Neural Networks Using Hausdorff Distance, SURF and Fisher Algorithms for Ear Recognition

Pedro Luis Galdámez, María Angélica González Arrieta, and Miguel Ramón Ramón

University of Salamanca, Plaza de los Caídos, 37008 Salamanca, Spain
{peter.galdamez,angelica,miguel.ramon}@dgp.mir.es

Abstract. The purpose of this paper is to offer an approach in the biometrics analysis field, using ears to recognize people. This study uses Hausdorff distance as a preprocessing stage adding sturdiness to increase the performance filtering for the subjects to use for testing stage of the neural network. Then, the system computes Speeded Up Robust Features (SURF) and Fisher Linear Discriminant Analysis (LDA) as an input of two neural networks to detect and recognize a person by the patterns of its ear. To show the applied theory in the experimental results; it also includes an application developed with Microsoft .net. The investigation which enhances the ear recognition process showed robustness through the integration of Hausdorff, LDA and SURF in neural networks.

Keywords: Neural Network, Hausdorff, LDA, SURF, Ear Recognition.

1 Introduction

The ear has been used as a mean of human recognition in forensic activities for a long time. During the investigation of several crime scenes, earprints commonly have been used to identify a suspect when there is no information of fingerprints. Recognition systems based on face and ears are very similar, however, the ears have some advantages; for example, their appearance does not change due to expression and is little affected by the aging process. Although the use of information from ear identification 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 behavioural trait can be used as biometric identification mechanism if it is universal, being distinctive and unique to each individual, invariant in time, and measurable automatically or manually; the ear accomplish all these characteristics.

This article is organized as follows: section two presents a brief review of the information in the literature about ear detection and recognition, section three includes a typical ear biometric system which introduce the research in this paper; sections four and five contain a review of various ear detection and recognition methods used. While sections six and seven discuss the results, conclusions and future work.

J.G. de la Puerta et al. (eds.), *International Joint Conference SOCO'14-CISIS'14-ICEUTE'14*, Advances in Intelligent Systems and Computing 299,
DOI: 10.1007/978-3-319-07995-0_24,

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 [18] 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. [9] 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 [21], 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 [22] 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 [17] 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. [23] 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 [27] with an accuracy of 98.05% and 97.05% respectively.

A. Cummings et al. [4] 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 [3], 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 [24] to get the exact location of ear contours, the technique has been tested over 700 ear images. As well as these techniques there are many other significant proposals.

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 [20], principal component analysis (PCA), is also suitable for use in ear recognition. PCA [11] is an orthogonal transform of a dataset which exploits the training data with the propose to find out a set of orthogonal basis vectors or a new axes that causes the projection onto the first axis (principal component) to represent one greatest variance in data, subsequent orthogonal axes to represent decreasing amounts of variance with minimum reconstruction mean square error.

Victor et al. [5] used PCA to perform a 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, they 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, greater depth can be found in the work of Anika Pflug, Christoph Busch [1] and Abaza et al. [2], the next section introduce an intent to solve the problem.

3 Ear Recognition System

Most of ear biometric articles have centered their attention on recognition using manually cropped ear images. This is due to the fact that ear detection of an image face profile is a complicated problem, especially because ear images vary in pose and scale under different conditions. However, for a robust and efficient system is desired to detect the ear from the face profile in an automatic way.

Recognition systems traditionally follow a set of standards, such as, acquiring images, preprocessing, feature extraction, and classification. 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. In this way, the system combines some algorithms that give significant results individually, and when they are combined, achieve a higher degree of robustness with improving in problems such as changes in brightness and perspective. Chart one shows the workflow that the project will follow, in next sections we will deepen in these steps.

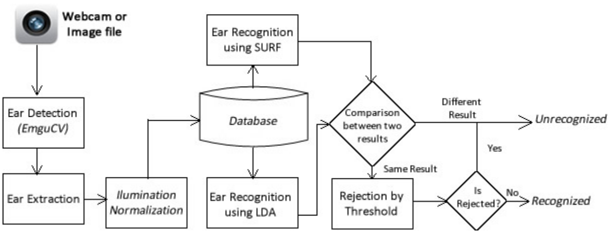

Fig. 1. System flow chart

4 Detecting and Tracking the Ear

There are some techniques which could be used to detect ear automatically. In fact, these techniques usually can detect the ear only when a profile face image do not contain a noisy or big background around the ear. These techniques are not useful, when profile face images are affected by scaling and rotation. This section proposes an useful ear localization technique which attempts to solve these issues.

