Journal of Applied Logic 17 (2016) 4–13

Contents lists available at ScienceDirect

Journal of Applied Logic

www.elsevier.com/locate/jal

A small look at the ear recognition process using a hybrid approach

Pedro Luis Galdámez *, Angélica González Arrieta, Miguel Ramón Ramón

Salamanca University, Plaza de los Caídos, 37008 Salamanca, Spain

ARTICLE INFO

Article history:
Available online 28 September 2015

Keywords:
Hausdorff
LDA
IRT
SURF algorithm
Ear recognition
Neural network

ABSTRACT

The purpose of this document is to offer a combined approach in biometric analysis field, integrating some of the most known techniques using ears to recognize people. This study uses Hausdorff distance as a pre-processing stage adding sturdiness to increase the performance filtering for the subjects to use it in the testing process. Also includes the Image Ray Transform (IRT) and the Haar based classifier for the detection step. Then, the system computes Speeded Up Robust Features (SURF) and Linear Discriminant Analysis (LDA) as an input of two neural networks to recognize a person by the patterns of its ear. To show the applied theory experimental results, the above algorithms have been implemented using Microsoft C#. The investigation results showed robustness improving the ear recognition process.

1. Introduction

The ears do not have a completely random structure. They have standard part as other biometric traits like the face. A recognition system based on images of the ears is very similar to a typical face recognition system, however, the ear has some advantages over the face, for example, their appearance does not change due to expression and is little affected by the ageing process, its color is usually uniform and the background is predictable.

Although the use of information from ear identification of individuals has been studied, it is still debatable whether or not the ear can be considered unique or unique enough to be used as a biometric. However, any physical or behavioral trait can be used as biometric identification mechanism if it is universal, that every human being possesses an identifier, being distinctive and unique to each individual, invariant in time, and measurable automatically or manually; the ear accomplishes all these characteristics.

* Corresponding author.
E-mail address: peter.galdamez@usal.es (P.L. Galdámez).

http://dx.doi.org/10.1016/j.jal.2015.09.004

2. Brief review of the literature

Significant progress has been made in the past few years in ear biometrics field. One of the most important techniques which are known to detect the ears is raised by Burge and Burger [6] who have made the process of detection using deformable contours with the observation that initialization contour requires user interaction. Therefore, the location of the ear is not fully automatic. Meanwhile Hurley et al. [11] used the technique of force field, this process ensures that it is not required to know the location of the ear to perform recognition. However, only applies when the technique has the specific image of the ear out of noise. In [22], Yan and Bowyer have used manual technique based on two previous lines for detection, where takes a line along the border between the ear and face while another line crosses up and down the ear.

Ansari and Gupta [1] presented a process based on the outer ear helices edges, they use 700 samples collected at IIT Kanpur, the strategy only relies on the outer helix curves. Yuan and Mu [23] have proposed a skin-color and contour information technique, they perform the ear detection considering ear shape elliptical and fitting an ellipse to the edges to get the accurate ear position. Attarchi et al. [2] have shown an ear detection process based on the edge map. It relies on the hypothesis that the longest path in edge image is the ear outer boundary. It works well only when there is not noisy background present around the ear and fails if ear detection is carried out in whole profile face image; they use two databases, USTB and Carreira-Perpinan with 308 and 102 images [20] with an accuracy of 98.05% and 97.05% respectively.

A. Cummings et al. [9] show a strategy using the image ray transform which is capable of highlighting the ear tubular structures. The technique exploits the helix elliptical shape to calculate the localization. Kumar et al. [14], have introduced a proposal where uses skin segmentation and edge map detection to find the ear, once they find the ear region apply an active contour technique [17] to get the exact location of ear contours. In the context of 3D images, Zhou et al. [24] presented a novel shape based feature set, called Histograms of Categorized Shapes (HCS), for robust 3D ear detection, using a sliding window approach and linear Support Vector Machine (SVM).

