PalArch's Journal of Archaeology of Egypt / Egyptology

IMPROVING THE PERFORMANCE OF HUMAN EAR-BASED AUTOMATED BIOMETRIC SYSTEMS USING FEATURES FUSION AND COMPARING WITH CONVOLUTION NEURAL NETWORKS

Omid Sakaki Ghazvini

Department Of Computer Engineering, Faculty of Computer Engineering and Mechanics,

Science and Research Brach, Islamic Azad University, Tehran, Iran.

Email: omidsakaki@yahoo.com

Omid Sakaki Ghazvini. Improving the Performance of Human Ear-Based Automated Biometric Systems Using Features Fusion and Comparing with Convolution Neural Networks -- Palarch's Journal of Archaeology of Egypt/Egyptology 18(18), 1577-1586. ISSN 1567-214x

Keywords: Biometric System – Recognition and Identification - Features Fusion - Convolution Neural Network - Human Ear

ABSTRACT

In this study, the features fusion method was used to improve the performance of human ear-based automated biometric system recognizing and identifying individuals without the need for user's interaction. The image ray transform (IRT) algorithm was used in the recognition stage. Subsequently, binarised statistical image features (BSIF) and weber local descriptor (WLD) were extracted from the normalized outputs of the recognition stage. The canonical correlation analysis (CCA) method was then used to combine the two extracted feature vectors and create a distinct feature vector. Finally, the K-nearest neighbors' algorithm (*k-NN*) was used for decision making. The accuracy of designing and implementing this system was 99.82%. Then a convolution neural network (CNN) was designed and trained with binarised statistical image features (BSIF) and weber local descriptor (WLD). The simulation accuracy of CNN was 100%. According to the results of this study, convolution neural network based on BSIF and WLD features has superior performance over state-of-the-art methods.

INTRODUCTION

Today, all kinds of computer systems and software play a vital role in human life. Humans are always looking for speed, accuracy and transparency in doing individual and social work. This trend has led to the spread and popularity of technologies based on computer systems. The application of computer systems is very diverse and wide. Aside from the benefits of computer systems, there is potential room for aggressive action. Humans use a variety of authentication methods to protect their privacy. With the advancement of technology and the emergence of new technologies, many methods of classical authentication are obsolete and ineffective. Today, biometric authentication systems have been introduced by experts and researchers as a suitable alternative to the old methods. In recent years, biometrics has emerged as a major trend in security systems [9]. Getting lost, forgotten and stolen in these systems is meaningless. Biometric systems are basically defined based on the behavioral or physiological traits of individuals. These traits have unique, distinctive, resilient and measurable characteristics [12]. However, the security challenge in these systems is very significant and a lot of research has been done in this field. One way to resolve this challenge is to use multipurpose systems. That is, systems performing on the basis of two or more behavioral or physiological traits. But the process of sampling several behavioral and physiological traits is time consuming and costly. So we have to use a biometric attribute that reduces system risk and increases security and speed.

Another challenge is that sampling and collecting biometric traits through interaction with people makes them sensitive and fearful. Some people even consider this process as a violation of their sanctity. Therefore, we must use a method that automatically extracts and collects biometric information while maintaining the respect and privacy of individuals. One of the biometric traits that takes into account the security challenges, speed and accuracy of biometric systems, as well as maintaining the dignity and privacy of individuals, is biometric data extracted from the ear.

Some researchers have shown that an ear-based identification system is a viable alternative to other older biometric systems such as the face, fingerprint, and iris [1]. In general, we can classify ear-based biometric systems into two main categories: (a) systems that detect the ear area within an image in great detail; (b) systems based on the geometric properties of the texture and biometric features of the ear, such as feature extraction of local binary patterns method [11]. Second class systems are more accurate in the identification process.