4.1 Ear Localization

OpenCV and its wrapper for .Net framework EmguCV includes different object detectors based on the Viola-Jones framework, most of them are been constructed to deal with different patterns as frontal face, eyes, nose, etc. Modesto Castellón-Santana et al. [7] have developed a haarcascade classifier to be used with OpenCV to detect ears. This classifier represents a first step to create a robust ear detection and tracking system. The Application is developed in C#.

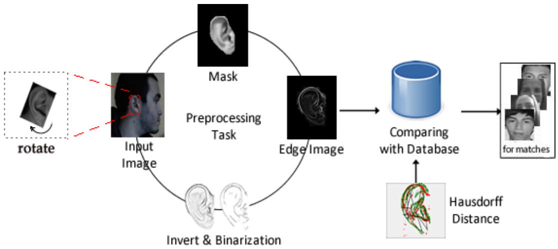

Fig. 2. Image preprocessing

With the ear identified we proceed to perform the preprocessing task, converting the image to gray scale and begin the normalization process, the first step is to perform the segmentation of the image applying a mask 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 pixels and h is the height of the image in pixels, the canny edge detector takes as input an array $w \times h$ of gray values and sigma. 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 and normalize each image, trying to put the lobule of the ear in the centre. This process is to try to get all the images whose shape is similar to the image to identify. We identify some points on the external shape of the ear and the angle created by the center of the line drawn before and the section in the ear's tragus with the major x value.

4.2 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. Accordingly, the edge detection algorithm used will have a significant effect on performance. Figure 2 shows the flow used in the

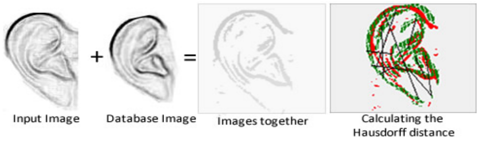

Fig. 3. Hausdorff pre-processing

application of the algorithm, and Figure 3 represent an example of the Hausdorff distance trying to put together two images, the algorithm tries to calculate the distance between the points, this task works like a filter choosing and discarding some images in order to strengthen the classification system.

The procedure involves removing the background of the image to obtain the edges using the Canny and Sobel filter, then, the image is reversed to operate with a white background and the ear is binarized. Similar procedure is applied to each image stored in the database. With the obtained objects we compare pixels to get how similar are the two figures, as if they were geometric figures performing a comparison process, 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 and recognition process using SURF algorithm, if the system procedures identify that the user is the same, even without exceeding the thresholds defined in each process, 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 the video.

4.3 Tracking the Ear

Speeded Up Robust Features (SURF) [10] is a scale and rotation invariant interest point detector and descriptor. It has been designed for extracting highly distinctive and invariant feature points (also called interest points or key-points) from images. One of the basic reasons to use SURF for the feature representation is to analyse 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 and scale, illumination changes and occlusion [10] as compared to other scale and rotation invariant shape descriptors such as SIFT [8] and GLOH [14].

In addition for the extracting SURF features from an image there are two main steps, which describe how to find key points and the calculation of their descriptor vectors. 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.

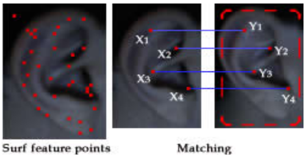

Fig. 4. Example of SURF Features

The way SURF process pairing is using the most proximate neighbour 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 neighbour in the other key-points from a second picture where the most proximate neighbour 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.

Fig. 5. Tracking Ear using SURF Features

The approach to isolate the ear in the image, the prototype we used for the ear identification should reveal the characteristics of scale and rotation immutability. To calculate such prototypes in a suggested method, an invariant shape characteristic to rotation and scale was used. Among numerous scale and rotation invariant shape characteristics, SURF [12] 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 detecting some exclusive feature points in it and then by describing them with the support of a unidirectional feature descriptor matrix.

5 Ear Recognition Using Neural Networks

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 preprocessing tasks, feature extraction and pattern recognition.

