EAR RECOGNITION SYSTEM BASED ON A NEW PROPOSED FEATURE EXTRACTION ALGORITHM

DR. ANNA PURANI ,K1, INDRANIL BANERJEE1, APARAJITA SINGH1
1Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, India

ABSTRACT:

Now-a-days, security of data and valuable articles is of utmost importance. There is an increased need for efficient and reliable security systems which can monitor and authenticate people who attempt to access a location. Various biometric traits have been used in security systems. The research on using the human ear as a biometric has gained momentum in the last decade. The ear is a good choice for a biometric because it is unique and does not alter much with age. In this work, the ear features tragus and anti-tragus, are extracted using the new proposed algorithm for feature extraction. The algorithm consists of creating a boundary of the ear and finding out a line (L1) with the maximum length between the coordinates of the boundary. The tragus is located at the middle point of L1. A mirror of the line (L2) is drawn. The antitragus is located at the middle point of L2. The extracted features are stored as templates for the next stage of ear recognition which is done using a convolution neural network. The accuracy obtained using tragus is 94.18% and using anti-tragus is 73.56%.

Keywords: Biometric, Feature Extraction, Tragus, Anti-Tragus, Convolution Neural Network, Human Ear, Accuracy

1. INTRODUCTION

Biometrics is a method for calculating physical and behavioral characteristics of the human body to extract unique human features. Biometrics can be used for identification and authentication. Therefore, it finds a place in surveillance and security systems. Biometric systems are a better method of authentication than simple alphanumeric passwords or pin codes [11]. Systems that incorporate biometric authentication are more reliable because biometric features are unique and generally cannot be duplicated or stolen. Biometrics was first studied and used as a technique to solve crime by Alphonse Bertillon [24]. His works were extended further by Alfred Iannarelli [25] who worked on ear feature extraction techniques.

In this section, advantages of ear as a biometric, the selection of ear features and the dataset are discussed. In the second section, a summary of the past and recent developments is given. In the third section, we propose an ear recognition system which consists of three stages - ear detection, ear feature extraction and ear recognition. We propose a new algorithm for ear feature extraction. In the fourth section, the results of the recognition phase are analyzed. In the sixth section and seventh section, the conclusion and potential future work are stated.

1.1 Advantages of Ear as Biometric

Ear is a comparatively new trait used for biometric systems. Any human physical or behavioral biometric trait needs to be unique. The human ear has been investigated in terms of its uniqueness and it has been found that it is unique across humans [3]. The technique used for testing the uniqueness of the ear by Ruma Purkait is based on feature vector representation in a seventeen dimensional feature space [3]. The findings of the investigation prove that the ear is a unique physical trait. Moreover, the ear exhibits permanence which makes it an appropriate choice for biometric systems. There are several advantages of using the ear rather than using other features such as face, fingerprint, retina and iris. The ear structure does not change
significantly throughout a person’s lifetime [7]. Therefore, there is no need to frequently update the database for ears. The surface of the ear is small as compared to the face which makes it easier to work with for image processing [11] [7]. At the same time, the ear is not as small as the iris. The iris requires high definition cameras to capture a clear image whereas the features of the ear can be captured using a normal camera. The face of a human varies with age, expression, pigmentation and illumination. The ear has a more uniform color distribution thereby conserving all information which is essential for processing of the image of the ear [14]. Moreover, the background of the ear is predictable i.e. it will always be at the side of a head, which makes it easier to detect. Furthermore, face recognition performs relatively poorly when the face is at an angle or when it is captured from the side view, while the ear recognition generally performs well even when the ear is captured from an angle [6]. The advantages of using ear as a biometric depicted in Fig. 1.