But new researches have been done on the development of human ear-based biometric systems that perform the human ear recognition and automation processes automatically and without user interaction. One of these systems is the method proposed in [15]. In this study, a not normalized IIT database is used. The proposed method first detects the area around the ear using image ray transform (IRT) and preprocesses the image of the ear to separate it from the background. After normalization, it uses local tissue descriptors such as local binary pattern (LBP) [11], binarised statistical image features (BSIF) [5] and weber local descriptor (WLD) [8] extract local characteristics of the ear [7]. In [15] BSIF and WLD have obtained the most accuracy. In fact, the WLD descriptor, based on accurate edge extraction, has increased the recognition accuracy rate compared to other methods.

The organization of this research is as follows. In section (2), the concepts and simulations performed in this study as well as the implementation steps of the proposed method are introduced. In section (3) the database used is introduced and then the numerical results, analysis and comparison of the proposed method with another previous research are described. The conclusion of the paper is presented in section (4).

Proposed Method

In this section, we explain the implementation details of the system, as shown in Figure 2-1.

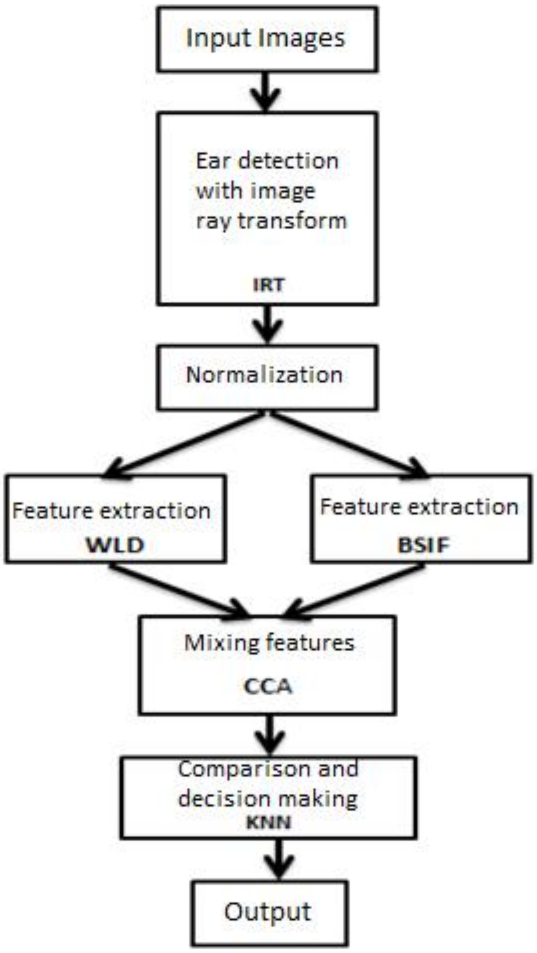

Figure 2-1: Implementation procedure steps

As shown in the study [15], the ear area is first determined by applying the image ray transform (IRT) algorithm to the input image. Image ray transform is a new way to enhance the structural features of images. In this method, optical laws are used to reveal the circular features in the image [2]. Then, the threshold limit method converts the original image to a binary image to determine its edges. Morphological operators then extract the image skeleton. To normalize the angle of the ear, we must first determine the angle of rotation. For this purpose, we determine the minimum and maximum points of the image and calculate the slope of the connecting line of these points (i.e. the angle of rotation). After this step, the normalized image is cropped as shown in Figure 2-2.

Figure 2-2: Ear recognition steps - a) Input image - b) Image ray transform - c) Binary image using the threshold limit method - d) Extraction of the image skeleton and removal of excess pixels - e) Calculating rotation angle - r) Output image

After the ear image recognition and normalization step, the BSIF and WLD features of the images are extracted, as suggested in [15]. The BSIF descriptor encodes each pixel of the biometric image based on the filtered response before the training. In this descriptor, two parameters, the filter size and the bit string length, are critical. The WLD descriptor consists of two components, the triggering differential and the orientation. The trigger differential is calculated for each pixel, which is equal to the brightness of the current pixel and the difference between the brightness of the current pixel and its neighbors. In addition, the direction of the current pixel gradient is also calculated [3]. The WLD descriptor extracts the edges of the image in the presence of extreme noise. The WLD descriptor has performed well in past research in the fields of tissue classification, facial recognition [3], and facial identification [6].

