Multimodal biometrics system based on face profile and ear

Iman S. Youssef[a] Ayman A. Abaza[a,b] Mohamed E. Rasmy[a] and Ahmed M. Badawi[a]

[a] Systems and Biomedical Engineering, Cairo University, Egypt;
[b] West Virginia High Tech Foundation, Fairmont, USA

ABSTRACT

Face recognition from a side profile view, has recently received significant attention in the literature. Even though current face recognition systems have reached a certain level of maturity at angles up to 30 degrees, their success is still limited with side profile angles. This paper presents an efficient technique for the fusion of face profile and ear biometrics. We propose to use a Block-based Local Binary Pattern (LBP) to generate the features for recognition from face profile images and ear images. These feature distributions are then fused at the score level using simple mean rule.

Experimental results show that the proposed multimodal system can achieve 97.98% recognition performance, compared to unimodal biometrics of face profile 96.76%, and unimodal biometrics of ear 96.95%, details in the Experimental Results Section. Comparisons with other multimodal systems used in the literature, like Principal Component Analysis (PCA), Full-space Linear Discriminant Analysis (FSLDA) and Kernel Fisher discriminant analysis (KFDA), are presented in the Experimental Results Section.

Keywords: Face Profile and Ear Recognition, Local Binary Pattern, and Biometric Fusion

1. INTRODUCTION

Traditionally, biometric systems that use only one biometric trait (uni-modal biometric systems) are very common. Many drawbacks for these uni-modal systems were reported.[1,2] Recently, some multimodal biometric systems in the literature were reviewed and successfully developed to significantly overcome some of these problems. These systems were proposed to fulfill the need for better recognition accuracy by consolidating more than one independent source of evidence to recognize individuals and ensuring identities. Moreover, some overcome spoofing to some extent by adding complexity to the personality characteristics.

Generally, Face profile and ear biometric systems consist of three main stages(as shown in Figure 1):

- Region detection / segmentation: Segmenting the face profile / ear region from the image.
- Feature extraction: Representing the face profile / ear structure by feature vectors.
- Classification: Matching the probe and the gallery feature vectors to verify the subject claimed identity (Verification mode) or to search a database to identify the admitted person (Identification), or other modes like match against a watch list, etc.

Face profile and ear multimodal biometrics systems have not been given much attention in the literature and especially the profile side views that can be acquired by surveillance cameras with no controlled environment or even with a controlled environment. Very limited research articles in this area were found in the literature. At the time of writing this paper, only 4 publications, that we are aware of.[3–6] These publications are summarized in section 2.

In this paper, we introduce a new approach for the fusion of face profile and ear biometrics. First, the image is divided into a number of regions or blocks. Then the Local Binary Pattern (LBP) features are extracted producing several histograms that are concatenated to represent the image. The chi-square distance is used for similarity measure generated by LBP operator.[7] Different experiments were conducted to tune the parameters of the LBP operator. Then testing experiments were performed on another database. Recognition rates reached 97.98% using the weighted-sum rule. Finally, a comparison against various techniques used in the literature of fusion of face profile and ear was conducted.

Further author information: (Send correspondence to Ayman Abaza)
A.A.: E-mail: aabaza@wvhtf.org, Telephone: +1 304 333 6452.

Biometric and Surveillance Technology for Human and Activity Identification XI,
edited by Ioannis A. Kakadiaris, Walter J. Scheirer, Christoph Busch, Proc. of SPIE
Vol. 9075, 907506 · © 2014 SPIE · CCC code: 0277-786X/14/$18 · doi: 10.1117/12.2050159

