Thermal and Visible Image Fusion for Ear Recognition

Syed Mohd Zahid Syed Zainal Ariffin[1], Nursuriati Jamil[1], Puteri Norhashimah Megat Abdul Rahman[2]

[1]*Digital Image, Audio and Speech Technology Research Group, Faculty of Computer and Mathematical Sciences,*
Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia.
[2]*Universiti Teknologi MARA, Tapah Campus, Perak, Malaysia*
liza@tmsk.uitm.edu.my

Abstract—**This paper discusses the possibility of fusing thermal and visible images to improve ear recognition ability in images of varying illuminations. Since thermal image is known to remain invariant to lighting changes and visible images are able to capture feature details, image fusion is proposed to accentuate the strengths of both spectra for ear recognition. Two popular image fusion techniques, weighted average (WA) and discrete wavelet transform (DWT) are used as a preliminary investigation. Eigenvectors are extracted from the fused image and recognition is performed using metric distance measure. With 67.5% recognition rate, DWT fused images performed better than WA fused images (63.75%). Thermal images, on the other hand, achieved 68.75% recognition rate. Even though thermal images performed slightly better than DWT fused images by 1.25%, the difference is deemed as insignificance due to the small dataset used and the primitive fusion rules employed. Further studies on the fusion techniques need to be done to improve fusion method.**

Index Terms—**Image Fusion; Ear Recognition; Ear Biometrics.**

I. Introduction

Research in ear recognition or ear biometrics has slowly gained attention in the past decade. Experiments by Iannarelli [1] proved the uniqueness of human ears. Human ears are relatively static in size and structure throughout an individual's life. Ear images can be obtained remotely without any direct contact with the sensor compared to thumbprint and iris pattern. Compared to fingerprint, ears are organ that are passively used in daily life, therefore is less prone to injury. In addition, it is also unaffected by facial expression which is one issue in facial recognition. The uniqueness of ears and its advantages make the ear as an alternative candidate in biometrics.

Similar to face recognition, there are three main problems that affect the performance of ear recognition which are pose variation, occlusions and illumination variation. This paper focuses on illumination variation problem. Currently, regular digital cameras are used to acquire images for ear recognition. The images acquired using this camera is called visible images as the acquisition is done based on light reflections in visible band of electromagnetic spectrum. Visible images represent the images better in well-lit environment [2]. However, in bad or dark illumination condition, the image's quality drop significantly, thus the ear features cannot be adequately represented.

Thermal images, on the other hand, are generated based on surface temperature of an object. The sensors detect the infrared (IR) radiation emitted from an object and the surrounding. Thermal images are not affected by illumination variation as it is based on temperature [3]. However, medical condition (e.g. fever) which risen the body temperature may affect the thermal images generated [4]. Figure 1 shows examples of images taken in good and bad illumination conditions using digital and thermal cameras.
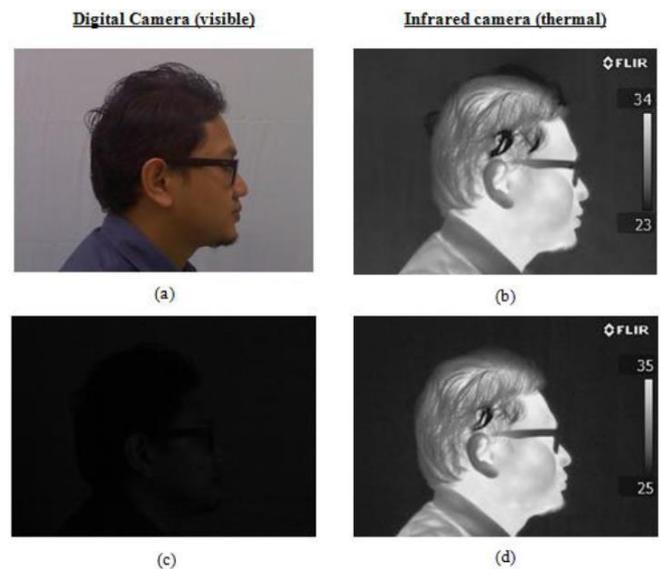
Figure 1: Examples of images taken in good illumination: (a) and (b), and image taken in bad illumination condition: (c) and (d)

Abaza and Bourlai in [4] tried used thermal image for ear recognition to overcome the illumination problem. The highest recognition rate obtained is 80.68% while other studies [6, 7] using visible images taken in controlled environments were able to achieve recognition rate of more than 95%. Since ear recognition to be deploy in real-life situation is more challenging, study using images in various illuminations needs to be done.