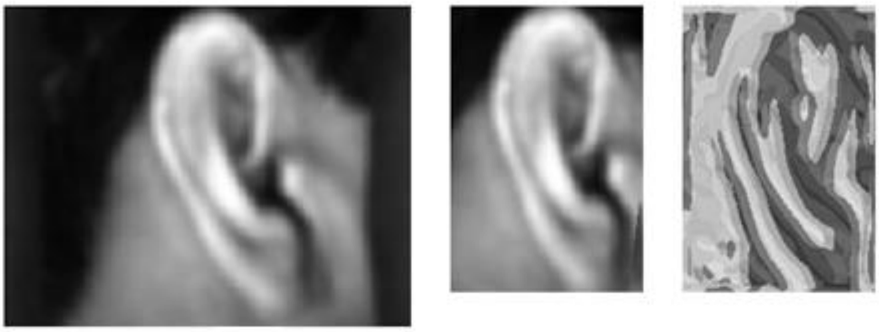

Figure 2-3: Feature extraction by BSIF method, a) original image, b) cropped and normalized image, c) performing BSIF on ear image

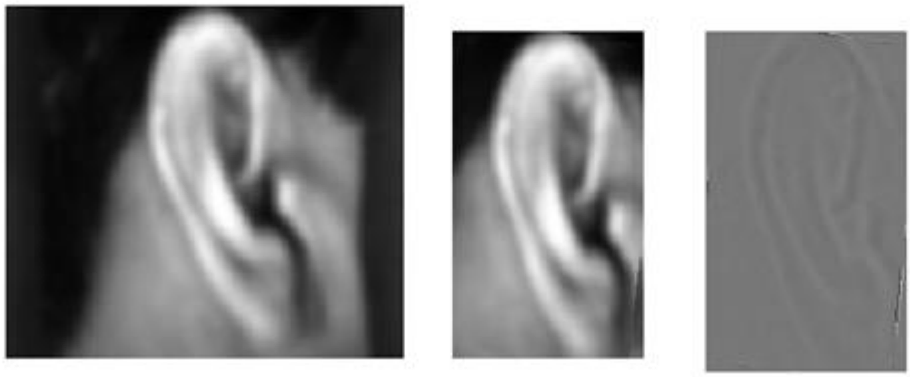

Figure 2-4: Feature extraction by WLD method, a) original image, b) cropped and normalized image, c) performing WLD on ear image

Canonical correlation analysis (CCA):

In this study, to improve the performance of the system presented in [15], the canonical correlation analysis method is used to fuse the useful features of WLD and BSIF. Using canonical correlation analysis (CCA), we combine two feature vectors to obtain a distinct feature vector. In multimodal recognition systems, CCA method is a powerful technique for statistical analysis [16]. The CCA technique is the most common form of general linear modeling, which examines the relationship between two sets of variables using multivariate statistical methods. In addition, the CCA technique evaluates the independent statistical relationships between two sets of variables by simultaneously analyzing sets and identifying elements strongly dependent on elements of another variable set.

Let X € Rp * n and Y € Rq * n represent two matrices, each with n features vectors of two different states. In other words, there are n instances. For each instance (p + q) features is extracted. Also, SXX € Rp * p and SYY € Rq * q represent the X and Y covariance matrices, and SXY € Rp * q represents their middle covariance matrix. The following covariance matrix contains information on the X and Y features:

$$s = \begin{pmatrix} \operatorname{cov}(x) & \operatorname{cov}(x, y) \\ \operatorname{cov}(y, x) & \operatorname{cov}(y) \end{pmatrix} = \begin{pmatrix} \boldsymbol{S}_{xx} & \boldsymbol{S}_{xy} \\ S_{yx} & \boldsymbol{S}_{yy} \end{pmatrix}$$

The main purpose is to find the linear combinations X * = WTx X and Y * = WTy Y that maximize the relationship between two standard variables:

$$\operatorname{corr}(X^*, Y^*) = \frac{\operatorname{cov}(X^*, Y^*)}{\operatorname{var}(X^*) \cdot \operatorname{var}(Y^*)}$$

Where cov (X *, Y *) = WTx Sxy WY, var (X *) = WTx Sxx Wx and var (Y *) = WTy Syy Wy.