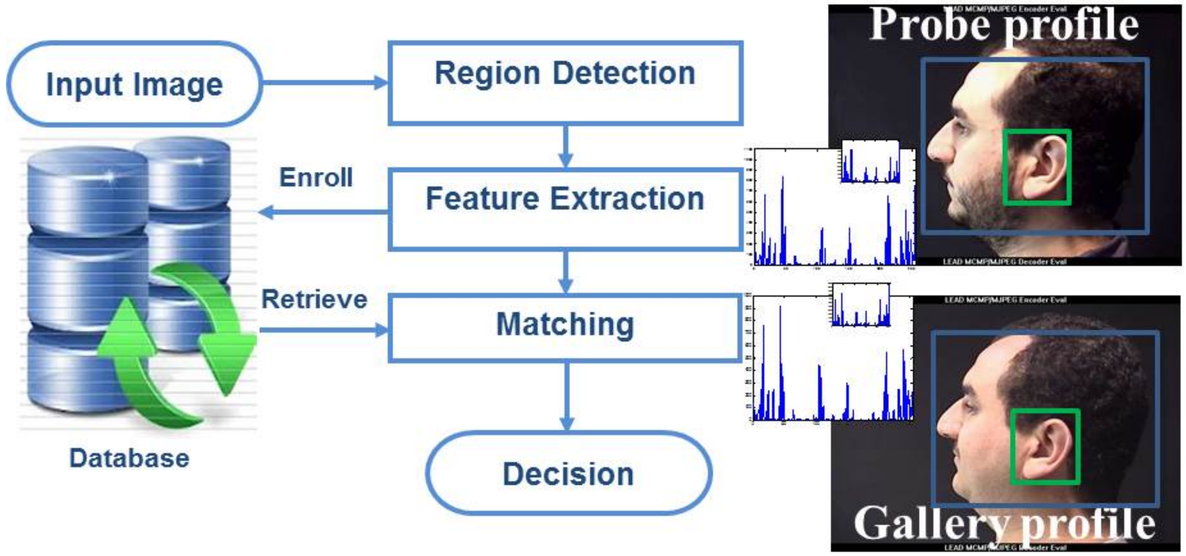

Figure 1. Face profile and ear recognition system

This article is organized as follows: Section 2 highlights some related work on 2D face profile and ear fusion. Section 3 presents a brief overview of the block-based Local Binary Pattern technique. Experimental results are presented in Section 4. Finally, Section 5 presents conclusions and sketches our future plans.

2. RELATED WORK

Rahman et al.[3] used the UND database which contains 18 subjects with 5 different variations. First during the image processing stage, the face profile and ear are extracted and cropped manually from each image. Then the image size is reduced to 32×32. To minimize the light effect, the image is then normalized. The second stage is the calculation of the eigenfaces and eigenears using Principal Component Analysis (PCA). Only one image per subject is used in training. The last stage is the testing in which 4 images per subject were used. They used face profile image without the ear part. Fusion results were 94.44%.

Xu et al.[4] employed a subset of USTB database with 294 images for 42 subjects. First during the preprocessing stage, the face profile and ear are cropped from each image and then filtered using Wiener filtering to emphasize the features. All the images are resized to 200×200 for the face profile and 80×50 for the ear. Then histogram equalization is used on the images. The second stage is where the Full-space Linear Discriminant Analysis (FSLDA) is applied for feature extraction and classification. Finally during the decision fusion level, the multi-modal biometric ear and face profile is integrated for both classifiers. This step is performed using the combination methods of Product, Sum and Median rules according to the Bayesian theory and a modified Vote rule for two classifiers. Their fusion results reached 97.62%.

Yuan el al.[5] combined the face profile with the ear on USTB database. They selected five profile images per subject with variations of the head position and facial expressions. They normalized the images sizes to 200×200 pixels. They used 3 images per subject for training FSLDA for feature extraction. For testing they used 79 images. The recognition results reached 96.2%.

Pan et al.[6] used USTB for 79 persons with variations of the head position, facial expressions and glasses. For their experiments, they selected nine images with rotation from $-20°$ to $20°$ with $5°$ interval. Images were cropped and filtered with Wiener filter to emphasize the features. Images were resized to 40×25 pixels, and the face profile images to 100×100 pixels. Then they applied histogram equalization, to produce images with equally distributed intensity values. Kernel Fisher discriminant analysis (KFDA) was applied and fusion at the feature extraction level was done. Results reached 96.84% for the weighted-sum rule.