![Fig. 1. Advantages of Ear as a Biometric](image)

The outer ear has many features as shown in Fig 2. For the purpose of our project, we have selected two features - the tragus and the antitragus. The features on the top of the ear - helix and scaphoid fossa - are usually covered by hair or caps and therefore, these features are difficult to extract. Features such as crus of helix, crura of antihelix and concha cava do not have a clear identified portion and therefore cannot be used. The lobule cannot be used as it is occluded by jewellery and does not have a clear boundary which can be extracted. The tragus is selected since it is one of the least occluded portions of the ear. Also, it has a clear boundary which can be extracted. Sometimes there might be hair present near the area of the tragus. Therefore, we select another feature, antitragus, to make the system more robust. The antitragus is also generally not occluded and can be extracted since it has a clear boundary.

1.3 Dataset used

The Mathematical Analysis of Images (AMI) Ear Database [20] is used for the purpose of this project. It was created by Esther Gonzalez. The database consists of 700 images taken under the same lighting in an indoor environment. For each
subject, there are 7 images in different orientations. One image is of the left ear while the other 6 are of the right ear. There are some images with occlusions such as hair, jewellery and spectacles as shown in Fig 3.

2. Summary of Past and Recent Developments

The human ear has been proven to be a unique trait therefore it can be used as a biometric. The ear has various advantages over other biometric traits. Different ear feature extraction methods were studied.

Table 1 Accuracy obtained for various databases

<table>
<thead>
<tr>
<th>Database</th>
<th>Publication</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMI [20]</td>
<td>[1]</td>
<td>99.92%</td>
</tr>
<tr>
<td></td>
<td>[9]</td>
<td>100%</td>
</tr>
<tr>
<td>IIT Delhi [22]</td>
<td>[5]</td>
<td>75% to 90%</td>
</tr>
<tr>
<td></td>
<td>[12]</td>
<td>96.3%</td>
</tr>
<tr>
<td></td>
<td>[1]</td>
<td>99.92%</td>
</tr>
<tr>
<td></td>
<td>[9]</td>
<td>98%</td>
</tr>
<tr>
<td></td>
<td>[15]</td>
<td>98%</td>
</tr>
<tr>
<td>NIST [21]</td>
<td>[11]</td>
<td>84.3% to 91.2%</td>
</tr>
<tr>
<td>UND [18]</td>
<td>[13]</td>
<td>96.1%</td>
</tr>
<tr>
<td></td>
<td>[10]</td>
<td>97.5%</td>
</tr>
<tr>
<td></td>
<td>[6]</td>
<td>93.34%</td>
</tr>
<tr>
<td></td>
<td>[15]</td>
<td>98%</td>
</tr>
<tr>
<td>USTB [19]</td>
<td>[4]</td>
<td>93.33%</td>
</tr>
<tr>
<td></td>
<td>[17]</td>
<td>87%</td>
</tr>
<tr>
<td></td>
<td>[6]</td>
<td>93.34%</td>
</tr>
<tr>
<td></td>
<td>[16]</td>
<td>94.73%</td>
</tr>
<tr>
<td></td>
<td>[15]</td>
<td>98%</td>
</tr>
<tr>
<td>XM2VTS [23]</td>
<td>[2]</td>
<td>99.2%</td>
</tr>
<tr>
<td></td>
<td>[10]</td>
<td>98.7%</td>
</tr>
</tbody>
</table>

Out of them the method of edge detection proposed by Annapurani et al [1] and the method of force field transformations given by Bourouba et al [12] also gives good results when it comes to local and global feature extraction. According to the results of the papers analyzed, ear recognition was best performed by Artificial Neural Networks. The ear recognition by ANN approach is dependent upon the feature extraction method used. Also, a greater number of training images results in improved accuracy. Accuracy obtained with ear images with hair occlusion was generally less than ear images without hair occlusion. A summary of the accuracy obtained in various datasets is given in Table 1.

3. Proposed Ear Recognition System

The objective of this work is to create a pipeline from ear detection to ear recognition. The proposed system consists of three main stages - ear detection, ear feature extraction and ear recognition. It is shown in Fig. 4. Each stage will be explored in the following sub sections.