Fusion of thermal and visible image can be one solution for this problem. Objects in thermal image have the advantage of being visible despite bad illumination, while objects in visible image can be captured in detail. Therefore, due to these viable characteristics of both images, we proposed thermal and visible image fusion for better ear recognition regardless illumination condition. Fusion of thermal and visible images has been proven to improve face recognition [7, 8, 9, 10] compared to using thermal or visible image alone. Since image fusion is known to be workable for face recognition, can fusion of thermal and visible images improve ear recognition rate? At the moment of this writing, no known

literature is found discussing thermal and visible image fusion for ear recognition. Therefore, this paper intends to answer the question by applying the available thermal and visible fusion techniques of face recognition for ear recognition.

The rest of this paper is organized as follows. Section 2 discusses about image fusion and the available image fusion techniques. The methodology of the study is elaborated in Section 3. The results of our experiments are explained in Section 4, while Section 5 concludes this paper with the proposal for future work.

II. IMAGE FUSION

Image fusion can be defined as the process of combining two or more images with certain rules to produce new image for better analysis and understanding. Rockinger and Fechner [11] listed out three generic requirements that can be imposed on the fusion result. Firstly, all relevant information from source images should be preserved in the fused images (pattern conservation). Second, no artefacts or inconsistencies that will distract human observer or the following processes should be introduced during fusion process. Finally, the fused result should be shift and rotational invariant.

The fusion process may happen at three different levels (i.e. pixel, feature and decision) [12]. For this study, we will focus on fusion at pixel level as pixel-level fusion preserved the relevant information of the source images in the fused image.

In general, pixel-level image fusion is consists of three main stages as shown in Figure 2. The first stage is pre-processing. Pre-processing intends to prepare the image for the consecutive steps. Common pre-processing steps include greyscale conversion and contrast adjustment. The second stage in image fusion is image registration. Images to be fused are usually acquired using different image acquisition devices. Therefore, image registration process is needed to align both images.

The final stage that is pixel-level image fusion is the fusion process itself. At this stage, registered source images are subjected to image fusion algorithm resulting in one fused image. Pixel-level fusion can be done in either spatial domain or transform domain. Spatial domain has the tendency to suppress salient information producing low contrast images [13, 14]. In contrast, multi-scale fusion in transform domain allows features in different spatial extends to be fused [13]. Example of spatial domain fusion is weighted average (WA), while discrete wavelet transform is one example of fusion in transform domain.

A. Weighted Average (WA)

This is one simple method for image fusion. Each source image is weighted before being fused using addition [7]. The process is described by Equation 1.

$$F(x, y) = wA(x, y) + (1 - w)B(x, y) \qquad (1)$$

where $F(x, y)$ is the fused image, $A(x, y)$ and $B(x, y)$ are the source images and w is the weight for the fusion. Figure 3 shows the example of WA process while Figure 4 shows the example of WA done for ear images. In Figure 4, the top row illustrates visible image, thermal image and WA fused image, respectively, taken in a well-lit environment. Meanwhile, bottom row shows visible image, thermal image and WA fused image captured in a dark setting.
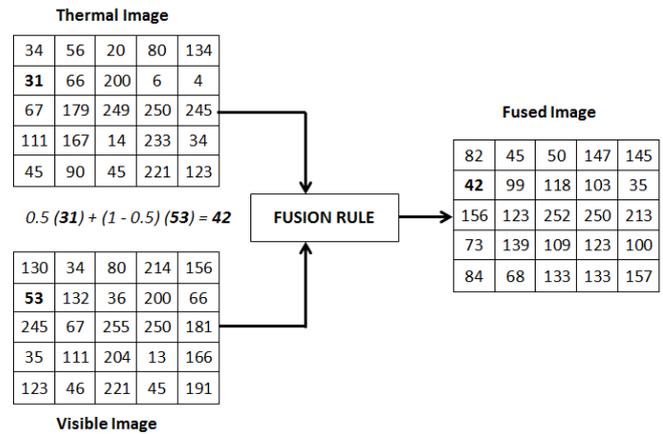
Figure 3: Example of weighted average process with $w = 0.5$