Fig. 6. Avila's Police School Database

Two neural networks are used in the system, the first one based on the SURF algorithm and the second using a classification based on LDA, both networks have been trained and proven using the database of the Police College of Ávila (figure 6). The training was performed using 3 poses of the ear of each person and the tests were done with 10-n poses of the same people. After calculating the features using SURF the projection vectors are calculated for the training set and then used to train the network. Similarly, after the calculation of the LDA projection vectors, the second neural network is trained.

5.1 SURF Neural Network

The Ear Image is recreated as a set of salient points, where each on is associated with a vector descriptor. Each can be of 64 or 128 dimensions. If 128 dimensional vector is chosen, It is more exacting in comparison to the 64 vector. So the 128 dimensional descriptor vector is considered the most exacting feature based in the knowledge that is always best to represent the image with the most powerful discriminative features.

A method to obtain unique characteristic fusion of one sole individual is proposed by combining characteristics acquired from various training instances. If we have n ear images 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. We had to use a small database made for the training purpose with 309 pictures matching to 3 ear captures from 103 persons. Having all the images processed, a collection was made with their respective tags describing the images and fusion vector indicating to whom the image belongs. After calculating the SURF features, and filtering the images by the Hausdorff distance, the unidirectional characteristic matrices are deposited in the database.

5.2 Linear Discriminant Analysis Neural Network

Linear Discriminant analysis or Fisherears method in our case, overcomes the limitations of PCA method by applying the Fisher's linear discriminant criterion. This criterion tries to maximize the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class.

So the PCA algorithm can find faulty components for classifying especially when we are working with image noise such as changes in the background, light and perspective. To prevent this problems, we implement the Fisher algorithm to compare results in the ear recognition process. The Fisher algorithm that we implement basically goes like the version exposed in [6,26].

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. Then, we proceed to 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 $(length(unique(C)))$

In the next step we calculate the between-classes scatter of the projection P as $Sb = \sum_{i=1}^{c} N_i * (mean_i - mean) * (mean_i - mean)^T$ 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

Also, we proceed to calculate the within-classes scatter of P using the next formula $Sw = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - mean_i) * (x_k - mean_i)^T$ 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 almost $(c-1)$, so there are only $(c-1)$ non-zero eigenvalues, cut off the rest. Finally we obtain the fisherears by $W = WPca * Wfld$ [26].

These vectors are used as inputs to train our neural network. In the training algorithm, the unidirectional vectors 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, we compute new descriptors. 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.

6 Experimental Results

The results obtained in the process of detection and recognition of the ear 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. That can be seen in 2D images or photographs the difference are not so evident, however when the process is done on video, the difference is almost 10 percentage points, and is only done when considering the location of the ear in the video in different pose and lighting conditions. 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.

Table 1. Ear Detection (Haar-Cascade and adding SURF Tracking)

	#Attemps	*EarLocalization*(%)	
		Haar − Cascade	*With SURF tracking*
2D Images	308	92.53	98.70
Real Time Video	314	86.69	95.13

In Table 2 we can observe the results of the recognition process and system performance. At this stage we have compared the results obtained with traditional algorithms such as PCA and our propose using the two neural networks with SURF and LDA to check the validity of our work. 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 the traditional algorithms in video. Summarizing with perspective and illumination in normal conditions, we get 86% of succeed in recognition with PCA, 93% with LDA-NN algorithm, using the neural network with SURF descriptors, the percentage increased to 97%, over more than 300 attempts of different individuals.

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

The method that has been 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 and use this system to identify criminals. The techniques studied provide a clearer picture to where should point this research in the future, observing some of the strengths and weaknesses of the algorithms proposed in order to strengthen preprocessing tasks.

7 Conclusion and Future Work

The integration of two algorithms is the main result of this paper. 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 with another Feed Forward Neural Network (LDA-NN). The feature projection vectors obtained are used as input values in the training and testing stages in both architectures. The proposed system shows improvement on the recognition rates over the conventional Fisher and PCA ear recognition that use the Euclidean Distance based classifier. Additionally, the recognition performance of SURF-NN is higher than the LDA-NN among the proposed system as shown in figure 7.