Finally, the fusion at the feature level is achieved through the following formula:

$$Z = \begin{pmatrix} X^* \\ Y^* \end{pmatrix} = \begin{pmatrix} W_x^{\mathrm{T}} \boldsymbol{X} \\ W_y^{\mathrm{T}} \boldsymbol{Y} \end{pmatrix} = \begin{pmatrix} W_X & 0 \\ 0 & \boldsymbol{W}_y \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{Y} \end{pmatrix}$$

In the following subsections, the k-nearest neighbors (KNN) classifier is used for comparison and decision making. This algorithm uses distance or similarity criteria to classify a new sample, and based on this criterion identifies the k-nearest neighbors of the new sample and then uses the majority vote between this K neighbors to determine the class of the new sample.

Convolutional Neural Network (CNN):

Convolutional neural networks are one of the special architectures of artificial neural networks. CNN design is inspired by the function of the visual cortex of the brain (the part of the brain that processes visual information) [17]. Convolution neural network is a type of feedforward network that was first used for image recognition. Today, the convolution neural network is used for a variety of tasks such as image classification, motion recognition, voice recognition, and natural language processing. Convolutional networks are a type of feedforward neural networks with the following capabilities [18]:

Convolution layer: To reduce the size of the data and reduce the number of calculations.Sparse connection: To reduce the number of connections between neurons in the current layer with the next layer.

Parameter sharing: Using common weights for a kernel in each level and sharing these weights with other levels to extract features. Pooling: A policy to reduce the size of input data and learn useful features.

Convolution neural network has three layers which are as follows.

Convolutional Layer

This layer performs the task of extracting features from the input data (image, text and audio). This operation is performed by a set of kernels / filters. In fact, each kernel is applied to the input according to the type of task and extracts certain features [19].

Pooling Layer:

The pooling layer is embedded in the architecture after a convolution layer and is used to reduce the size of the feature map and network parameters. The use of max pooling and average pooling layers is very common in convolutional networks.

Fully Connection Layer

After the last pooling layer, fully connection layers are embedded that convert 2D feature maps to one-dimensional feature vectors. This layer also performs the task of calculating the scores of classes of objects. Fully connection layers act like their counterparts in traditional neural networks and including approximately 90% of the parameters of a CNN network. The fully connection layer shows the network output in the form of a vector of a specified size. This vector can be used to classify objects or to enhance the processing process.

Convolution neural networks (CNNs) operate by collecting features at each layer. They start their operation in an image by finding the edges, shapes and then the real objects. Deeper layers (close to the input data) learn simpler features such as edges and color changes. Higher layers, on the other hand, combine simple features with complex ones. Finally, the last layer combines different features to produce a classification result. In this study, we used local texture features, such as BSIF and WLD descriptors in convolution layers based on the settings in the table below.

Table 2-1: Values set for general simulation parameters

Value	Parameter	
80 percent	Training data ratio	Model
20 percent	Test data ratio	
Sigmoid	Fully connected layer activation function	
0.4	Error control rate	
0.001	Learning rate	

RESULTS

Database

The IIT Delhi Database is a standard database of ear images of 221 people aged 14 to 28 years. At least 3 ear images were recorded for each person. The total number of images in this database is 471. This database has two subsets. The first subset includes images of the ear with a resolution of 180 x 50 pixels, and the second subset includes images of the ear with a resolution of 204 x 272 pixels. Images with a resolution of 180 x 5 are cropped and normalized images of original images.