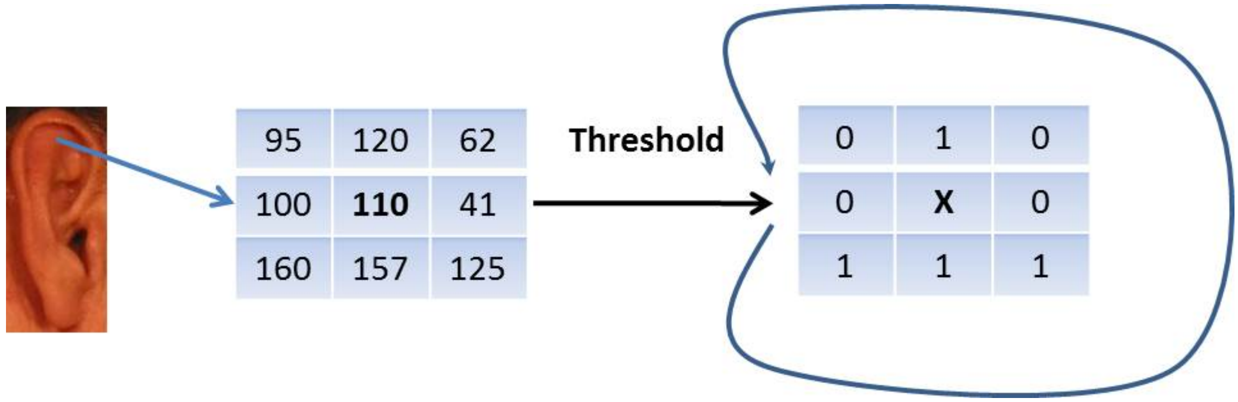

Figure 2. Basic LBP operator: (1) For a given input image pixel and its 8 neighbors, (2) Each neighbor pixel greater than or equal to the center pixel is assigned 1 otherwise it is assigned 0, (3) These binary values are arranged to form a binary number (01110010), which is transferred to a decimal equivalent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

3. LOCAL BINARY PATTERNS

Ojala et al.[8] quantify the intensity patterns in local pixel neighborhood patches such as spots, line ends, edges, corners, and other distinct texture patterns. The basic Local Binary Patterns (LBP) operator assigns a decimal value to each pixel in the image by thresholding (P=8) neighbor pixels at distance (R=1), as shown in Figure 2. The histogram (H) of these decimal values represents the feature vector. Ojala et al.[9] extended the LBP operator by using neighborhood of various sizes, generally, "P" neighborhood pixels, at distance "R". Bilinear interpolation is used for points out of grid to calculate an approximation of a pixel's intensity based on the values at surrounding pixels.

3.1 Uniform and Non-uniform Local Binary Patterns

According to Ojala et al.,[10] what determines a local binary pattern to be called uniform or non-uniform is the number of bitwise transitions from 0 to 1. The uniform LBP contains at most two bitwise transitions, otherwise it is non-uniform. For example, the patterns "00000000" and "11111111" have 0 transitions, while "00000111" and "10000001 have 2 transitions. Whereas the patterns "11001001" and "01010011" are non-uniform, because they include 4 and 6 transitions respectively.

In the computation of the LBP histogram, uniform patterns are used so that the histogram has a separate bin for every uniform pattern and all non-uniform patterns are assigned to a single bin. The following notation will be used to define an LBP operator: $LBP^{U}_{P,R}$ where the subscript represents using the operator in a (P,R) neighborhood; P sampling points on a circle of radius R. We omit the superscript U, stands for using uniform patterns, as we will only consider uniform pattern for the rest of the paper.

3.2 Block Based Division

In this approach, the image is divided into a number of blocks (as shown in Figure 3). These blocks can be of arbitrary size and can overlap. Then the LBP operator is applied to each block separately, and their corresponding histograms are calculated. Integration of these blocks can either be at the feature level or at the score level. For integration at the feature level the histograms extracted from various blocks are concatenated and the overall histogram is used for matching. On the other hand, integration at the score level involves matching the histogram extracted from each block alone then fusion of scores from these blocks. Sub-blocks are expected to be more discriminative than using the whole image.

