AN EMPIRICAL STUDY AND EVALUATION ON AUTOMATIC EAR DETECTION

Resmi K. R. 1), Raju G. 2)

1) School of Computer Science, Mahatma Gandhi University, Kerala, India, resmykr@gmail.com
2) Department of Computer Science and Engineering, Christ, Bengaluru, India, kurupgraju@gmail.com

Abstract: Biometric is one of the growing fields used in security, forensic and surveillance applications. Various types of physiological and behavioral biometrics are available today. Human ear is a passive physiological biometric. Ear is an important biometric trait due to many advantages over other biometric modalities. Because of its complex structure, face image detection is very challenging. Detection deals with finding or localizing the position of ear in the given profile face image. Various methods like manual, semiautomatic and automatic techniques are used for ear detection. Automatic ear localization is a complex process compared to manual ear cropping. This paper presents an empirical study and evaluation of four different existing ear detection techniques with our proposed method based on banana wavelets and circular Hough transform. A comparative analysis of the five algorithms in terms of detection accuracy is presented. The detection accuracy was calculated by means of manual as well as automatic verification.

Copyright © Research Institute for Intelligent Computer Systems, 2020. All rights reserved.

1. INTRODUCTION

Authentication of an individual is important in various security applications. Traditional security methods are based on tokens (like ID card) and knowledge (Passwords, PIN etc.). These methods have the drawback since sometimes these tokens can be stolen or duplicated by someone else or they can be forgotten. The difficulties of traditional methods are removed by using biometric traits which use universal, permanent and unique characteristics of a person. Biometric authentication is mainly based on face, fingerprint, iris, retina, ear, voice and speech. Compared to other physiological biometrics, ear biometric has gained attention due to the following advantages.

According to medical studies [1] the structure of ear is constant in the age from 8 to 70 years. During the period of 4 months to 8 years it stretches its shape and it further changes its shape when a person reaches the age of 70 years and over.

Ear recognition is not affected by changes in pose and facial expressions [2]. The recognition performance of biometrics like face changes with pose and facial expressions.

Ear is a passive biometric which means it is contactless. Capturing biometric data does not require user cooperation. Biometrics like iris, retina, etc. needs user cooperation for data capture and it is affected with problems like anxiety, hygiene and privacy. Ear is relatively immune to these problems [2].

The color distribution of ear is almost uniform [3]. This helps to preserve almost all information.

Due to smaller structure of ear, it is possible to work faster [4].

Biometrics like finger/palm print suffers from such problems as wear and tear. Ear is relatively free from these problems.

Multimodal biometrics using ear enhances the results of other popular biometrics [4].

Ear can be used in applications like security, access control, surveillance, etc.

The major problems in case of ear detection are illumination and occlusions like hair, spex, ear ring, etc. The important requirement for personal identification from ear images is the detection and localization of ear from the given image. Various methods are used for ear detection. The simplest
technique used is manual ear cropping. Several semi-automatic and automatic detection methods are also in practice. Automatic ear localization or detection is still a difficult task. Recent studies in automatic ear detection and recognition using CNN have shown very good results with the use of lots of training data and high GPU power.

In this paper four selected ear detection algorithms are implemented. One among them [5] is modified. The ear detection accuracy of the five algorithms is analyzed with standard dataset. Further a method is suggested and experimented for automatic evaluation of ear detection methods.

2. EAR DETECTION

The different stages [6] in an ear recognition system are shown in Fig. 1.

![Figure 1 – Stages in an ear recognition system](Image 57x468 to 292x552)

The biometric sample obtained from image acquisition module undergoes preprocessing. Preprocessing includes enhancement of the original image by noise removal or histogram equalization. Detecting an ear involves finding/localizing the ear position in an image. Various features obtained from detected ear are used for building a verification/identification model using one or more classification techniques.

The quality of the detected ear depends on the ear recognition performance. Detection can be performed using manual, semi-automatic or automatic methods. In the rest of the section, an overview of prominent ear detection algorithms are presented. Burge and Burger [7] developed one of the most important ear detection technique based on deformable contours. The main drawback of their method is that contour initialization requires user interaction. Their localization method is not automatic.

Template based approach for ear detection was proposed by Prakash and Gupta [8]. They first perform skin segmentation using color based method. A standard ear template was created by taking selected images from the database. Ear was localized by moving the template over the image and finding the distance transform and cross correlation value at every pixels. Detection accuracy of 95.2% was reported on IITK database with 150 images.

Wavelets based on shapes called banana wavelets were proposed in [5]. Banana wavelets gave responds to curved regions in an image. Since face contains many curved regions, the wavelets find multiple regions in face image and a circular Hough transform addition to this improves the detection accuracy. With 300 images their method reports a detection accuracy of 83%.