The modules of the proposed ear recognition system are shown in Fig. 5. All input images must be of the same dimensions so that it is possible to work with the dataset. In the preprocessing stage, the images are resized to a common set of dimensions. Then, the images are converted to its grayscale form and its binary form. After preprocessing, ear detection is carried out. In this, a program is run to find if an ear is detected in an image or no. If an ear is detected, then a rectangle is drawn around the ear where it is detected. From the detected ear, features are extracted using the new proposed algorithm. Once the features are extracted, they are stored as
templates for the next stage of training. Convolution neural network architecture is built for the training phase. All the stored templates are supplied to the neural network where the network learns on its own by adjusting weights. The network is trained repeatedly while changing parameters to get the highest accuracy and the lowest loss possible. Once it is trained, the network is stored separately as a trained convolution neural network. For the stage of recognition, test images are supplied to the trained convolution neural network and it identifies the person on the basis of the ear.

3.1 Preprocessing
Preprocessing of images is one of the most important steps when images are to be used for machine learning. In the preprocessing phase, many preprocessing tasks can be performed on the images such as resize, grayscale, edge detection, binary image, blur and denoise. Resizing images to specific dimensions is important to have a uniform size of images across the dataset. For the purpose of ear detection, the images of the dataset were converted to grayscale form. Conversion to grayscale is important because color information usually does not help with identifying features and images. Conversion to grayscale also reduces the processing time and thus makes tasks such as edge detection and feature extraction faster. Fig. 6(a) shows the grayscale version of an original image. For the purpose of feature extraction, the images of the dataset were converted to grayscale form and then converted to binary image format. Binary image format helps to segment the image into regions, out of which some regions are of interest and some are not. Fig. 6(b) shows the binary version of an original image.
3.2 Ear Detection

Ear detection is carried out by using Haar cascade files in OpenCV. An image with detected ear is shown in Fig. 7. Several iterations were done to detect ears in the dataset. The ear was detected in 699 out of 700 images.

![Image of ear detection](image)

Fig.6. Preprocessing (a) grayscale, (b) Binary format

![Image of detected ear](image)

Fig.7. Image with detected ear

3.3 Ear Feature Extraction

After detecting the ear, the image with the
Let this line be L1. Another perpendicular line L2 is drawn at the midpoint of line L1 and this intersection point lies on the tragus of the ear.

For all the images in the AMI dataset this algorithm is applied and the resulting rectangular extracted region is saved. The extracted region is further passed in convolution neural network architecture for recognition process.

3.3.1 New Proposed Ear Feature Extraction Algorithm

rectangle outside the detected ear is supplied to a MATLAB code for feature extraction. Several points inside ear could be used as a unique feature but we have selected tragus and antitragus out of them. Tragus and antitragus appear much clearly as a part of the ear and could be used as a feature for extraction. The proposed algorithm with a slight variation can be used for both tragus and antitragus extraction. Many feature extraction algorithm are already available such as Principal Component Analysis, contouring, active contouring etc. Snake - Active contouring is another algorithm that could be used to extract feature. This algorithm was proposed by Chan-Vese [8]. The algorithm takes the original image and binary mask image as its arguments. This algorithm creates a segmentation of the image. It segments it in a foreground and a background image. Then the algorithm traverses on the boundary of the objects as if a snake is traversing the boundary. This creates boundary of all the objects in the image, segmented in background as well as in foreground. This algorithm does the segmentation in an iterative manner and by default it does 100 iterations. The proposed ear feature extraction algorithm is based on the concept of creating a boundary of the ear surface. After
specifying the boundary of the ear we locate the tragus and antitragus using intersection of two longest perpendicular lines. A grayscale version of the ear image is generated with a certain threshold. Using this grayscale image a binary version of the image is generated. All the executions are done over the binary image and the result is displayed on the original image. Using the function bwboundaries(), a boundary around the ear is generated. This function generates the boundaries around the blobs found inside the ear as well. After creating a boundary of the ear we iterate for all the pair of points on the boundary and check which is longest line corresponding to the pair of points.