4. EXPERIMENTAL RESULTS

In this section, we present various experiments to evaluate the recognition performance of score level fusion of face profile and ear features. First, we present experiments to tune the different parameters of LBP. Second, we present another experiment to adjust the fusion weight for face and ear. Third, we compare the performance of the proposed method against techniques used in the literature.

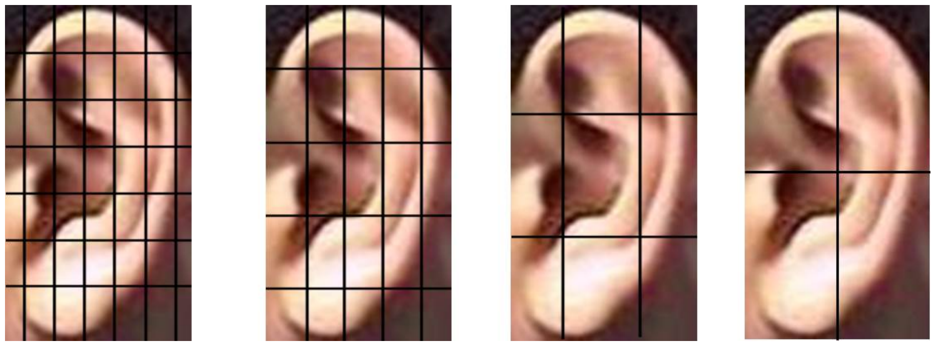

Figure 3. Block Based Division: Ear image divided into 7x7, 5x5, 3x3 and 2x2 rectangular regions.

For all the experiments, the system is working in a verification mode, where the system matches the probe and the gallery feature vectors to verify the subject claimed identity. To evaluate the system performance, we used either the following performance measures:

- Equal Error Rate (EER): the biometric system can change its threshold value; which changes its False Acceptance Rate (FAR) and its False Rejection Rate (FRR). When these rates are equal, the common value is referred to as the equal error rate.
- Recognition Rate is defined as (100-Equal Error Rate). The lower the equal error rate value, the higher the accuracy of the biometric system.

4.1 Face Profile / Ear Databases

Two databases were used for various experiments:

1. The University of Notre Dame (UND) Biometrics databases* are available for public use.[11] We use Collection E, because it contains 464 visible-light face side profile (ear) images from 114 human subjects. We refer to this data set as UND. It contains 102 subjects to maintain 2 images per subjects, and we use it in the testing phase.
2. The University of Science and Technology Beijing (USTB) databases † are available for academic research.[12] Database III holds right side profile full images photographed with color CCD camera under white background and constant lighting. Image resolution is 768×576, 24-bit true color and BMP format. This database contains 790 images of 79 volunteers. Images within USTB database III include rotation from $0°$ to $60°$ with variable $5°$ intervals to the right side. We use this set to tune various parameters of LBP, as well as to select the weight between face profile and ear in the weight sum fusion rule.

4.2 Preprocessing

Before conducting the experiments, images were preprocessed as follows (as shown in Figure 4):

- Twelve subjects with occluded ears were excluded from testing.
- RGB color images were converted to 256-bit grey-scale.
- Face profile images were segmented to remove any extra background, and then resized.
- Two more steps are added for ear images; rotation and cropping. It is to be mentioned that face profile images were rotated first then ear segments were cropped, and then resized.

*http://www3.nd.edu/ cvrl/CVRL/Data_Sets.html
†http://www1.ustb.edu.cn/resb/en/index.htm

Figure 4. Example of preprocessing: (a) Original image, (b) Cropped face profile image, (c) Rotated face profile, and (d) cropped ear image.

