Journal of Applied Logic • • • (• • • •) • • •–• • •

Contents lists available at ScienceDirect

Journal of Applied Logic

www.elsevier.com/locate/jal

A brief review of the ear recognition process using deep neural networks

Pedro Luis Galdámez *, William Raveane, Angélica González Arrieta

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

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Ear recognition
Convolutional Neural Network
CNN

ABSTRACT

The process of precisely recognize people by ears has been getting major attention in recent years. It represents an important step in the biometric research, especially as a complement to face recognition systems which have difficult in real conditions. This is due to the great variation in shapes, variable lighting conditions, and the changing profile shape which is a planar representation of a complex object. An ear recognition system involving a convolutional neural networks (CNN) is proposed to identify a person given an input image. The proposed method matches the performance of other traditional approaches when analyzed against clean photographs. However, the F1 metric of the results shows improvements in specificity of the recognition. We also present a technique for improving the speed of a CNN applied to large input images through the optimization of the sliding window approach.

1. Introduction

The study of biometrics is based on physical or behavioural characteristics of an individual in order to verify his or her identity. Features like fingerprints, face, and iris have received major attention for long time. Researchers consider fingerprints and iris are more precise in biometric analysis than the face, but the face has other qualities like being easily obtained in real conditions without user interaction. However the face by itself is not as flexible as it should be due to illumination and expression changes.

Ear images can be obtained with the same approach like the face, this scenario suggests that it could be used as a complement in a recognition system. Multiple researchers have affirmed that the ears are indeed unique enough to identify a person and they could have a practical use as a biometric feature [17].

2. Background

Fundamentally most of the ear detection approaches rely on properties in the ear's morphology, like the occurrence of certain characteristic edges or frequency patterns. Significant progress has been made in the

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

http://dx.doi.org/10.1016/j.jal.2016.11.014
1570-8683/

past few years in the ear biometrics field. One important technique known in ear detection was introduced by Burge and Burger [1]. They proposed a technique for detection using deformable contours, the main problem is that this method requires user interaction for contour initialization; therefore, the task of localization is not fully automatic. Hurley et al. [10] applied force fields, this process does not require to know the location of the ear to perform detection; however, this only applies in controlled environments without any kind of noise.

One of the most impressive techniques known to detect the ears is raised by A. Cummings et al. [8] who show a strategy using the image ray transform (IRT) which is capable of highlighting the ear tubular structures. The technique exploits the helix elliptical shape to calculate the localization. In [19], 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. In the context of 3D images, Zhou et al. [20] presented a novel shape based feature set, called Histograms of Categorized Shapes (HCS), for robust 3D ear detection. Using a sliding window approach and a linear Support Vector Machine (SVM) classifier.

Chen and Bhanu propose three different approaches for ear detection. In the first of these, they trained a classifier that recognizes a specific distribution of shape indices [4]; however, this only works on profile images and is very sensitive to any kind of rotation, scale or pose variation in the image. Later, they worked on image regions with a large local curvature using a technique known as step edge magnitude [3], where a template containing the typical shape of the outer helix and the anti-helix, is fitted to clusters of lines. Finally, they narrowed the number of possible ear candidates by detecting the skin region; first before applying the helix template matching to the curvature lines [5]. There are many proposals to solve the problem, this paper only has done a small review from some of them, in order to deepen about the literature review in the ear biometrics researches you can refer the work of Pflug et al. [15].

The Convolutional Neural Network (CNN) [7] has become a general solution for image recognition with variable input data. CNNs consist of two stages one for automated feature learning, and another for classification, both of them can be successfully trained in tandem through gradient descent of the error surface [14]. Its results have consistently outclassed other machine learning approaches in large scale image recognition tasks [11], outperforming even human inspection of extensive datasets [6].

Compared to other feature-based computer vision methods such as SIFT [13] or HOG [7], CNNs are much more robust and tolerant to shape and visual variations of the images or objects intended to be recognized. However, contrary to such methods, an execution of a CNN will only recognize features on a single image block of size equal to the input dimensions of the network. As CNNs are usually trained with small image patches, this recognition area is likewise small. As a result, to run image recognition over a larger image size, it is necessary to repeatedly apply the same network over multiple regions. This is a very common technique named sliding windows, albeit a time consuming one as the execution time naturally grows in proportion to the number of sampled blocks.