Figure 2: General image fusion process
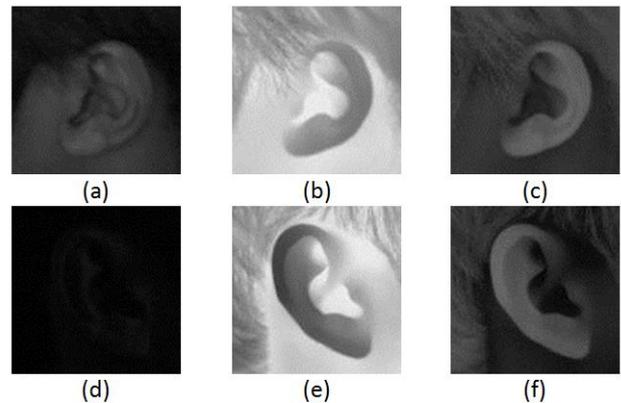
Figure 4: Examples of WA Fusion. (a) and (d) are visible images, (b) and (e) are thermal images, (c) and (f) are WA fused images

B. Discrete Wavelet Transform (DWT)

For DWT image fusion, source images are decomposed by discrete wavelet transform function up to one n level resulting in 3 sub bands and one approximate band. For this study, both visible and thermal images are decomposed to one level.

Once the images have been decomposed, fusion rules are applied to each band. In this study, we employed average fusion rule during image fusion. The same fusion rules can be applied to all sub bands or varying rules to each sub band depending on the needs and applications. After the fusion rules are applied, inverse DWT function is used to compose a new image from the modified sub bands. Figure 4 illustrates the general process of image fusion using DWT while Figure 5 shows the example DWT image fusion.

Figure 5: Block diagram of 1-level 2-dimensional DWT image fusion.

III. MATERIALS AND METHODS

This study consists of two experimental stages that are image fusion and ear recognition. The fusion stage starts with image acquisition followed by image pre-processing, image registration and finally image fusion. Ear recognition stage is meant to evaluate the performance of image fusion techniques based on recognition rate.

A. Fusion of Image

For this study, a new data set of ear images is developed. Currently, 40 subjects volunteered for this study and for each subject, 4 ear images are acquired. Thus, a total of 160 ear images are collected.
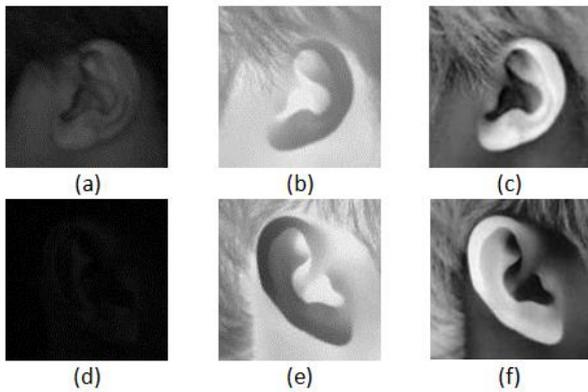
Figure 6: Examples of DWT Fusion. Top row are images taken in good illumination while bottom row taken in bad illumination. (a) and (d) are visible images, (b) and (e) are thermal images, (c) and (f) are DWT-fused images

a. Image Acquisition

Ear images are acquired based on the setup as shown in Figure 7, adopted from [4]. The camera for this purpose is FLIR E60 which has temperature sensitivity of 0.05°C and is also able to capture both thermal and visible images simultaneously.

The acquisition process starts when subject is seated on the provided chair with the ear facing the camera. Illumination of the environment is measured using lux meter. The lux meter sensor is placed directly on the ear to ensure accurate reading.

Then, both visible and thermal images are captured simultaneously. The process is repeated in different illumination conditions. The illuminations are manipulated by controlling the light sources and the position of subjects. There is a total of five illumination conditions per subject ranging from 3 lux to 8000 lux.
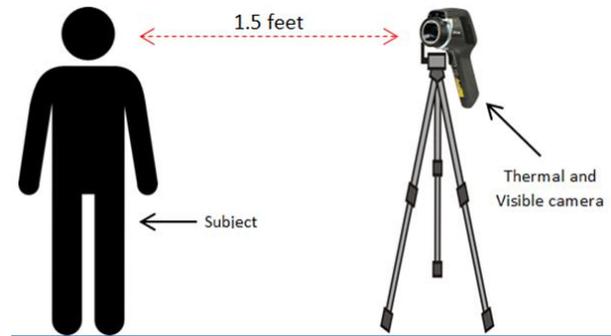
Figure 7: Image acquisition setup using camera with visible and thermal IR sensors

b. Image Pre-processing

The acquired images further underwent several pre-processing steps to prepare them for image fusion and later recognition. The visible images are resized to 800 x 600 pixels from the original 2048 x 1056 pixel to match the thermal images that have lower resolution (i.e. 360 x 240 pixels). Then, both thermal and visible images are converted to greyscale before being registered.

