Human Identification based on Ear Recognition

S. Gangaram¹, and S. Viriri¹,²

1. INTRODUCTION

Due to the fact that conventional means of human recognition, such as passwords, ID cards, etc. can be stolen, faked, or forgotten, there has been much research in the field of using biometric characteristics for this purpose. Biometric characteristics are universal, unique, permanent, and measurable. Many methods for identifying humans have been researched, including those based on face, iris, fingerprint and gait recognition, however, these are considered to be slightly invasive because they require the co-operation of the individual. Ear recognition is non-invasive, and is not affected by factors such as mood or health, unlike facial recognition. The appearance of the ear is also unaffected by age, making it suitable for long-term identification.

Therefore, this paper presents an overview of the state-of-the-art of various feature extraction techniques and classification methods for improved ear recognition.

2. LITERATURE REVIEW

The French criminologist Alphonse Bertillon was the first to become aware of the potential use for human identification through ears, more than a century ago [3]. Much research has been done to show that the appearance of the outer ear is unique and relatively unaffected by age. While it has not been proven that each person’s ear is unique, studies shown in [10] gave supporting evidence.

Iannarelli inspected more than 10 000 ears and discovered that they are all distinguishable, and created a method where only 12 measurements are used to differentiate individuals. In 1995, Carreira-Perpiñán investigated using artificial neural networks with linear nodes for feature extraction.

In a study by D. J. Hurley et al. [8], a force field feature extraction approach based on simulated potential energy fields was used. The force field transformation sees the image as a collection of mutually attracting particles that act as the sources of a Gaussian force field. This approach was performed on 63 individuals and the results showed that it was an improvement on PCA-based methods; however, it is only applicable when there is a small background in the ear image.

In 2005, M. Choras presented a study on ear biometrics using geometric feature extraction [6]. This involved image normalization, edge detection, calculation of the centroid, coordinate normalization and two steps of geometrical feature extraction. The geometric features extracted used the points of intersection between circles of different sizes with the calculated centroid as their centre and the contours extracted from the ear image. However, this method was carried out on extremely high quality images with ideal conditions. Because the environment was ideal for recognition, the results were error-free. We cannot expect the same results in the case where the images are of lower quality or there are changes in illumination.

B. Moreno et al. [7] carried out two experiments with neural network classifiers. For the first experiment, features were extracted by performing edge detection and getting seven known feature points of the outer ear to form the feature vector. For the second experiment, the intersection points between h horizontal cuts, v vertical cuts and 2(h+v) diagonal cuts over an h x v size image forms a morphology vector. The first method resulted in an extremely low 43% recognition rate, while the second approach resulted in an 83% recognition rate. This indicates that ear morphology is much more effective than the feature point approach, although ear morphology is still not ideal as there is much room for improvement.
Prakash and Gupta recently described a new approach on ear recognition using edges [5]. They used skin segmentation and classified the edges into 2 groups: convex and concave. Thereafter, the edges in the skin region are broken up into edge segments, which form an edge connectivity graph. The convex hull of all edges is computed from this connectivity graph. The enclosed region is the ear region. This study used full profile images and a 96.63% detection rate was attained.

H. Alastair et al. [4] proposed the ray transform approach, which detects the ear in different positions and ignores straight edges in the image (which are caused by glasses or hair). This method uses a light ray analogy to scan the image for cylindrical and curved structures, such as the outer helix. The simulated ray is reflected in bright tube-shaped regions, highlighting these regions in the transformed image. Since glasses have straight edges, they are not highlighted by the ray transform. This method had a 98.4% recognition rate.

3. METHODOLOGY
The method uses a training dataset and testing dataset of ear images. The basic procedure is outlined in Figure 1 below:

3.1 Preprocessing
Preprocessing involves converting the image to grey scale, performing histogram equalization, and Gaussian filtering. Preprocessing is essential in order to remove noise and smooth the image.

3.1.1 Converting to grey scale
Converting to grey scale involves mapping colour RGB triplets to a single value representing the grey scale intensity. Each color pixel is described by a triple (R, G, B) of intensities for red, green, and blue. A weighted average of these values is calculated as the grey scale intensity as follows:

\[ I = 0.21R + 0.72G + 0.07B \]

A weighted average is used for human perception, because humans are more sensitive to green than other colours, therefore green carries the largest weight.