3. System description

The network on which our system is based upon is a standard CNN composed of alternating convolutional and max-pooling layers for the feature extraction stage, and one or more linear layers for the final classification stage. Fig. 1 depicts the layer structure of such a network, and it is the reference architecture used here to describe the concepts of the framework presented. The first layer in the network consists of one or more neurons containing the image data to be analyzed, usually composed of a single grayscale channel of the incoming image.

Recognition systems traditionally follow a set of standards, such as, acquiring images, pre-processing, feature extraction, 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,

Please cite this article in press as: P.L. Galdámez et al., A brief review of the ear recognition process using deep neural networks, J. Appl. Log. (2016), http://dx.doi.org/10.1016/j.jal.2016.11.014

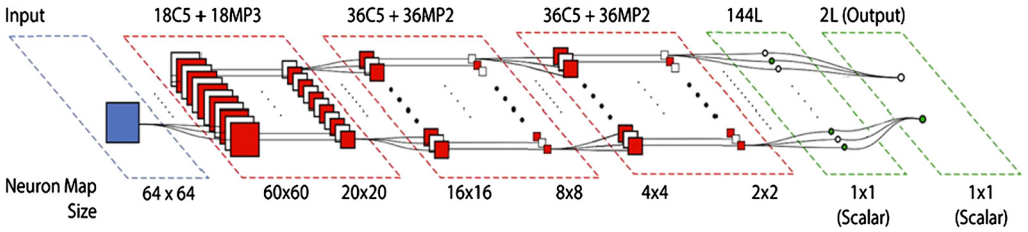

Fig. 1. Architecture of the Convolutional Neural Network used in the system.

it is important to notice that the process 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 clean images using identification through the ear.

The convolutional neural network used is a very standard one which has a few considerations made in its architecture that help for our particular use case. Given that our end purpose for this system is the recognition of ears in video streams, we require a system that executes rapidly. Therefore, we require a highly optimized architecture which is still useful to recognize the trained data.

The target classes we seek to recognize with the neural network are only two: (i) User, (ii) Other Users, and Background – from herein referred to by their abbreviations US and BG. The intra-class data is, ultimately, very similar with all ears following a similar set of distinctive features. As a result, the network does not need to learn an unbounded amount of different features as would be the case in the large CNNs that are utilized for large general image recognition. Instead, a modestly sized neural network, with a limited number of neurons turns out to be quite sufficient to learn the type of data required for this task. The approach followed in our strategy is one-to-many, it means that the system creates one neural network by individual, this is applicable only in a controlled scenario with few users, like the research that we are conducting, where our goal is to identify if it is indeed possible to use this kind of networks in combining with other algorithms in self-contained experiments.

Given the nature of the images, it was decided that an optimum input size for the network would be 64×64, as ear images of that size are large enough to carry enough identifying information to properly define the shape of an ear, but they are not too large that would require too large convolution kernels or pooling layers to properly analyze.

Lastly, given the requirement to do sliding windows over full images through the Shared Maps method, the maximum accumulated pooling factor should not be too large, such that the output window map skips large portions of the input image per window. Therefore, a maximum of 3 convolutional + pooling layers was decided upon.

3.1. The sliding window method

Recognition of images larger than the NN input size is achieved by the sliding window approach (see Fig. 2). This algorithm is defined by two quantities, the window size S, usually fixed to match the NN's designed input size; and the window stride T, which specifies the distance at which consecutive windows are spaced apart. This stride distance establishes the total number of windows analyzed W for a given input image. For an image of size $I_w \times I_h$, the window count is given by:

$$W = \left(\frac{I_w - S}{T} + 1\right)\left(\frac{I_h - S}{T} + 1\right) \quad \Longrightarrow \quad W \propto \frac{I_w I_h}{T^2} \tag{1}$$

Fig. 2 shows this method applied on an input image downsampled, extracting windows of $S = 32$ for the simple case where $T = S/2$. A network analyzing this image would require 40 executions to fully analyze all

Please cite this article in press as: P.L. Galdámez et al., A brief review of the ear recognition process using deep neural networks, J. Appl. Log. (2016), http://dx.doi.org/10.1016/j.jal.2016.11.014

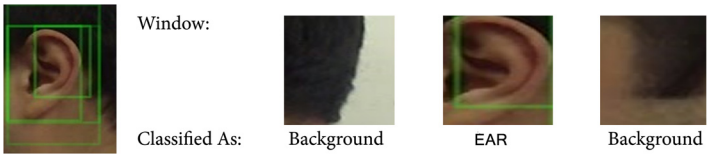

Fig. 2. An overview of the sliding window method.