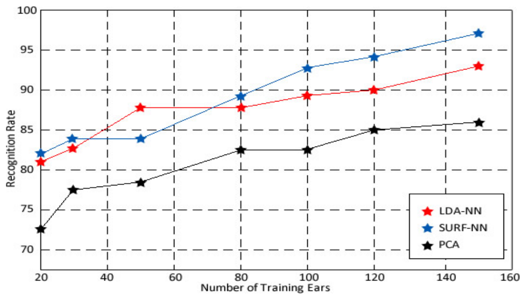

Fig. 7. Recognition rate vs number of Training ears

The Neural network using SURF Descriptors appears to be better over variation in lighting. The LDA-NN and SURF-NN perform better than the PCA over changes on illumination and perspective. Changes in preprocessing process allows 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 contains some noise.

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, they show a significant progress to approach the final purpose, recognition based on these earprints.

References

1. Pflug, A., Busch, C.: Ear Biometrics: A Survey of Detection, Feature Extraction and Recognition Methods. IET Biometrics 1(2), 114–129 (2012)
2. Ayman, A., Arun, R., Christina, H., Harrison, F., Ann, M., Nixon, M.S.: A Survey on Ear Biometrics. ACM Computing Surveys 45 (2013)
3. Kumar, A., Hanmandlu, M., Kuldeep, M., Gupta, H.M.: Automatic ear detection for online biometric applications. In: NCVPRIPG, pp. 146–149 (2011)
4. Cummings, A., Nixon, M., Carter, J.: A novel ray analogy for enrolment of ear biometrics. In: BTAS, pp. 1–6 (2010)
5. Victor, B., Bowyer, K., Sarkar, S.: An evaluation of face and ear biometrics. In: ICPR, vol. 1, pp. 429–432 (2002)
6. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear Projection, New Haven. Yale University (1997)
7. Castrillón-Santana, M., Lorenzo-Navarro, J., Hernández-Sosa, D.: An Study on Ear Detection and Its Applications to Face Detection, pp. 313–322 (2011)
8. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
9. Hurley, D.J., Nixon, M.S., Carter, J.N.: Force field feature extraction for ear biometrics. Computer Vision and Image Understanding (2005)
10. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (SURF). Computer Vision and Image Understanding, 346–359 (2008)
11. Jolliffe, I.T.: Principal Components Analysis, 2nd edn. Springer, New York (2002)
12. Bustard, J., Nixon, M.: 3D morphable model construction for robust ear and face recognition. In: CVPR, pp. 2582–2589 (2010)
13. Calvo-Rolle, J.L., Corchado, E.: A Bio-inspired knowledge system for improving combined cycle plant control tuning. Neurocomputing 126, 95–105 (2014)
14. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence (2005)
15. Lin, K.-H., Lam, K.-M., Siu, W.-C.: Spatially eigen-weighted Hausdorff distances for human face recognition. Polytechnic University, Hong Kong (2002)
16. Chang, K., Bowyer, K.W., Sarkar, S., Victor, B.: Comparison and combination of ear and face images in appearance-based biometrics (2003)
17. Yuan, L., Mu, Z.-C.: Ear detection based on skin-color and contour information. In: ICMLC, vol. 4, pp. 2213–2217 (2007)
18. Burge, M., Burger, W.: Ear biometrics in computer vision. In: ICPR, vol. 2, pp. 822–826 (2000)
19. Woźniak, M., Graña, M.: Emilio Corchado A survey of multiple classifier systems as hybrid systems. Information Fusion 16, 3–17
20. Turk, M., Pentland, A.: Eigenfaces for Recognition (1991)
21. Yan, P., Bowyer, K.W.: Empirical evaluation of advanced ear biometrics. In: Proceedings of International CVPR Workshop, vol. 3, pp. 41–48 (2005)
22. Ansari, S., Gupta, P.: Localization of ear using outer helix curve of the ear. In: ICCTA, pp. 688–692 (2007)
23. Attarchi, S., Faez, K., Rafiei, A.: A new segmentation approach for ear recognition. In: Blanc-Talon, J., Bourennane, S., Philips, W., Popescu, D., Scheunders, P. (eds.) ACIVS 2008. LNCS, vol. 5259, pp. 1030–1037. Springer, Heidelberg (2008)
24. Lankton, S., Tannenbaum, A.: Localizing region-based active contours. IEEE Transactions on Image Processing, 2029–2039 (2008)
25. Prakash, S., Gupta, P.: An Efficient Ear Localization Technique (2012)
26. Wagner, P.: Fisherfaces, `http://www.bytefish.de/blog/fisherfaces/` (January 13, 2013)
27. USTB Database, University of Science and Technology Beijing