3.1.2 Histogram Equalization
Histogram equalization involves transforming the histogram of the image in order to increase the contrast. The intensities will be distributed evenly on the histogram. Suppose the intensity values range from 0 to \( L-1 \), then let \( p_n \) denote the normalized histogram of the grey scale image as follows:

\[ p_n = \frac{\text{no. of pixels with intensity } n}{\text{total no. of pixels}} \quad n = 0, 1, ..., L - 1 \]

\( p_n \) creates a mapping function for the pixel intensities, \( T \), as follows:

\[ T(K) = (L - 1) \left( \sum_{n=0}^{K} p_n \right) \quad K = 0, 1, ..., L - 1 \]

The intensity value \( K \) will be mapped onto \( T(K) \) in the output image.

3.1.3 Gaussian filtering
Gaussian filtering is used to reduce noise in the images by decreasing the intensity variation among pixels. It uses a 2D sliding window matrix (kernel) that sequentially traverses the image and uses the pixel values contained in the sliding window to replace the centre pixel value. The equation of the Gaussian function is:

\[ G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \]

Gaussian filtering uses this 2-D distribution as a point-spread function, and this is achieved by convolution. This function is used to set values of the sliding window and performs the sum of products.

3.2 Region of interest detection
Region of interest detection involves identifying the boundary of the ear in the image and extracting it. To implement this, I used a Haar Feature-based Cascade Classifier.

3.2.1 Haar Feature-based Cascade Classifiers
Haar Feature-based Cascade Classifiers are trained with a few hundred samples of an object, in this case, ear images, called positive images, and arbitrary images of the same size, called negative images. Afterwards, the classifier is applied to an image and outputs a 1 or 0, depending on whether the image contains the object of interest (in this case, the ear). This approach is not 100% effective, and in many cases, the original image has to be used.
3.3 Feature Extraction

Feature extraction deals with isolating distinct features of the ear in the image. I implemented PCA, LBP and spatial histograms to accomplish this.

3.3.1 Principal Component Analysis (PCA)

PCA is a method of dimensionality reduction, used to reduce the number of features to only those with a large variation between them. Firstly, each pixel in an image is taken row by row and converted to a row vector of the intensity values. All the row vectors of the training set (or testing set) are combined to form a matrix. Thereafter, the covariance matrix is calculated as follows:

$$
cov(x_i, x_j) = E[(x_i - \mu_i)(x_j - \mu_j)]$$

for $i, j = 1, 2, 3… n$, where $E$ is the mathematical expectation.

Principal components analysis (PCA) is an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. This is done on the covariance matrix ($C$) by satisfying the relation $Ce_i = \lambda_e e_i$ where $e$ and $\lambda$ are the corresponding eigenvectors and eigenvalues respectively. Thereafter a matrix $A$ is constructed from the eigenvectors sorted by decreasing eigenvalues, and each training image is projected onto the PCA subspace. A test image will be projected onto the PCA subspace, and then compared with the training images to classify it.

3.3.2 Local Binary Patterns (LBP)

LBP does not look at the image as a whole, but instead isolates local features of an object. Each pixel is compared with only its neighbourhood. A pixel is taken as the center and used as a threshold value. If the intensity of the neighbour pixel is greater than or equal the centre pixel, then denote it with 1. Otherwise denote it with a 0. The result is a binary number for each pixel, eg.11001111, called an LBP code.

![Example calculation of LBP](image)

Local Binary Patterns reveal the texture of an image, and the features extracted have a low dimensionality.

3.3.3 Spatial Histograms

Spatial histograms are used to preserve local information in an image. Instead of computing one histogram for the whole image, the image is divided into $n \times n$ smaller regions, and a histogram is calculated for that each region separately. Thereafter, all the regional histograms are concatenated to form the spatial histogram. The region size used was $8 \times 8$.

![Image divided into regions to compute histograms](image)

3.4 Identification

Identification is the final act of classifying an ear image as belonging to a certain individual. It involves using the set of features that were extracted and comparing them to the database to determine which image matches the closest to it. In order to achieve this, I implemented the SVM classifier, and the K-nearest neighbour algorithm.

3.4.1 Support Vector Machine (SVM)

SVM is a supervised machine learning classification technique. It is based on the concept of decision planes that define decision boundaries. A decision plane is one that separates a set of objects belonging to different classes. SVM determines the optimal separating hyperplane in order to categorize data.

![Optimal hyperplane for classifying data](image)

The example above depicts binary classification. However, ear recognition is a multi-class classification problem, which the algorithm takes care of by repeatedly performing binary classification. The SVM algorithm learns from the training set of images, determines the optimal hyperplanes of each separation, and uses this to classify a given test image.