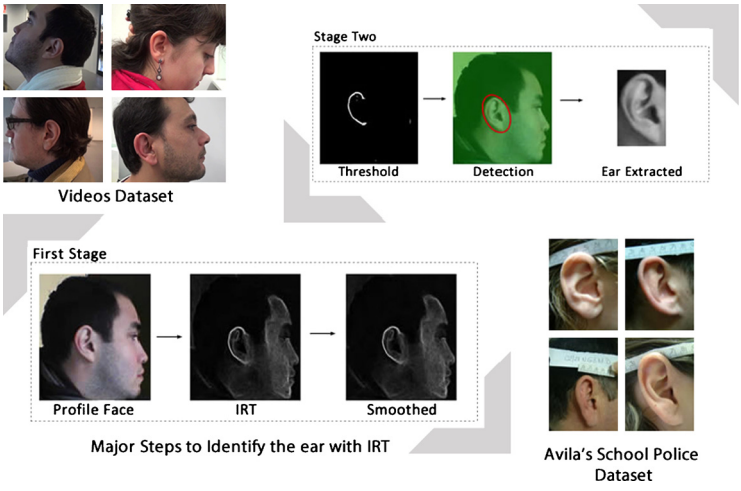

Fig. 3. Datasets and Major Steps to identify the ear with IRT.

extracted windows. The computational requirement is further compounded when a smaller stride is selected – an action necessary to improve the resolving power of the classifier: at $T = S/8$.

3.2. Detecting

The Recognition step is developed by CNNs but also it is important to highlight the difficulty in precisely detecting and locating an ear within an image which is the first step to tackle in an ear-based biometric recognition system. We have been worked with different object detectors based on the Viola–Jones framework, most of them have been constructed to deal with different patterns like frontal face, eyes, nose, etc. Modesto Castellón-Santana et al. [2] have developed a haarcascade 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 even, although it has major problems in uncontrolled conditions like change of viewpoint. The purpose of integrating a classifier is to improve the timing of the sliding window algorithm to make it fast and only if the ear has not been detected execute the rest of the process.

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 can not identify the ear, the system will compute the IRT (see Fig. 3). The original images from the Ávila's Police School database created for this research with 1200 ears, 12 images per subject undergoes using the technique proposed by Cummings et al. [8] 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. This approach works very well on clean images in defined environments but has major issues in no collaborative situations.

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 to extract only the ear, then the image is converted to an edge map using the canny edge filter. If w is the

Please cite this article in press as: P.L. Galdámez et al., A brief review of the ear recognition process using deep neural networks, J. Appl. Log. (2016), http://dx.doi.org/10.1016/j.jal.2016.11.014

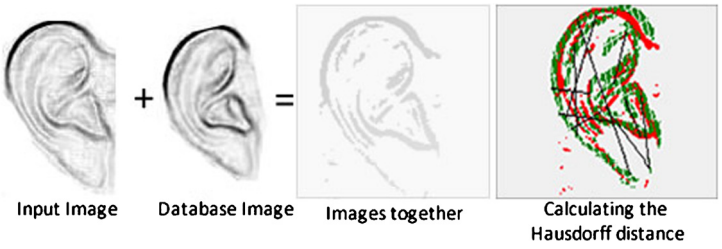

Fig. 4. Hausdorff pre-processing.

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.

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 centre. 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.

3.3. Application of the Hausdorff distance

On the assumption that the ear has regions with different degrees of importance where features such as helix, antihelix, tragus, lobe, and the ear contour play an important role in the process of ear recognition (see Fig. 4). The algorithm here introduced is based on the work presented in [12].

The Hausdorff distance operates comparing edge maps. The advantage of using this maps to match templates, is that the representation is robust to illumination change. Fig. 4 represents an instance of the algorithm trying to put together two ears, the technique tries to calculate the distance between the points in order to choose a group of images of our database, this task works like a filter discarding some images to strengthen the classification.

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.

4. Experimental setup

In order to test our approach we basically test the deep neural network vs a standard feed forward neural network fusioned with traditional algorithms like PCA, LDA and SURF Features. In fact we collect 44 videos with a time average of 90 seconds and 30 frames per seconds, all of them taken in the controlled setup with the same illumination and perspective, this task gave us 118,800 ears of 11 subjects.

Please cite this article in press as: P.L. Galdámez et al., A brief review of the ear recognition process using deep neural networks, J. Appl. Log. (2016), http://dx.doi.org/10.1016/j.jal.2016.11.014

The parameter settings of the neural network used in this method are dynamic, the output 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 [9].