In other terms a biometric recognition system requires the discovery of unique features that can be measured and compared in order to correctly identify subjects. There are some known techniques for ear recognition specially in 2D and 3D images, as the strategies based on appearance, force transformation, geometrical features, and the use of neural networks.

The most used technique for face recognition principal component analysis (PCA) was applied for ear recognition by Victor et al. [3]. They used PCA to perform an comparative analysis between face and ear, concluding that the face performs better than the ear. However, Chang et al. [16] also have accomplished a comparison using PCA and found that ears provided similar performance, concluded that ears are essentially just as good as faces for biometric recognition. There are many proposals to solve the problem, in this paper only has done a small review from some of them, the next section introduce an intent to solve the problem of ear recognition using a practical way, applying some interesting concepts for 2D images and real time video.

3. Ear recognition system

Recognition systems traditionally follow a set of standards, such as, acquiring images, pre-processing, feature extraction, and the classification and/or recognition of the respective object. All of these tasks will be described in upcoming sections connecting important algorithms in order to complete its goal. Nevertheless, it is important to notice that the process that we are about to describe is based in the combination of some existing methods in order to build a robust system, allowing to perform, detection, tracking, and recognition in a real time video using identification through the ear.

In this way, the system combines a series of algorithms that give significant results individually, and when they are combined, achieve a higher degree of robustness with significant improving in problems such as

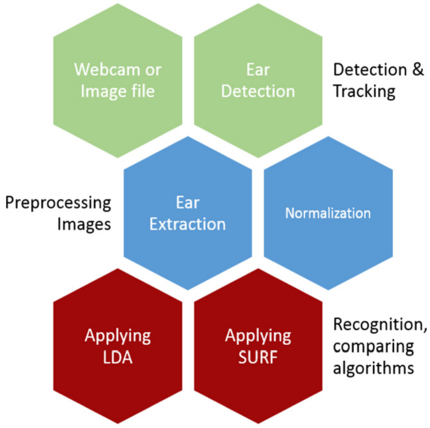

Fig. 1. System flow chart.

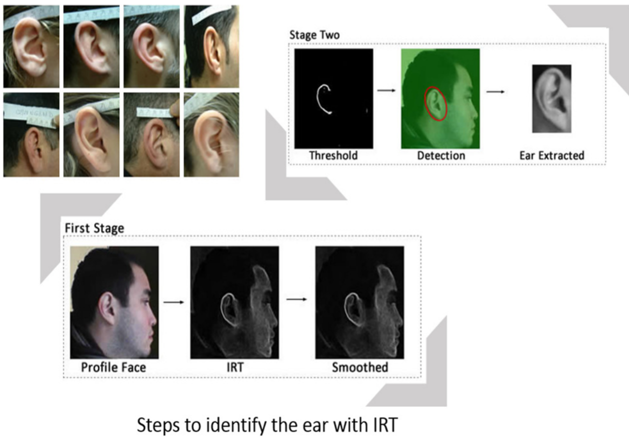

Fig. 2. Database and steps to identify the ear with IRT.

changes in pictures brightness and perspective. Fig. 1 shows the workflow that the project will follow, next sections will deepen these steps.

3.1. Detecting and preprocessing ears

OpenCV and its wrapper for .Net framework EmguCV [12] includes different object detectors based on the Viola–Jones framework, most of them are been constructed to deal with different patterns like frontal face, eyes, nose, etc. Modesto Castellón-Santana et al. [8] have developed a haar-cascade classifier to be used with OpenCV to detect left and right ears. This classifier represents a first step to create a robust ear detection and tracking system.

The haar classifier is used to identify the face profiles, with this captures we proceed to obtain the ear using the same haar approach, if this technique cannot identify the ear, the system will compute the IRT. The original images from the Ávila's Police School database created for this research with 300 ears, 3 images per subject undergoes using the technique proposed by Cummings et al. [9] where the system computes the IRT, then it applies the Gaussian Smoothing in order to reduce noise and remove gaps in the helix. Then the image is thresholded to remove as much of the image as possible whilst trying to leave the helix intact before to use an elliptical template to match the image (see Fig. 2).