3.4.2 K-Nearest Neighbours

K-nearest neighbours is a classification algorithm that uses a measure of similarity to classify objects. A distance is calculated between a test sample and all the training samples to determine which $k$ training samples have the smallest distance from the test sample. The test sample will be placed into the class of the majority of its nearest neighbours. If $k = 1$, the test sample will simply be assigned to the class of its nearest neighbour. The distance formula used is the Euclidean distance:
\[ d = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \]

where \(x_i\) are the components of the training sample vector and \(y_i\) are the components of the test sample vector, and \(n\) is the number of components in each vector.

4. RESULTS AND DISCUSSION

The experiments were tested using the testing set of 125 ear images. The dataset of ear images used was the IIT Delhi Ear Database version 1.0. This database contains ear images of 125 different individuals, acquired during October 2006 – June 2007 using a simple imaging setup. Each individual has 3 – 6 ear images, 1 of which was used in the testing set, and the rest used in the training set. All individuals are in the age group 14 – 58 years. The images use an integer to identify each individual, ranging from 1 to 125. The resolution of these images is 272 x 204 pixels and all these images are available in jpeg format.

The images were preprocessed, which includes conversion to grey scale, histogram equalization and Gaussian filtering. Thereafter the region of interest was detected and feature extraction was performed. Each of the 3 feature extraction methods were tested, along with both the classification methods that were implemented. The results are summarized in Table 1 below, displaying the success rates of each combination.

![Figure 7: (a) Image 004_01.jpg from dataset. (b) Image 004_02.jpg from dataset.](image)

The most notable feature extraction method was spatial histograms, which resulted in the highest success rate because it uses local descriptors as opposed to the entire image. However, these results depend on the dataset used, as factors such as illumination variation in the images and training set size could drastically affect the outcome. Presumably, when using a different dataset, the preprocessing carried out might need to be altered, as well as the feature extraction techniques used.

The features that are extracted is what affects the success rate of the system most, therefore, in future work, other feature extraction methods should be investigated. Methods such as Scale Invariant Feature Transform (SIFT), Histogram of Oriented Gradients (HOG) and Speeded Up Robust Features (SURF) should be applied to the problem of ear recognition to determine if the success rates can be improved. In addition, other classification methods should be applied, eg. Random/Decision trees, Multi-Layer Perceptron and k-Means Clustering.

5. CONCLUSION AND FUTURE WORK

This paper investigated the effectiveness of using the ear as a biometric trait to identify humans, and compared the accuracy of using different methods for feature extraction and classification. The most notable feature extraction method was spatial histograms, which yielded the highest success rate because it uses local descriptors as opposed to the entire image. However, these results depend on the dataset used, as factors such as illumination variation in the images and training set size could drastically affect the outcome. Presumably, when using a different dataset, the preprocessing carried out might need to be altered, as well as the feature extraction techniques used.

The success rates of using the PCA feature extraction technique are significantly lower than the other two. This is because PCA determines components with maximum variance over all the classes. These may not necessarily be useful for classifying the images. Therefore, in images with a lot of illumination variation, the wrong components were identified and used to define the PCA subspace. Another factor is the small number of images for each individual in the training set, which causes the covariance estimates for the subspace to be slightly incorrect. In order to achieve a 90% success rate using PCA, at least 8 images per person are needed, however, there are only 3-5 images in this case.

The LBP technique yielded success rates of 69.6% and 71.2%, which are acceptable. Because LBP does not see an image as a high-dimensional vector but only considers local features, factors such as illumination variation and small training set size have very little effect. However, what does affect LBP are factors like rotation, scale and translation, which may have caused the average success rates.

The spatial histograms feature extraction method resulted in the highest success rates, 85.6% and 87.2%. Just like LBP, this method does not look at images as a high-dimensional vector but instead considers a group of regions or grids into which the image is divided. Therefore, it is not affected by the large variation in illumination or the small training set size, and very slightly affected by scale, rotation and translation. Therefore, spatial histograms are the most effective method of feature extraction.

Both the SVM and K-nearest neighbours classification algorithms resulted in very similar success rates, given the method of feature extraction.

6. ACKNOWLEDGMENTS

The people who put together the IIT Delhi Ear Database are hereby acknowledged and thanked for allowing the use of this database for research and investigation.

7. REFERENCES


