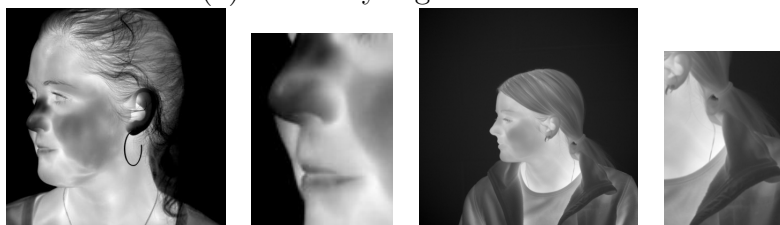


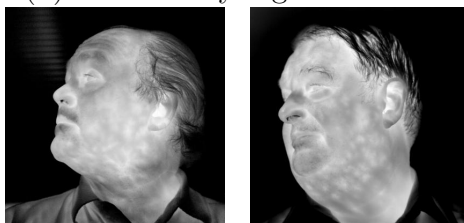
Figure 11: Cumulative Match Curves (CMC) for various feature extraction methods using: (a) manually segmented thermal ear images; (b) automatically segmented thermal ear images.



(a) Correctly segmented ears

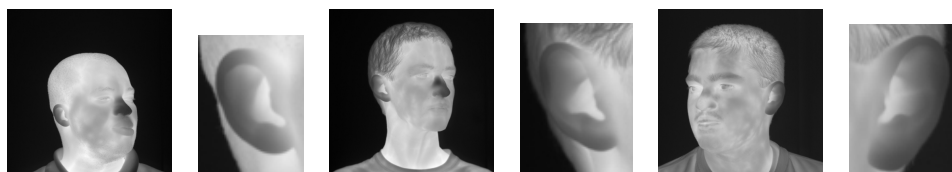


(b) Incorrectly segmented ears

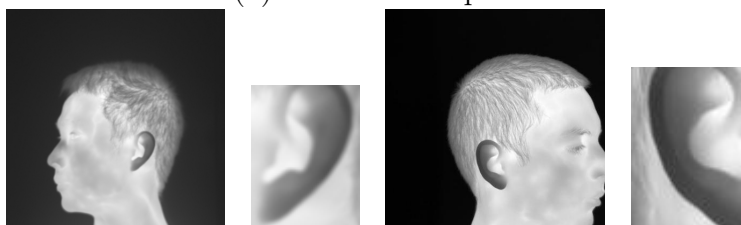


(c) Missed ears

Figure 12: Automated Thermal Ear Detection.



(a) Error due to pose



(b) Under segmentation error

Figure 13: Analysis of Ear Identification Error.

Table 4: Identification rate (R1:Rank1) using several fusion schemes at the score level, namely SUM:average rule, MIN: minimum rule, MAX: maximum rule, and PROD: production rule.

Fusion Schemes	Left Set (R1%)	Right Set (R1%)	Both (R1%)
LTP rule	81.40	80.00	80.68
LBP rule	76.74	64.44	70.45
SUM rule	79.07	77.78	78.41
MIN rule	81.40	68.89	75.00
MAX rule	76.74	80.00	79.55
PROD rule	79.07	75.56	77.27
With automated detection [14]			
LTP rule	72.09	64.44	68.18
LBP rule	60.47	55.56	57.95
SUM rule	76.74	68.89	72.73
MIN rule	69.77	60.00	60.23
MAX rule	74.42	66.67	68.18
PROD rule	76.74	68.89	72.73

improved by about 5%.

6. Conclusions and Future Work

This paper has presented a study on the problem of ear recognition using thermal profile face images (TPF) database with head yaws (y-axis) angle that range from $+/- 50^\circ$ to $+/- 90^\circ$. Experimental results show that: (i) our proposed ear recognition method can operate on MWIR thermal band with promising results; (ii) LTP was proven to yield the best performance 80.68% using manually segmented ear regions, and 68.18% using automated ear detection regions; and (iii) The proposed fusion method yielded 72.73% recognition accuracy using automated ear detection regions.

Although our approach is relatively successful in identifying ears under various rotation angles and using relatively sizable database, it would be desirable to: (i) scale up the database, and (ii) enhance system performance under variable angles. Another extension of our work would be to test cross-band (thermal and visible) reliable ear features.