**Pseudo Code: tragus and antitragus extraction**

1. Convert original image to grayscale form.
2. Convert to binary image and obtain boundary of the ear.
3. Find the distance using Distance = sort ((x (k) - x)^2 + (y (k)-y)^2)
   And hence find the longest line (L1) using the coordinates of the boundary obtained.
4. After obtaining the longest line, the midpoint (M1) of line L1 is calculated using Midpoint = mean(x (index1)-y (index2))/(x (index1)-x (index2))
5. A perpendicular is drawn to the line L1 at point M1 by calculating negative slope of line L1 using Slope = - x (index1) - x (index2))/ ((y (index1) - y (index2)).
6. A rectangle is drawn using the point obtained. This rectangle contains the tragus of the ear.
7. Another line (L2) is drawn which is the mirror of L1 and its axis is shifted. The midpoint (M2) of line L2 is calculated and a perpendicular is drawn.
8. Another rectangle is drawn with the point M2 obtained. This rectangle contains the antitragus of the ear.

The new proposed algorithm works on the technique of creating a boundary on the ear detected and generating a pair of longest perpendicular line in it. The intersection of both these lines will give the co-ordinates of the center of mass of tragus and antitragus respectively. The whole extraction process is done in MATLAB environment. The algorithm initially checks if the given image is present in the specified MATLAB directory. If it’s available then it reads it and stores it in a variable. This variable will store the image in the form of a matrix. Then the grayscale format of the image is generated and its grayscale threshold is also generated. Using this threshold a binary equivalent of the image is generated. In the next step, bwboundaries() function is used to generate a boundary of the ear image. This will create a boundary outside the ear image.

![Fig.8. Feature extraction](image)

(a) Extracted Antitragus, (b) Extracted Tragus
Now the next step is to find the longest line inside the generated boundary. A particular point \((x, y)\) is selected on the boundary. The rest of the points on the boundary are represented by \((x(k), y(k))\) where \(k\) varies. The following equation is used to calculate the distance between two points on the boundary of the ear:

\[
\text{Distances} = \text{sort} ( (x(k) - x)^2 + (y(k) - y)^2). \]

The pair of points which gives the maximum distance that is, the longest line, is stored. This pair of points is represented by \((x(\text{index1}), y(\text{index1}))\) and \((x(\text{index2}), y(\text{index2}))\). A line is drawn using these points. Let this line be \(L1\). The slope of the line is also calculated. Using the co-ordinates acquired the midpoint of the line is calculated from where the other longest line will pass. The slope of the perpendicular line is calculated using previously generated slope. Let the perpendicular line be \(L2\). The following equation is used to calculate the slope of the perpendicular line.

\[
\text{Slope} = (x(\text{index1}) - x(\text{index2}))/ (y(\text{index1}) - y(\text{index2})). \]

These two lines when drawn on the ear image forms an intersection point which gives us the position of the tragus. Using this intersection point a rectangular region of suitable area is generated which contains the tragus. It is shown in Fig. 8(a). All these steps are repeated for each and every ear image in the database. The rectangular regions generated are stored as templates which will be used as inputs in the convolution neural network. For the extraction of antitragus another line \(L3\) is generated which is the mirror image of line \(L2\) and its axis is shifted by 150 units. The mid-point of \(L2\) is calculated and a perpendicular is drawn. This intersection point gives co-ordinates of the antitragus. It is shown in Fig. 8(b). Again a rectangle is drawn using this co-ordinate and is stored as a template.

### 3.4 Stored Templates

After the feature extraction code is executed, it gives output of the extracted tragus and antitragus within the rectangulated region. The code is executed on the dataset and all the extracted features are saved. Fig. 9(a) shows an extracted template of tragus and Fig. 9(b) shows an extracted template of antitragus.