Table 1. Experiment to tune LBP parameters for face profile recognition, using Equal Error Rate (EER)

Operator		Equal Error Rate(EER)
Whole Image	$LBP_{8,1}$	6.76%
Whole Image	$LBP_{8,2}$	8.30%
Whole Image	$LBP_{16,2}$	**6.11%**
Whole Image	$LBP_{16,2}$	6.11%
(2×2)	$LBP_{16,2}$	**5.05%**
(3×3)	$LBP_{16,2}$	9.22%
(5×5)	$LBP_{16,2}$	12.20%
(7×7)	$LBP_{16,2}$	17.11%

4.3 Tuning Local Binary Patterns

To tune various parameters of Local Binary Patterns (LBP) method, we use the training set from USTB database. For all the subjects in the USTB data sets, we use: (i) One image per subject as a gallery; (ii) One image per subject as a probe. To measure the similarity between the probe histogram H^p and gallery histogram H^g generated by the LBP operator, we used the chi-square distance:

$$S_{Chi}(H^p, H^g) = \Sigma_{j,i}\omega_j * \frac{(H^p_{i,j} - H^g_{i,j})^2}{(H^p_{i,j} + H^g_{i,j})} \tag{1}$$

where i and j refer to the i^{th} bin in histogram corresponding to the j^{th} block.

We set up several experiments to tune LBP operators for face profile recognition (details in Table 1). The first experiment selects the number of neighbors points (P) and the radius of these points from the center (R). This experiment shows that the $LBP_{16,2}$ operator achieves the best performance. Hence we decide to use the $LBP_{16,2}$ operator for the remaining experiments. The second experiment selects the number of blocks (2×2, 3×3, 5×5, and 7×7) versus using the whole image. This experiment indicates that using block size of (2×2) yields the best performance. Hence we decide to divide the face profile images into (2×2) blocks for the remaining experiments.

We set up another set of experiments to tune LBP operators for ear recognition (details in Table 2). The first experiment selects the number of neighbors points (P) and the radius of these points from the center (R). This experiment shows that the $LBP_{16,2}$ operator achieves the best performance. Hence we decide to use the $LBP_{16,2}$ operator for the remaining experiments. The second experiment selects the number of blocks (2×2, 3×3, 5×5, and 7×7) versus using the whole image. This experiment indicates that using block size of (3×3) yields the best performance. Hence we decide to divide the face profile images into (3×3) blocks for the remaining experiments.

4.4 Fusion of face profile and ear

Fusion can be applied at match score level, where match scores output by different biometric matchers are consolidated. This approach has been widely used since match scores are easy to access and combine. However, match scores output by different biometric matchers may not be homogeneous, which is not the case in the following experiments. Thus, no normalization step is needed. Several integration rules can be used to implement score level fusion. A fusion rule which is

Table 2. Experiment to tune LBP parameters for ear recognition, using Equal Error Rate (EER)

Operator		Equal Error Rate (EER)
Whole Image	$LBP_{8,1}$	19.35%
Whole Image	$LBP_{8,2}$	10.33%
Whole Image	$LBP_{16,2}$	**9.32%**
Whole Image	$LBP_{16,2}$	9.32%
(2×2)	$LBP_{16,2}$	7.96%
(3×3)	$LBP_{16,2}$	**7.52%**
(5×5)	$LBP_{16,2}$	8.06%
(7×7)	$LBP_{16,2}$	11.85%

Table 3. Experiment to tune the weights employed in the weighted-sum rule, using Equal Error Rate (EER)

α	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
β	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0
EER	8.06	6.82	5.39	6.16	6.26	6.89	**5.15**	5.53	9.42	9.83	12.20

commonly used in the literature is the *simple mean* formulated in the following equation:

$$S_{mean} = (\sum_{k=1}^{K} S_k)/K \tag{2}$$

where s_k is the match score output by the k^{th} matcher.

Another integration rule is the weighted-sum rule, where equal weights means the simple mean. To tune the weights employed in the weighted-sum rule, an experiment is carried out to find the amount of contribution of face profile and ear in the identification procedure:

$$S_{WS} = \alpha S_{fp} + \beta S_{ear} \tag{3}$$

where α is the face profile weights and β is the ear weights and $\alpha + \beta = 1$.

These experiments show that for the weighted-sum rule, the weight set as α = 0.6 "ear score weight" and β = 0.4 "face profile score weight", yields the best performance according to Table 3.