c. Image Registration

Even though the images are taken using the same device, they are however acquired by separate thermal and visible image sensors. Therefore, these images need to be registered before image fusion. Registration is done by using Triangular Fossa and the Incisure Intertragica as reference points [14]. The location of Triangular Fossa and the Incisure Intertragica are shown in Figure 8(a). Once registered, the ear region is then cropped to 125 x 125 pixels resolution as shown in Figure 9.
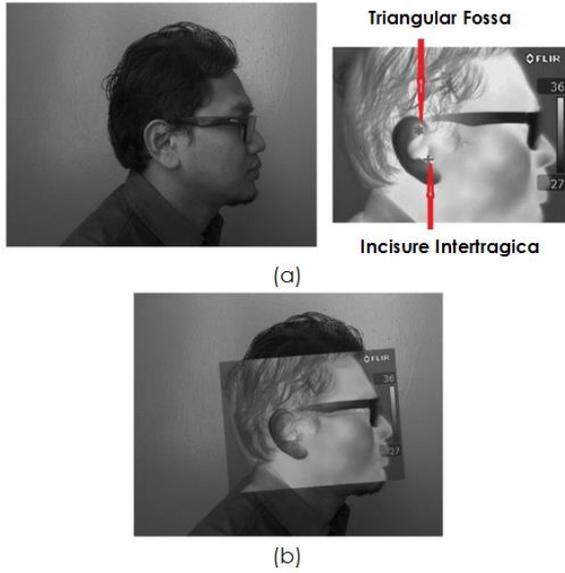
Figure 8: Example of image registration where (a) is the pre-processed images while (b) is registered image
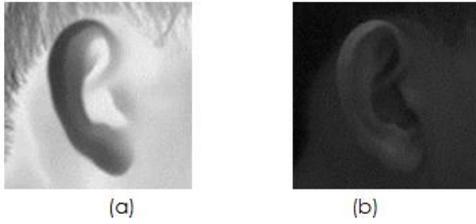
Figure 9: Example of the cropped registered image where (a) is thermal image and (b) visible image

d. Image Fusion

The fusion process took place once those images are registered. Images are fused using WA [7], average DWT and weighted DWT [2] fusion techniques. Referring to equation 1, the weight used in WA is 0.5. For average DWT, the fusion rule will be similar to WA however the fusion happened at each sub bands instead of the pixel value as in WA. Equations (2) to (5) shows the fusion rule for average DWT:

$$LL_f = 0.5(LL_t) + 0.5(LL_v) \qquad (2)$$
$$HL_f = 0.5(HL_t) + 0.5(HL_v) \qquad (3)$$
$$LH_f = 0.5(LH_t) + 0.5(LH_v) \qquad (4)$$
$$HH_f = 0.5(HH_t) + 0.5(HH_v) \qquad (5)$$

where LL_f, HL_f, LH_f and HH_f are sub bands of the new fused image. LL_t, HL_t, LH_t and HH_t are sub bands of the thermal image, while LL_v, HL_v, LH_v and HH_v are sub bands of the visible image.

Weighted DWT was adapted from [2]. The fusion rule is based on approximation coefficient (from low-pass filter, $W_\varphi(m,n)$, and detail coefficient (from high-pass filter), $W_\gamma(m,n)$. Equations (6) and (7) show the fusion rules.

$$W_\varphi(m,n) = \alpha_1 W_\varphi^{visible}(m,n) + \beta_1 W_\varphi^{thermal}(m,n) \qquad (6)$$
$$W_\gamma(m,n) = \alpha_2 W_\gamma^{visible}(m,n) + \beta_2 W_\gamma^{thermal}(m,n) \qquad (7)$$

To emphasize on both thermal and visible features, [2] have proposed the weight for α_1, α_2, β_1 and β_2 to be 0.5, 0.5, 1.0 and 1.0 respectively.