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 [16].

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}.

$$\omega_k = u_k^T(\Gamma_{new} - \Psi) \quad k = 1, \ldots, M'$$
$$\Omega^T_{new} = [\omega_1\ \omega_2\ \ldots\ \omega_{M'}] \tag{2}$$

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 [16]. After the eigen vectors are computed they are stored to use them as an input to the feed forward neural network.

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 [18]:

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:

$$Sb = \sum_{i=1}^{c} N_i * (mean_i - mean) * (mean_i - mean)^T \tag{3}$$

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:

$$Sw = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - mean_i) * (x_k - mean_i)^T \tag{4}$$

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

Please cite this article in press as: P.L. Galdámez et al., A brief review of the ear recognition process using deep neural networks, J. Appl. Log. (2016), http://dx.doi.org/10.1016/j.jal.2016.11.014

Table 1
Average confusion matrix of the training data fold.

Classified as/ real class	User identified correctly	Background	Total in class	Accuracy (%)
User identified correctly	26071	329	26400	98.75
Background	190	26210	26400	99.28
		Total	52800	99.02

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 at most $(c-1)$, so there are only $(c-1)$ non-zero eigenvalues, cut off the rest. Finally obtain the fisherears by $W = WPca * Wfld$ [18]. These vectors are used as inputs to train the neural network in the same way that as the previous section.

4.3. Speeded up robust features (SURF)

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 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.

5. Experimental results

The experimental configuration of this research, was built with the purpose of identifying whether it is possible to use deep neural networks in the recognition of people by ears through images. Preliminary results are encouraging, Tables 1 and 2 show the average of the confusion matrix for the data training and testing of the 17 deep neural networks, built for the same number of individuals in order to perform recognition among a large number of negative data in an approach of one to many.

In each of the matrices, it can be seen the significant difference between users correctly classified versus the false negatives and false positives. Although the data are encouraging, it is important to note that training of deep neural networks were performed using images taken under the same lighting conditions and outlook, with similar quality, same size of the ear; summarizing, without any kind of noise. This setting determines the effectiveness of the system in no collaborative environments.

As part of the evaluation process was taken 44 videos, which comprise 90 seconds on average, where approximately 60 seconds maintains the conditions described above, and 30 seconds of the video shows changes of perspective, each video has 30 frames per second. It is logical to imagine that the use of images

Please cite this article in press as: P.L. Galdámez et al., A brief review of the ear recognition process using deep neural networks, J. Appl. Log. (2016), http://dx.doi.org/10.1016/j.jal.2016.11.014

Table 2
Average confusion matrix of the testing data fold.

Classified as/ real class	User identified correctly	Background	Total in class	Accuracy (%)
User identified correctly	12890	310	13200	97.65
Background	211	12989	13200	98.40
		Total	26400	98.03

Table 3
Recognition performance of our CNN system vs PCA, LDA and SURF – NN.

	Avila's Police School Dataset				Bisite Videos Dataset			
	Precision (%)	Recall (%)	Accuracy (%)	F1 score (%)	Precision (%)	Recall (%)	Accuracy (%)	F1 score (%)
CNN	97.71	80.85	84.82	88.48	79.17	43.65	40.13	56.27
PCA-NN	98.00	61.03	66.37	75.22	48.38	28.46	21.83	35.84
LDA-NN	98.40	69.13	68.36	81.21	52.31	36.47	27.37	42.98
SURF-NN	98.95	77.39	76.75	86.85	53.18	42.03	30.68	46.95

under the conditions described above affects all algorithms that have been tested here, the Table 3 shows a comparative analysis between the two datasets used in this paper, where it can be clearly observed how the F1 score decreases drastically when a random sample of ears in all the videos are choosing. Showing that for the proper operation of the deep neural network is necessary to train with data which must be as similar as possible with the future objects to recognize.

It is important to highlight the differences between the F1 score in ideal conditions and when the dataset is highly fuzzy. The result is not really significant. However, maintains an important positive difference, when images are closer to the real world and although its accuracy may be considered disappointing, shows that has an important potential of improvement.

6. Conclusion and future work

In this paper we have presented a new approach to ear recognition based on convolutional neural networks. CNNs work by image and shape perception, which tends to be a much more robust approach than traditional computer vision systems which rely on hand-crafted features to be manually defined for each specific task. This is mostly the case when tested against variable perceptions and occlusions, both of which are very common occurrences in natural images featuring a person's ear, indeed this affirmation represent a major future line because as it has been seen along the present document it is a huge problem.