We conducted an experiment to study the recognition performance for the proposed face profile and ear recognition system using LBP technique applied on UND set. The recognition performance for the mentioned experiment: (i) unimodal biometrics of face profile 96.76%, (ii) unimodal biometrics of ear 96.95%, and (iii) multimodal face profile and ear 97.98%. Figure 5 shows Receiver operating characteristic (ROC) curves for each unimodal in comparison to their weighted sum score fusion.

4.5 Comparing the proposed technique to the literature

Table 4, shows the recognition performance of the proposed block-based LBP technique compared to the PCA, FSLDA and KFDA techniques. For this experiment $LBP^{U}_{16,2}(3 \times 3)$ operator for ear recognition, and $LBP^{U}_{8,1}(2 \times 2)$ operator for face recognition, and weighted-sum rule for fusion were used. We used the UND set in this test because: (i) it has more cases compared to USTB, and (ii) We completely separate the training set (USTB) from the testing set (UND).

These experiments prove that the proposed system achieve better performance compared to similar system in the literature, given that we only use one gallery image per subject (the more the gallery images per subject, the higher the achieved accuracy). The recognition performance for the proposed system:

- unimodal biometrics of face profile 96.76% in comparison to the KFDA technique 93.46%;
- unimodal biometrics of ear 96.95% in comparison to FSLDA technique 94.05%;
- multimodal face profile and ear 97.98% in comparison to FSLDA technique 97.62%.

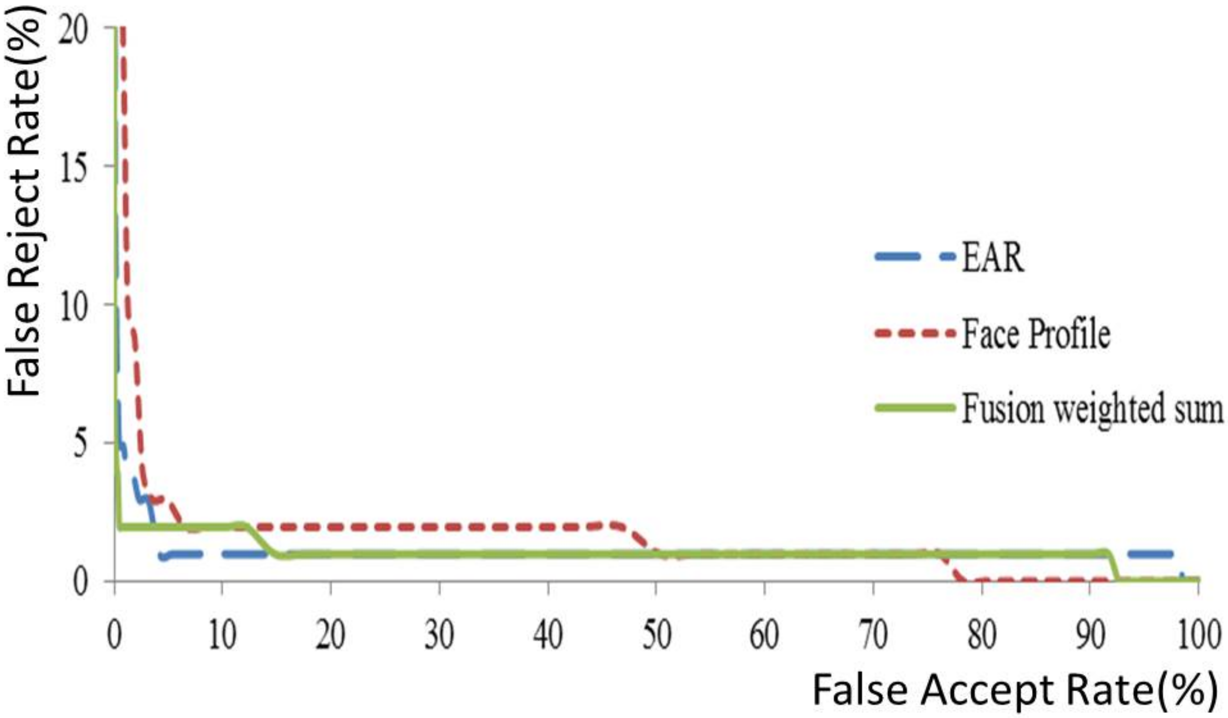

Figure 5. Receiver operating characteristic (ROC) curve for unimodal face profile, unimodal ear, compared to their score-level fusion

Table 4. Comparison between different techniques used in fusion of face profile and ear with our proposed technique, using recognition performance.