With the ear identified we proceed to perform the pre-processing task, converting the image to gray scale and we begin the normalization process, first we perform the segmentation of the image applying a mask

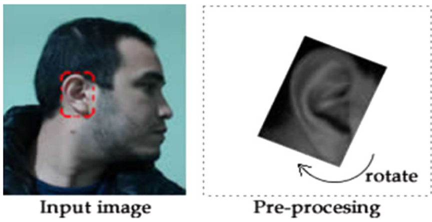

Fig. 3. Ear detection.

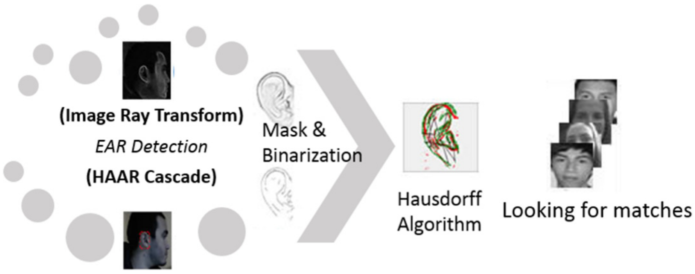

Fig. 4. Image pre-processing.

Fig. 5. Hausdorff pre-processing.

to extract only the ear, then the image is converted to an edge map using the canny edge filter. If w is the width of the image in pixel and h is the height of the image in pixel, the canny edge detector takes as input an array $w \times h$ of gray values and sigma (see Fig. 3).

The output is a binary image with a value 1 for edge pixels, i.e., the pixel which constitute an edge and a value 0 for all other pixels. We calculate a line between major and minor y value in the edge image to rotate each image, trying to put the lobule of the ear in the center. This process is trying to get all the images whose shape is similar to the unknown image. Once the pre-processing is complete, we proceed to compute a Match using the contours of the ear form, with this we are trying to reduce the candidates for the recognition process, this task is performed using the Hausdorff distance (see Fig. 4).

3.2. Application of the Hausdorff distance

The Hausdorff distance measure used in this document is based on the assumption that the ear regions have different degrees of importance where characteristics such as helix, antihelix, tragus, antitragus, concha, lobe and ear contour play the most important role in ear recognition. The algorithm applied is based on what is stated in [15].

In applying the Hausdorff distance, basically operates the comparison of edge maps. The advantage of using edges to match two objects, is that this representation is robust to illumination change. Fig. 5 represents an example of the Hausdorff distance trying to put together two images, the algorithm try to calculate the distance between the points and with this distance we choose a group of image of our database, this task works like a filter choosing and discarding some images in order to strengthen the classification system.

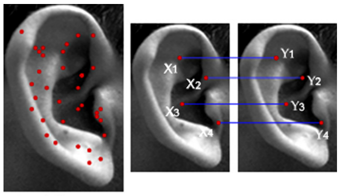

Fig. 6. Example of SURF features.

The procedure involves removing the background of the image as it was performed in the preprocessing original, added some steps after image masking, we proceed to obtain the edges using the Canny and Sobel filter, the image is reversed to operate with a white background, then the ear is binarized, similar procedure is applied to each image stored in the database. With the objects obtained we compare pixels to get how similar are the two figures, as if they were geometric figures performing a comparison process, calculating the Hausdorff distance, we compare pixels to know how similar are the two figures, resulting in a collection of values that contain the distance of the input image with respect to each item in the database.

The object can be presented as an option having the smaller relative distance; if not exceeds the minimum threshold value and identifies the user, otherwise the problem is considered as an unsolved. In the developed system, the Hausdorff algorithm is presented as an complementary preprocessing task to increase the performance of the neural network using the SURF algorithm, if the system procedures identify that the user is the same, the image is accepted to belong to user input identified by all three techniques combined. In this stage we also compute the SURF features to track the ear in video (see Fig. 6).