B. Ear Recognition

Before being evaluated using the ear recognition system, we categorized the ear images into three classes based on their lux readings. Currently, there are no known literatures that categorize image quality based on illumination or lux readings. Therefore, we are proposing the categorization based on the ear visibility in the image. In Class 1, the ears are either not seen or barely visible to the naked eyes. Images fall in this class have lux readings of 0 to 20. Images with 21 to 100 lux are categorized into Class 2 where the ear can easily be seen by the naked eyes. The ear features are most visible in images of Class 3 which are captured in illumination more than 100 lux. Summary of the categories is shown in Table 1.

Table 1
Summary of ear categorization

Class	Lux	Ear visibility
Class 1	0 – 20 lux	Poor
Class 2	21 – 100 lux	Moderate
Class 3	> 100 lux	Excellent

In recognition process, we took three samples (one from each class) as training set for one subject. Therefore, a total of 120 images are used as training dataset. Then, one image of the same subject from any of the 3 classes is chosen as test image. Therefore, 14 images from Class 1, 13 images from Class 2 and 13 images from Class 3 are used as test images from each 40 subjects.

The ear recognition used in this study employed eigenvectors as ear features known as eigen-ears, adopting the work in [15]. The process starts by initializing a set of ear images or the training set. Eigen-ears of the training set are then calculated. Only M images corresponding to the highest eigenvalues are kept, where M is the number of images in the training set. These M images define the ear spaces. Then the corresponding distribution in M-dimensional weight space is calculated for each known individual by projecting their ear images onto ear space.

The recognition begins by calculating a set of weight based on the input image and the M eigen-ears. This is done by projecting the input images onto each of the eigen-ears. Next, the recognition process checks whether or not the input image is an ear by referring to the ear space. The identity of the ear's owner is determined using Euclidean distance between the probe image and each image classes in the database.

IV. RESULTS AND DISCUSSIONS

Table 2 shows the result of ear recognition using visible, thermal, WA, average DWT and weighted DWT fused images while Figure 10 illustrates the result. Based on the result, visible images performed poorly with an average recognition rate of 20% for both side of ears. As mentioned earlier, features from visible images taken in a good illumination are easily extracted as the ears are well represented compared to dark images. The illumination variations in visible images caused data inconsistencies, thus severely affected the recognition rate.

Thermal images on the other hand, performed the best with 68.75% average recognition rate for both ears. As thermal images used emitted thermal IR radiation by the face instead of the reflected light, images obtained are invariant to illumination changes. The representation of thermal images is more consistent. Therefore, recognition rate of thermal images is better than visible images.

Table 2
Average ear recognition rate (%) for visible, thermal, WA fused and DWT fused image

	Left	Right	Both
Visible	15	25	20
Thermal	72.5	65	68.75
WA	60	67.5	63.75
Avg. DWT	65	70	67.5
Weighted DWT	62.5	70	66.25
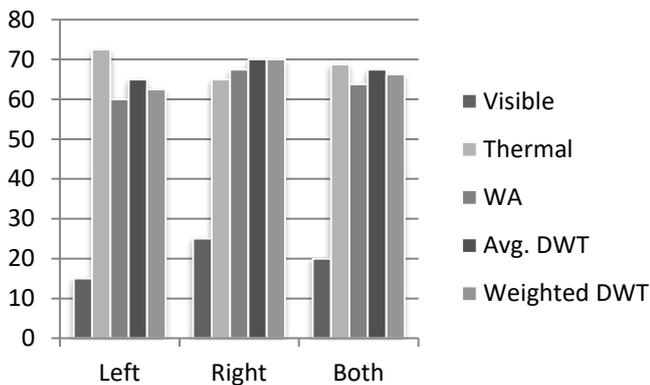
Figure 10: Ear recognition rate.

As for fused images, average DWT fused images performed better with 67.5% average recognition rate compared to weighted DWT and WA fused images which only achieved 67.5% and 63.75% recognition rate, respectively. Average DWT fused images also achieved higher recognition rate for left and right ears compared to weighted DWT and WA fused images. As can be seen in Figure 4 and Figure 6, images underwent DWT based fusion represent the ear details better than WA fused images, thus contribute to better ear recognition result.