7. Acknowledgments

This work was sponsored through CITEr award number 1003702CR to West Virginia University. The authors are grateful to FLIR Thermal Imaging for providing us a demo MWIR camera to perform our pilot data collection and to all faculty and students at West Virginia University for their contributions. Special thanks to Mr. Nnamdi Osia for his valuable participation and assistance with data collection.

8. References

- [1] A. Jain, P. Flynn, A. Ross, Handbook of Biometrics, Springer, 2007.
- [2] A. Iannarelli, Ear Identification, Forensic Identification Series, Paramount Publishing Company, USA, 1989.
- [3] M. Ibrahim, M. Nixon, S. Mahmoodi, The effect of time on ear biometrics, in: International Joint Conference on Biometrics (IJCB 2011), Washington-DC, USA.
- [4] L. Yuan, Z. chun Mu, Ear recognition based on 2D images, in: First IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS 2007), Washington-DC, USA.
- [5] A. Kumar, C. Wu, Automated human identification using ear imaging, Pattern Recognition 45 (2012) 956–968.
- [6] A. Abaza, A. Ross, C. Herbert, M. Harrison, M. Nixon, A survey on ear biometrics, 2013. To appear on ACM Computer Survey.
- [7] T. Bourlai, Mid-wave ir face recognition systems, SPIE Newsroom Magazine - Defense & Security (2013).
- [8] T. Bourlai, Z. Jafri, Eye detection in the middle-wave infrared spectrum: Towards recognition in the dark, in: International Workshop on Information Forensics and Security (2011), Foz do Iguacu, Brazil.
- [9] T. Bourlai, A. Ross, C. Chen, L. Hornak, A study on using middle-wave infrared images for face recognition, in: SPIE Conference on Defense, Security, and Sensing (2012), Baltimore-MD, USA.

- [10] N. Osia, T. Bourlai, Holistic and partial face recognition in the mwir band using manual and automatic detection of face-based features, in: IEEE International Conference on Technologies for Homeland Security, Waltham-MA, USA.
- [11] L. Wolff, D. Socolinsky, C. Eveland, Face recognition in the thermal infrared, in: B. Bhanu, I. Pavlidis (Eds.), Computer Vision Beyond the Visible Spectrum, Advances in Pattern Recognition, Springer London, 2005, pp. 167–191.
- [12] D. Berry, R. Yemm, Variations in skin temperature of the face in normal subjects and in patients with mandibular dysfunction, British Journal of Oral Surgery 8 (1970) 242 – 247.
- [13] T. Bourlai, P. Buddharaju, I. Pavlidis, B. Bass, On enhancing cardiac pulse measurements through thermal imaging, in: International Conference on Information Technology and Applications in Biomedicine (ITAB 2009), Larnaca, Cyprus.
- [14] A. Abaza, T. Bourlai, Human ear detection in the thermal infrared spectrum, in: SPIE Conference on Defense, Security, and Sensing (2012), Baltimore-MD, USA.
- [15] P. Viola, M. Jones, Robust real-time face detection, Journal of Computer Vision 57 (2004) 137 – 154.
- [16] S. Islam, M. Bennamoun, R. Davies, Fast and fully automatic ear detection using cascaded adaboost, in: IEEE Workshop on Applications of Computer Vision (2008), Copper Mountain-CO, USA.
- [17] M. B. J. Movellan, T. Sejnowski, Face recognition by independent component analysis, IEEE Transactions on Neural Networks (NN) 13 (2002) 1450 – 1464.
- [18] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3 (1991) 71–86.
- [19] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman, Eigenfaces vs. fisherfaces: Recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 19 (1997) 711 – 720.

- [20] D. G. Lowe, Object recognition from local scale-invariant features, in: International Conference on Computer Vision (ICCV 1999), Washington-DC, USA, pp. 1150 – 1157.
- [21] T. Ojala, A. Hadid, M. Pietikainen, Face description with local binary patterns: application to face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 28 (2006) 2037 – 2041.
- [22] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, in: IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2007), Rio de Janeiro, Brazil, pp. 168 – 182.