Numerical Results

In the reference research [15], the main purpose was to design a biometric system based on the human ear to perform automatically diagnostics and identification operations without the need to interact with users. For this reason, the authors used human-normalized images in the IIT database, contrary to research [7]. The authors in [15] used IIT database images and proved that the use of WLD descriptors increases the recognition accuracy rate. It has also been shown in [7] [15] that the BSIF descriptor performs well and provides better results. Therefore, in this research, we will use two descriptors WLD and BSIF. After the recognition and normalization of the ear image, we extract the WLD and BSIF features, then combine these features using the canonical correlation analysis (CCA) method and with the KNN algorithm based on the settings in [7] And [15] evaluate the accuracy of system identification.

Finally, it was found that the best case is to combine the BSIF features with the 3 * 3 filter window size and the 6-bit string length with the WLD features. The results are shown in Table 3-1.

Table 3-1: Comparison of the results of the proposed method with various researches in the field of human ear-based biometric systems

References	Feature Extraction	Classifier	Recognition Rate
Kumar [13]	Shape Feature	KNN	30/62
Hurley [14]	Force Field	KNN	66/67
Kumar [13]	Gabor Phase	KNN	84/46
Sakaki [15]	IRT+BSIF	KNN	93/88
Hezil [7]	BSIF	KNN	98/90
Our Approach	IRT+ CCA (WLD+BSIF)	KNN	99/82

Next, the convolution neural network is used, which uses BSIF and WLD features in its inner layers (according to the settings presented in Table 2-1).

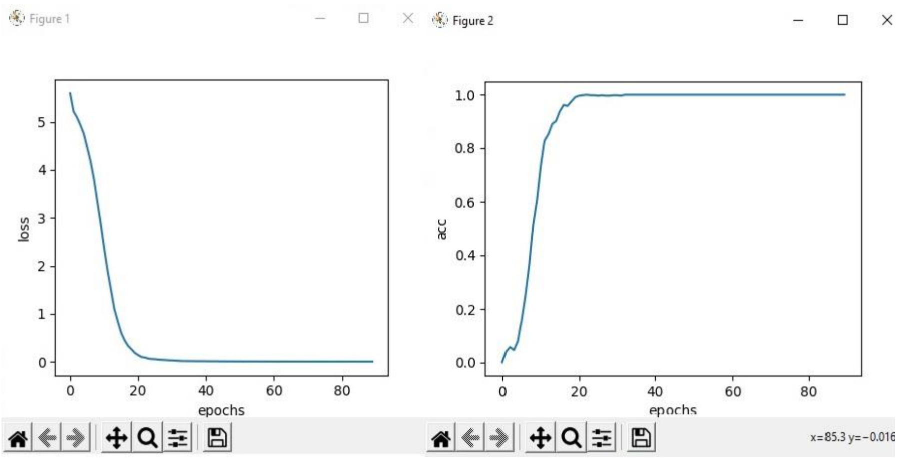

Figure 3-1: Plot of accuracy and error obtained in simulation

As shown in Figure 3-1, the convolution neural network simulation with the convolution layers BSIF and WLD has achieved 100% accuracy.

Table 3-2: Comparison of the simulation results of this study

Model	Accuracy
Fusion of features with CAA and decision making with KNN	99.82
Convolution neural network with feature layers BSIF and WLD	100

ANALYSES AND CONCLUSION

As suggested in the study [15], ear recognition operation and ear characteristics-based recognition operation require each other. Hezil and Boukrouche [7] used human-normalized images and then obtained the highest accuracy rate by the BSIF feature extraction. In [15] an automated system has been developed that detects and normalizes images and people by machine. In this paper, the BSIF feature extraction method has the highest accuracy rate. However, the WLD feature extraction method has gained better accuracy than [7] due to the preservation of edges in machine recognition and normalization operations.