3.3. Tracking the ear

Speeded Up Robust Features (SURF) [4] is a scale and rotation invariant interest point detector and descriptor. It has been designed for extracting highly distinctive and invariant feature points from images. One of the basic reasons to use SURF for the feature representation is to analyze how the distinctive characteristics works in images, and at the same time is to found more robust with respect to change, taking into account the point of view rotation, scale, illumination and occlusion [4] as compared to other scale and rotation invariant shape descriptors such as SIFT [18] and GLOH [13].

The result for the feature vectors SURF is the relative measured to the dominant orientation to generate each vector that represent an invariant with respect to rotation of the image. The way SURF process pairing is using the most proximate neighbor ratio pairing to get the greatest pairing match for a key-point of a picture inside in another picture is elucidated by detecting the most proximate neighbor in the other key-points from a second picture where the most proximate neighbor is defined as the key-point with the least euclidean distance from the known key-point of the first picture between their characteristic unidirectional matrices. Due to the fact that these SURF vectors are invariant to the image rotation, the process of ear detection combining the previous Viola–Jones approach with the SURF vectors becomes robust and efficient.

Among numerous scale and rotation invariant shape characteristics, SURF [7] offers respectable distinctive features and at the same time it is robust to variations in viewing circumstances, rotations and scales. SURF denotes a picture by first detecting some exclusive feature points in it and then by describing them with the support of a unidirectional feature descriptor.

3.4. Feature extraction and recognition

This section expose the feature extraction algorithms, and the parameter settings that the neural networks require to compute the ear recognition.

3.4.1. Principal Component Analysis (PCA)

The ear recognition algorithm with eigenears is described basically saying that the original images of the training set are transformed into a set of eigenears E. Then, weights are calculated for each image on the (E) set, and then are stored in the (W) set. Observing an image X unknown, weights are calculated for that particular image, and stored in the vector W_X. Subsequently, W_X compared to the weights of images [19].

The process of classifying a new ear in the Γ_{new} to another class (known ears) is the result of two steps. First, the new image is transformed into its eigenear components. The resulting weights forms the weight vector Ω^T_{new}.

$$\begin{aligned} &\omega_k = u_k^T(\Gamma_{new} - \Psi) \quad k = 1, \ldots, M' \\ &\Omega^T_{new} = [\omega_1 \; \omega_2 \; \ldots \; \omega_{M'}] \end{aligned} \tag{1}$$

The euclidean distance between two vectors $d(\Omega_i, \Omega_j)$ provides a measure of similarity between the corresponding images i and j. If the distance between Γ_{new} and the rest of images on average exceeds a certain threshold value, through this can be assumed that Γ_{new} is not a recognizable ear [19].

3.4.2. Linear Discriminant Analysis (LDA)

LDA or fisherears, overcomes the limitations of PCA method by applying the fisher's linear discriminant criterion. The PCA algorithm is a linear combination of functions that maximizes the variance of the information. This can result in poor performance, especially when we are working with image noise such as changes in the background, light and perspective. So the PCA can find faulty components for classifying. To prevent this problem, we implement the fisher algorithm to compare results in the ear recognition process. The fisher algorithm that we implement basically goes like this [5,21]:

We construct the Image matrix x with each column representing an image. Each image is assigned to a class in the corresponding class vector c. Project x into the $(N - c)$ dimensional subspace as P with the rotation matrix $WPca$ identified by a PCA, where N is the number of samples in x. c is unique number of classes $(length(unique(C)))$ and we calculate the between-classes scatter of the projection P as:

$$Sb = \sum_{i=1}^{c} N_i * (mean_i - mean) * (mean_i - mean)^T \tag{2}$$

Where $mean$ is the total mean of P, $mean_i$ is the mean of class i in P, N_i is the number of samples for class i. Then, we need to calculate the within-classes scatter of P as:

$$Sw = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - mean_i) * (x_k - mean_i)^T \tag{3}$$

Where x_i are the samples of class i, x_k is a sample of x_i, $mean_i$ is the mean of class i in P. We apply a standard Linear Discriminant Analysis and maximize the ratio of the determinant of between-class scatter and within-class scatter. The solution is given by the set of generalized eigenvectors $Wfld$ of Sb and Sw corresponding to their eigenvalue. The rank of Sb is utmost $(c - 1)$, so there are only $(c - 1)$ non-zero eigenvalues, cut off the rest. Finally obtain the fisherears by $W = WPca * Wfld$ [21]. These vectors are used as inputs to train a neural network.