### 3.5 Ear Recognition

We have used a convolution neural network for ear recognition. We have executed various architectures with different parameters such as epochs and batch size for training.

#### 3.4.1 Convolution Neural Network Architecture

The architecture used for ear recognition is shown in Fig 10. Four two-dimensional convolution layers are used. Three of these layers are followed by max pooling layers. The last two-dimensional is followed by a global max pooling layer. After that, a dense layer is added, followed by softmax function. Many variations were tested for the architecture. Four convolution layers gave the best results. The number of filters and kernel size were also varied in each convolution layer. The first convolution layer gives an output of 16 filters and has a kernel size of 3 x 3. The second convolution layer gives an output of 32 filters and has a kernel size of 2 x 2. The third convolution layer gives an output of 64 filters and has a kernel size of 2 x 2. The fourth convolution layer gives an output of 128 filters and has a kernel size of 2 x 2. Padding is added across all convolution layers and Rectified Linear Unit is used as the activation function in all layers. Dropout and flatten layers were also experimented with but they did not improve performance of the system. Therefore, these layers were removed.

### 5. Results

For ear detection, Haar cascades were used. Using Haar cascades, ear was detected in 699 images out of 700 images. Using the new proposed feature extraction, features were extracted in 100% of the dataset for the features, tragus and antitragus. For the purpose of recognition, a convolution neural network was trained and tested separately for tragus and antitragus. Training
images consist of 700 pictures, validation images consist of 465 pictures and test images consist of 499 pictures. The maximum accuracy reached for ear recognition using tragus is 94.18%. The maximum accuracy reached for ear recognition using antitragus is 73.56%.

It is found out that the tragus is a better feature than the antitragus since it has a more distinct and clear boundary. However, to make the system robust, we use both the features. In case the tragus is occluded completely, the antitragus can be used for recognition. For recognition using tragus, the accuracy vs. epoch graph is given in Fig. 11(a). For recognition using antitragus, the accuracy vs. epoch graph is given in Fig. 12(a).

6. Conclusion

In this work, a new algorithm for feature extraction for an ear recognition system is proposed. The recognition phase is dependent on the feature extraction phase. Therefore, to make the recognition rate higher, we need to select appropriate features and a good feature extraction algorithm. The outer ear has a number of features but we have selected two unique features, the tragus and antitragus. These features were selected as they can be clearly identified and have a clear boundary. They are also the features which are least occluded by jewellery, caps, hair and spectacles. For the extraction of these features, a new algorithm is proposed. This algorithm is based on finding a boundary of the ear and finding the
longest line which can be made using the points that lie on the boundary. For the ear recognition phase, a convolution neural network is used as a classifier. The recognition is performed separately for tragus and antitragus. The future possibility of this work is that it can be used for more images. Only one dataset is used in this work. We can test the system for images with more occlusions. Working with occluded images will give an idea of how the situation will be in real time application of the system. The system could also be analyzed for images of the ear taken from different angles, from different distances and in different settings. All of the pictures in our current dataset are taken in an indoor setting with the same lighting and from approximately the same distance. Since machine learning is used, a greater number of images could result in more accuracy. The more number of images that are used for training a convolution neural network, the better is the training and learning of the parameters. For the recognition phase, the future scope is to experiment with more variations of the convolution neural network architecture. In today’s time, there is an increased need for efficient and reliable security systems which can monitor and authenticate people who attempt to access a secure location. A future task is to extend the recognition system to be used in surveillance and security systems.
7. References


[20] ‘Mathematical Analysis of Images (AMI) Ear Database’ ctim.ulpgc.es/research_works/ami_ear_database/


[22] ‘IJIT Delhi Ear Database’ www4.comp.polyu.edu.hk/~csajaykr/ITD/Database_Ear

[23] ‘XM2VTS Database’ http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/