	Rahman et al.[3]	Yuan el al.[5]	Pan et al.[6]	Xu et al.[4]	Proposed Technique
Technique Applied	(PCA)	(FSLDA)	(KFDA)	(FSLDA)	(LBP)
Database Used	UND	USTB	USTB	USTB	UND
Subject Number	18	79	79	42	102
Face Profile Results	88.88%	N/A	93.46%	88.10%	96.76%
Ear Results	77.77%	N/A	91.77%	94.05%	96.95%
Fusion Results	94.44%	96.20%	96.84%	97.62%	97.98%

5. CONCLUSION AND FUTURE WORK

In this paper, we proposed a new multimodal system of face profile and ear biometrics. We used block-based local binary pattern technique to extract features from face profile and ear regions, then we integrated the resulted scores using weighted sum. Results were shown to yield about 1% enhancement compared to those involving only uni-modal biometric system of face profile or ear alone. The proposed system can work real-time or near real, given that using Matlab implementation it takes about 500 ms to perform face profile and ear detection and feature vectors extraction.

The multimodal system of face profile and ear achieved the best recognition performance 97.98% compared to other techniques in the literature such as PCA 94.44%, FSLDA 97.62% or KFDA 96.84%, given that we used the largest database.

In future work, we plan to (i) automate the segmentation and alignment of face profile and ear; (ii) enhance the performance (Recognition rate) of the proposed methods includes by studying more advanced methods for dividing the images, and assigning different weights to the extracted blocks.

REFERENCES

[1] Jain, A. and Ross, A., [*Handbook of Biometrics*], ch. 1 Introduction to Biometrics, 1–22, Springer (2007).

[2] Jain, A., Ross, A., and Prabhakar, S., "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology* **14**(1), 4 – 20 (2004).

[3] Rahman, M. M. and Ishikawa, S., "Proposing a passive biometric system for robotic vision," in [*Proc. of the* 10^{th} *International Symposium on Artificial Life and Robotics (AROB)*], (Feb 2005).

[4] Xu, X. and Mu, Z.-C., "Multimodal recognition based on fusion of ear and profile face," in [*Proc. of the* 4^{th} *International Conference on Image and Graphics ICIG*], 598–603 (2007).

[5] Yuan, L., Mu, Z.-C., and Liu, Y., "Multimodal recognition using face profile and ear," in [*Proc. of the IEEE International Symposium on Systems and Control in Aerospace and Astronautics ISSCAA*], 887–891 (2006).

[6] Pan, X., Cao, Y., Xu, X., Lu, Y., and Zhao, Y., "Ear and face based multimodal recognition based on KFDA," in [*Proc. of the International Conference on Audio, Language and Image Processing ICALIP*], 965–969 (2008).

[7] Abaza, A. A. and Harrison, M. F., "Ear recognition: A complete system," in [*SPIE, DSS, Biometric Technology for Human Identification*], (2013).

[8] Ojala, T., Pietikinen, M., and Harwood, D., "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition* **29**(1), 51 – 59 (1996).

[9] Ojala, T., Pietikainen, M., and Maenpaa, T., "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* **24**(7), 971 – 987 (2002).

[10] Ahonen, T., Hadid, A., and Pietikainen, M., "Face description with local binary patterns: application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* **28**(12), 2037 – 2041 (2006).

[11] UND, "University of Notre Dame (UND) Databases." available at: $http://www.cse.nd.edu/~cvrl/CVRL/Data\_Sets.html$.

[12] USTB, "University of Science and Technology Beijing (USTB) Database." available at: $http://www.ustb.edu.cn/resb/en/subject/subject.htm$.