3.4.3. SURF algorithm

The ear image is recreated through the SURF algorithm as a set of salient points, where each on is associated with a vector descriptor. Each can be of 64 or 128 dimensions. The 128 dimensional descriptor vector is considered the more exacting feature based in the knowledge that is always best to represent the image

Table 1
Haar-cascade and adding SURF tracking.

	#Attemps	Ear localization (%)	
		Haar-Cascade	With SURF tracking
2D images	308	92.53	98.70
Real time video	314	86.69	95.13

with the most powerful discriminative features possible. A method to obtain a unique characteristic fusion of one sole individual is proposed by combining characteristics acquired from various training instances.

If we have n ear images of an individual for training, a fused prototype is gained by fusing the feature descriptor array of all training images collected, considering the redundant descriptor array only once. Having all the images processed, a collection was made with tags indicating to whom, each image and fusion vector calculated before, belongs. After calculating the SURF features, and filtering the images to use by the Hausdorff distance, the unidirectional characteristic matrices of the ears are deposited in the database.

These vectors are used as inputs to train the network. In the training algorithm, the unidirectional matrices of values belonging to an individual, are taken as positive returning 1 as the neuron output assigned to that user and 0 to other neurons.

When the new image has been captured, the feature vectors are calculated, we compute new descriptors of the unknown ear. These descriptors are entered into the neural network, the outputs of individual neurons are compared, and if the maximum output level exceeds the predefined threshold, then it is determined that the user belongs to the ear assigned to the neuron with the index activated.

The parameter settings of the neural network used in this method are dynamic, the output of neurons depends on Hausdorff Distance filter stage where the algorithm selects some possible answers to the recognition problem in order to reduce the amount of candidates. The hidden layer is created dynamically, respecting that the number of hidden neurons should be between the size of the input layer and the size of the output layer, should be 2/3 the size of the input layer, plus the size of the output layer; and less than twice the size of the input layer based on the research of Jeff Heaton [10].

4. Experimental results

The results obtained in the process of detection and recognition are presented in this section. Table 1 shows the percentages of accuracy when only using the Viola–Jones classifier included in OpenCV vs the potentiation accomplished by adding the tracking with SURF features. This improvement can be seen in 2D images, however, the difference is not so evident, when the process is done on video.

If we take into consideration the time it succeeds in maintaining trying to identify the object, the algorithm combined with SURF tracking is much more accurate because these features allow you to place the image even if it has a 180 degrees event that does not happen with the ears.

Summarizing with perspective and illumination in normal conditions, we get 86% of succeed in recognition with PCA, 87% with fisher algorithm, using the neural network with SURF descriptors, the percentage increased to 92%, over more than 300 attempts of different individuals, but these numbers do not give us enough information to conclude which algorithm is effectively the best to classify our test set. There is a different way of combining precision recall. It is called the f score and it uses:

$$2 * \frac{(Precision * Recall)}{(Precision + Recall)} \tag{4}$$

So, we would tell from these f scores and we will say SURF has the highest f score (94%). LDA-NN has the second highest (84%) and PCA has the lowest (74%) and so, if we go by the f score, we would pick SURF over the others.

Table 2
(%)Performance of conventional PCA vs LDA-NN and SURF-NN.