Even though thermal images performed better than average DWT fused image, the 1.25% difference is barely significant. Despite thermal images achieving better recognition rate of 72.5% for left ears, average DWT and weighted DWT fused images however did outperformed thermal images for right ears with 70% recognition rate. Therefore, further investigations on the fusion rules and approach need to be done as no conclusive decision can be done based on the results.

V. CONCLUSION AND FUTURE WORKS

This paper discussed the possibility of applying image fusion for ear recognition and which fusion technique is best to do the job. Based on the result of the experiment, thermal images performed better than DWT fused images with narrow and barely significant margin. Besides, the recognition rate is also inconsistent where thermal images achieved higher than DWT fused images for left ears and vice versa for the right ears. Therefore, no concrete conclusion can be drawn to which fusion technique is best for ear recognition in various illuminations. Therefore, further studies need to be done to determine which type of images and fusion technique are better for robust ear recognition.

We are also expanding our ear images database with more subjects for more conclusive results. Nonetheless, we still believe fusion of images thermal and visible images is better than thermal images for illumination invariant ear recognition as it did for face recognition. Hence, modifications to the current fusion technique will be proposed for better recognition result. Besides, other feature extraction methods may also be considered for the ear recognition system.

REFERENCES

[1] Iannarelli, A. V., 1989. Ear Identification, in Forensic Identification Series, Freemont, California.: *Paramont Publishing Company*.

[2] Kong, S. G., Heo, J., Boughorbel, F., Zheng, Y., Abidi, B. R., Koschan, A., Yi, M. and Abidi, M.A., 2007. Multiscale Fusion Of Visible And Thermal IR Images For Illumination- Invariant Face Recognition, *International Journal of Computer Vision*. 71(2):215–233.

[3] Abaza, A. and Bourlai, T., 2012. Human Ear Detection In The Thermal Infrared Spectrum, *in Thermosence : Thermal Infrared Application XXXIV*. 8354:83540X.

[4] Abaza, A. and Bourlai, T., 2013.On Ear-Based Human Identification In The Mid-Wave Infrared Spectrum, Image and Vision Computing. 31(9):640–648.

[5] Arbab-Zavar, B. and Nixon, M. S., 2008. Robust Log-Gabor Filter For Ear Biometrics, *2008 19th International Conference on Pattern Recognition*.1–4.

[6] Kumar, A. and Wu C., 2012. Automated Human Identification Using Ear Imaging, Pattern Recognition. 45(3):956–968.

[7] Hanif, M. and Ali, U. , Optimized Visual And Thermal Image Fusion For Efficient Face Recognition, *2006 9th International Conference on Information Fusion, FUSION,* 2006.

[8] Bebis, G., Gyaourova, A., Singh, S. and Pavlidis, I., 2006. Face Recognition By Fusing Thermal Infrared And Visible Imagery, *Image Vision Computing*. 24(7):727–742,

[9] Moon, S., Kong, S. G., Yoo J.-H. and Chung, K., 2006. Face Recognition with Multiscale Data Fusion of Visible and Thermal Images, i*n IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety*.16–17.

[10] Singh, R., Vatsa M. and Noore, A., 2008. Hierarchical Fusion Of Multi-Spectral Face Images For Improved Recognition Performance, *Information Fusion*. 9(2):200–210.

[11] Rockinger, O. and Fechner, T., 1998. Pixel-Level Image Fusion: The Case Of Image Sequences, *Proceeding SPIE*. 3374:378–388.

[12] Kayani, B. N., Mirza A. M., Bangash, A., and Iftikhar, H. 2007. Pixel & Feature Level Multiresolution Image Fusion Based On Fuzzy Logic, *in Innovations and Advanced Techniques in Computer and Information Sciences and Engineering, Netherland: Springer*, 129–132.

[13] Yang, B., Jing, Z. L., and Zhao, H. T. 2010. Review Of Pixel-Level Image Fusion, *Journal of. Shanghai Jiaotong University*. 15 (1): 6–12.

[14] Nanni, L. and Lumini, A., 2007. A Multi-Matcher For Ear Authentication, *Pattern Recognition Letter*. 28(16):2219–2226.

[15] Turk, M.A. and Pentland, A. P., 1991. Face Recognition Using Eigenfaces, *Journal of Cognitive Neuroscience*. 3(1).