In the system proposed in this research, the recognition block is the same as in the research [15]. But in the identification block, first BSIF and WLD features were extracted and then the features were combined using canonical correlation analysis (CCA) method. The identification operation was then performed using a feature vector (consisting of two feature vectors BSIF and WLD) and achieved an accuracy of 99.82. The accuracy of the proposed system is higher than the systems presented in [7] and [15]. Next, we trained a convolution neural network that used the BSIF and WLD features in its inner layers and eventually achieved 100% accuracy. According to the results, convolution neural network based on BSIF and WLD features provides more accuracy than other researches in the field of human ear-based biometric systems.

REFERENCES

Chang, K, et al., et al. *Comparison and combination of ear and face image in appearance-based biometrics.* 9, 2003, IEEE transactions on pattern Analysis and Machine Intelligence, Vol. 25, pp. 1160-1165.

Cummings, A, Nixon, M and Carter, J. *A novel ray analogy for enrolment of ear biometrics.* Washington, DC, USA : Biometrics: Theory Applications and Systems (BTAS), 2010.

Chen, J, et al., et al. *WLD: a robust local image descriptor.* 9, s.l. : IEEE Transaction on pattern Analysis and Machine Intelligence, 2010, Vol. 32, pp. 1705-1720.

Emersic, Z, Struc, V and Peer, P. *ear recognition: more than a survey.* 13, s.l. : Neurocomputing, 2017, Vol. 255, pp. 26-39.

Ghiani, L, et al., et al. *Fingerprint liveness detection using binarized statistical image features.* Arlington, VA, USA : Biometrics: Theory, Applications and System (BTAS), 2013.

Li, S, Gong, D and Yuan, Y. *Face recognition using the weber local descriptor.* 25, s.l. : Neurocomputing, 2013, Vol. 122, pp. 272-283.

Hezil, N and Boukrouche, A. *Multimodal biometric recognition using human ear and palmprint.* 5, s.l. : IET Biometrics, 2017, Vol. 6, pp. 351-359.

Vegte, J V. *Fundamentals of Digital Signal Processing.*

Mamta and Hanmandlu, M. *Multimodal biometric system built on the new entropy function for feature extraction and the refined scores as a classifier.* 7, s.l. : Expert Systems with Applications, 2015, Vol. 42, pp. 3702-3723.

Burge, M and Burger, W. *ear biometrics in computer vision.* Barcelona, Spain, Spain : Pattern Recognition, 2000.

Ojala, T, Pietikainen, M and Maenpaa, T. *Multiresolution gray-scale and rotation invariant texture classification with local binary patterns.* 7, s.l. : IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, Vol. 24, pp. 971-987.

Galdamez, P L, Raveane, W and Arrieta, A G. *A brief review of the ear recognition process using deep neural networks.* s.l. : Journal of Applied Logic, 2017, Vol. 24, pp. 62-70.

Kumar, N and Wu, C. *Automated human identification using ear imaging.* 2011.

Hurley, D j, Nixon, Carter. *Force field feature extraction for ear biometrics.* 2005.

Omid Sakaki ghazvini, and A.Koochari. 2021. "*IMPLEMENTING AND DESIGNING BIOMETRIC SYSTEM FOR DETECTION AND IDENTIFICATION WITHOUT LIMITATION OF HUMAN EAR*". *PalArch's Journal of Archaeology of Egypt / Egyptology* 18 (10):2076-84

Sun, Zen, Liu, Heng, Xia. 2005. "*a new method of feature fusion and its application in image recognition.".* pattern recognition.

Wani, M.A., et al. *"User emotion analysis in conflicting versus non-conflicting regions using online social networks".* Telematics and Informatics, 2018.

Goodfellow, I., Y. Bengio, and A. Courville. *"Deep learning".* 2016. Book in preparation for MIT Press.

Liu, W., et. *"A survey of deep neural network architectures and their applications".* Neurocomputing 2017.