Training images	Testing images	PCA	LDA-NN	SURF-NN
20	80	73	81	82
30	71	77	83	84
50	87	78	88	84
80	104	83	88	89
100	149	83	89	93
120	186	85	90	94
150	305	86	93	97

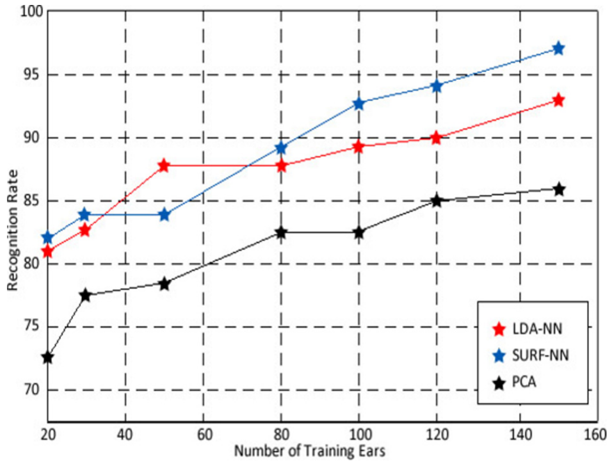

Fig. 7. Recognition rate vs number of training ears.

Table 3
Normal conditions.

	PCA		LDA-NN		SURF-NN	
	Positive	Negative	Positive	Negative	Positive	Negative
Positive	131	49	197	37	269	11
Negative	41	118	38	41	23	107

Table 4
Changing illumination and perspective conditions.

	PCA		LDA-NN		SURF-NN	
	Positive	Negative	Positive	Negative	Positive	Negative
Positive	164	69	169	58	178	27
Negative	89	128	49	111	32	137

In Table 2 we can observe the results in normal conditions with controlled lighting. At this stage we have compared the results obtained with PCA traditional algorithm with our LDA-NN and SURF-NN to check the validity of our work (see Fig. 7). In this sense the results are encouraging, using SURF features as input of a neural network with different test subjects, we get a recognition percentage higher than traditional algorithms in video. Tables 3 and 4 present the resulting confuse matrices.

All algorithms perform perfectly with frontal light, however, the change of perspective made a significant performance difference. The algorithm with less errors classifying the ear when we change the illumination and perspective was the neural network using the SURF descriptors as input. When we change conditions the error rate increase, leaving the PCA algorithm with a f score rate of only 67%, SURF-NN with 86% and LDA-NN in 76%.

5. Conclusion and future work

The integration of two algorithms in the ear recognition system is the main result of this paper. The method that has being used in this research is to try to put together some of the most common approaches in the recognition process, the project is not presented as unique and exceptional, but upon the approaches that other researchers have proposed, combining and comparing them, and trying to select a combination of these approaches to successfully implement a fully functional system capable of recognizing a person across its ear. The first technique is based on the SURF preprocessing followed by a Feed Forward Neural Network based classifier (SURF-NN), and the second is based on the LDA preprocessing (LDA-NN). The feature projection vectors obtained through the SURF and LDA techniques are used as input values in the training and testing stages in both architectures. The recognition performance of SURF-NN is higher than the LDA-NN among the proposed system.

The Neural network using SURF Descriptors as Input appears to be better over variation in lighting. Changes in pre-processing process allow better results specially using Hausdorff Distance as a filter stage. Results have shown that approximately 95.03% of ear recognition accuracy is achieved with a simple 3-layer feed-forward neural network with back-propagation training even if the images contain some noise.

Neural networks provide a great alternative to many other conventional classifiers. This type of algorithms represent powerful tools that can be trained to perform complex tasks and functions in computer vision applications, either in pre-processing tasks, feature extraction and pattern recognition. As future work, the most interesting and useful tool for the police is to achieve the development of an application not only able to propose candidates from the image of an ear, but also to achieve the identification and recognition of a criminal using an earprint. The results of this research are pointing towards that goal, although preliminary, they show a significant progress to approach the final purpose, recognition based on these earprints.