This article can be considered as a first approach to the use of convolutional neural networks for ear recognition. It represents an important step, because such networks, are taken very relevant position in the area of computer vision. Future lines of research point to, for instance, to improve the detection process, followed by using a complete set of data in real conditions for the training, enhance the architecture and compare the results against robust algorithms. The usefulness of this research is beyond the results, demonstrating the viability of using deep neural networks in ear recognition. In future researches the purpose is attempt to obtain access to different public datasets and some robust measure able to take into account the effect of the chance in order to use them to test a unique convolutional neural network which performs all the process here introduced.

References

[1] Mark Burge, Wilhelm Burger, Ear biometrics in computer vision, in: ICPR'00, vol. 2, 2000, pp. 822–826.
[2] M. Castrillón-Santana, J. Lorenzo-Navarro, D. Hernández-Sosa, Study on Ear Detection and Its Applications to Face Detection, 2011.

Please cite this article in press as: P.L. Galdámez et al., A brief review of the ear recognition process using deep neural networks, J. Appl. Log. (2016), http://dx.doi.org/10.1016/j.jal.2016.11.014

[3] H. Chen, B. Bhanu, Contour matching for 3d ear recognition, in: Proceedings of the Seventh IEEE Workshop on Applications of Computer Vision (WACV/MOTION), 2005, 38, 39, 50, 51, 52.
[4] H. Chen, B. Bhanu, Shape model-based 3d ear detection from side face range images, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition – Workshops (CVPR), June 2005, p. 122, 38, 39, 155.
[5] H. Chen, B. Bhanu, Human ear recognition in 3d, IEEE Trans. Pattern Anal. Mach. Intell. 29 (4) (April 2007) 718–737, 38, 39, 50, 67, 68, 80.
[6] D.C. Ciresan, U. Meier, L.M. Gambardella, J. Schmidhuber, Convolutional Neural Network Committees for handwritten character classification, in: 11th International Conference on Document Analysis and Recognition, ICDAR, 2011.
[7] D.C. Ciresan, U. Meier, J. Masci, L.M. Gambardella, J. Schmidhuber, Flexible, high performance convolutional neural networks for image classification, in: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, 2011, pp. 1237–1242.
[8] A. Cummings, M. Nixon, J. Carter, A novel ray analogy for enrolment of ear biometrics, in: BTAS' 10, 2010, pp. 1–6.
[9] J. Heaton, Introduction to Neural Networks for C#, 2nd edition, 2010.
[10] David J. Hurley, Mark S. Nixon, John N. Carter, Force field feature extraction ear biometrics, Comput. Vis. Underst. (2005).
[11] A. Krizhevsky, I. Sutskever, G.E. Hinton, in: ImageNet Classification with Deep Convolutional Neural Networks, vol. 25, 2012, pp. 1106–1114.
[12] L. Kwan-Ho, L. Kin-Man, S. Wan-Chi, Spatially Eigen-Weighted Hausdorff Distance for Face Recognition, Hong Kong, 2002.
[13] S. Lazebnik, C. Schmid, J. Ponce, Beyond bags of features: spatial pyramid matching for recognizing natural scene categories, in: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, CVPR 2006, 2006, pp. 2169–2178.
[14] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE (November 1998).
[15] A. Pflug, C. Busch, Ear biometrics: a survey of detection, feature extraction and recognition methods, IET Biometrics 1 (2) (2012) 114–129.
[16] Dimitri Pissarenko, Eigenface-Based Facial Recognition, 2002.
[17] Barnabas Victor, Kevin Bowyer, Sudeep Sarkar, An evaluation of face and ear biometrics, in: ICPR'02, vol. 1, 2002, pp. 429–432.
[18] P. Wagner, Fisherfaces, http://www.bytefish.de/blog/fisherfaces/, January 13, 2013.
[19] Ping Yan, Kevin W. Bowyer, Empirical evaluation of advanced ear biometrics, in: Conf. on Computer Vision and Pattern Recognition, 2005.
[20] J. Zhou, S. Cadavid, M. Abdelmottaleb, Histograms of Categorized Shapes for 3D Ear Detection, BTAS, Washington-DC, USA, 2010.

Please cite this article in press as: P.L. Galdámez et al., A brief review of the ear recognition process using deep neural networks, J. Appl. Log. (2016), http://dx.doi.org/10.1016/j.jal.2016.11.014