Acknowledgements

This work has been carried out by the project Sociedades Humano-Agente: Inmersión, Adaptación y Simulación, TIN2012-36586-C03-03, Ministerio de Economía y Competitividad (Spain). Project co-financed with FEDER funds.

References

[1] Saeeduddin Ansari, Phalguni Gupta, Localization of ear using outer helix curve of the ear, in: ICCTA'07, 2007, pp. 688–692.
[2] S. Attarchi, K. Faez, A. Raffei, A new segmentation approach for ear recognition, in: Proceedings of International Conference on Advanced Concepts for Intelligent Vision Systems, 2008, pp. 1030–1037.
[3] Victor Barnabas, Kevin Bowyer, Sudeep Sarkar, An evaluation of face and ear biometrics, in: ICPR'02, vol. 1, 2002, pp. 429–432.
[4] Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool, Speeded-up robust features, Comput. Vis. Image Underst. 110 (3) (2008) 346–359.
[5] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern. Anal. Mach. Intell. 19 (7) (1997) 711–720.
[6] Mark Burge, Wilhelm Burger, Ear biometrics in computer vision, in: ICPR'00, vol. 2, 2000, pp. 822–826.
[7] John Bustard, Mark Nixon, 3D morphable model construction for robust ear and face recognition, in: CVPR'10, 2010, pp. 2582–2589.
[8] M. Castrillón-Santana, J. Lorenzo-Navarro, D. Hernández-Sosa, Study on ear detection and its applications to face detection, in: CAEPIA 2011, in: LNCS, vol. 7023, 2011.
[9] A. Cummings, M. Nixon, J. Carter, A novel ray analogy for enrolment of ear biometrics, in: BTAS'10, 2010, pp. 1–6.
[10] J. Heaton, Introduction to Neural Networks for C#, 2nd edition, Heaton Researh, 2010.
[11] David J. Hurley, Mark S. Nixon, John N. Carter, Force field feature extraction ear biometrics, Comput. Vis. Image Underst. 98 (3) (1997) 491–512.
[12] Intel OpenCV wrapper EmguCV, http://www.emgu.com/wiki/index.php/Main_Page, April 08, 2013.
[13] M. Krystian, S. Cordelia, A performance evaluation of local descriptors, IEEE Trans. Pattern Anal. Mach. Intell. 27 (10) (2005) 1615–1630.
[14] Amioy Kumar, Madasu Hanmandlu, Mohit Kuldeep, H.M. Gupta, Automatic ear detection for online biometric applications, in: NCVPRIPG 2011, 2011.

[15] L. Kwan-Ho, L. Kin-Man, S. Wan-Chi, Spatially eigen-weighted Hausdorff distance for face recognition, Pattern Recogn. 36 (8) (2003) 1827–1834.
[16] C. Kyong, W.B. Kevin, S. Sudeep, V. Barnabas, Comparison and combination of ear and face images in appearance-based biometrics, IEEE Trans. Pattern Anal. Mach. Intell. 25 (9) (2003) 1160–1165.
[17] S. Lankton, A. Tannenbaum, Localizing region-based active contours, IEEE Trans. Image Process. 17 (11) (2008) 2029–2039.
[18] David G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[19] Dimitri Pissarenko, Eigenface-based facial recognition, 2002.
[20] USTB Database, University of Science and Technology Beijing.
[21] P. Wagner, Fisherfaces, http://www.bytefish.de/blog/fisherfaces/, January 13, 2013.
[22] Ping Yan, Kevin W. Bowyer, Empirical evaluation of advanced ear biometrics, in: Conf. on Computer Vision and Pattern Recognition, 2005.
[23] Li Yuan, Zhi-Chun Mu, Ear detection based on skin-color and contour information, in: Proceedings of International Conference on Machine Learning and Cybernetics, vol. 4, ICMLC'07, 2007, pp. 2213–2217.
[24] J. Zhou, S. Cadavid, M. Abdelmottaleb, Histograms of categorized shapes for 3D ear detection, in: BTAS'10, Washington, DC, USA, 2010.