In [9] they proposed another technique based on wavelets and log gabor filters. They reported a detection rate of 88.4% on XM2VTS database. An affine invariant detection of ear based on fast line based housdorff distance was proposed in [10]. The detection was done in images with changes in pose, lightning and complex background. An ear detection approach based on Modified CAMSHIFT was proposed in [11]. Detection was done in two stages. First profile face is tracked using CAMSHIFT from video. Secondly ear was detected using contour fitting method.

Recent detection works with the use of deep learning give superior detection rate compared to traditional methods. Emsric et al. [12] used convolutional encoder decoder for segmenting ear based on ear and non-ear pixels. SegNet [13] was used for segmentation. Pixel wise details are used for identifying ear location. Average detection accuracy of 99.21% was reported on uncropped AWE database.

In [14] they used multiple scale faster region-based CNN for ear detection. A detection rate of 98%, 100% and 98.22% was reported on a test set of web image dataset, UND-J2 and UBEAR database respectively. In [15] and [16] ear detection using CNN proved very good results using faster R-CNN and geometric morphometrics respectively. This work mainly uses five ear detection techniques based on traditional approaches in computer vision. Because of CNN’s growing popularity and applications, our future research includes automatic detection and recognition works based on deep learning using CNN.

3. IMPLEMENTATION DETAILS OF PROMINENT EAR DETECTION ALGORITHMS

In this work four existing ear detection algorithms are selected for implementation. They are described below.

3.1 METHOD I-AUTOMATIC DETECTION USING MORPHOLOGICAL OPERATORS

Kumar and Wu [2] proposed morphological operators for automatic ear localization. Images were pre-processed by using a Gaussian filter with
parameters $\mu=20$ and $\sigma=5$. Application of this filter removes noise in the image. Histogram equalization is applied to this filtered image for contrast enhancement. The preprocessed image is binarized with otsu threshold. Various morphological operations are also performed on the pre-processed image for ear detection. The details of Method I are described in Algorithm 1. Finally the images are normalized to 50×180 size. Fig. 2 shows the algorithm output applied on a sample image in IIT Delhi ear database.

**Algorithm 1 Method I Ear Detection**

1: Pre-process the image $I(i,j)$ with Gaussian filter ($\mu=20$ and $\sigma=5$) followed by Histogram equalization.
2: Convert the pre-processed image to binary image $I_1(i,j)$ based on Otsu threshold.
3: Use top hat operation to the same pre-processed image.
4: Apply opening operation to the image in step 3 and create mask $M(i,j)$.
5: Multiply mask with original image to obtain $I_2(i,j)=I(i,j) \times M(i,j)$
6: Binarize the above image to obtain $I_3(i,j)$.
7: Multiply the binary image obtained by Otsu and binary image obtained by morphological operations to obtain $I_4(i,j)=I_1(i,j) \times I_3(i,j)$.
8: Ear image is dilated to connect broken edges.
9: Eliminate small object if its count is less than 80 by considering 8 connectivity to extract ear boundary

3.2 METHOD II: EAR LOCALIZATION USING SNAKE MODEL

Anwar et al. [17] proposed an ear localization using snake model. First the input image is pre-processed by applying Gaussian filter. Snake model is used to detect the ear in an image. Initially select the control points of the snake by manual clicking. Mainly the control parameters used for snake are $\alpha$ (alpha) which specifies the snake elasticity, $\beta$ (beta) which specifies the contour rigidity, $\gamma$ (gamma) denotes the step size, $\kappa$ (kappa) the scaling factor, and the weighing factors $W(Eline)$, $W(Eterm)$, $W(Edge)$. Algorithm 2 describes detailed working of this method.

**Algorithm 2 Method II Ear Detection Using Snake Model**

1: Apply Gaussian filter to the original image for noise removal and smoothing.
2: Perform ear extraction using snake model.
3: Change ear image to region mask.
4: Isolate Ear image only by multiplying original image with mask.
5: Apply median filter of size 5×5.
6: Convert ear image to binary image by applying Global threshold.
7: Apply canny operator to locate edges by non-maximal suppression.

The steps in the algorithm are illustrated in Fig. 3.

3.3 METHOD III: DETECTION BASED ON TEMPLATE MATCHING

Prakash et al. [8] used template matching for ear detection. Their method first consists of preprocessing which include segmentation of skin pixels from non-skin pixels. The edges are detected from the segmented regions based on canny edge detection. An offline ear template was initially created by averaging selected normalized ear samples from the database. This template is moved over the edge map to locate ear based on cross correlation value. Fig. 4 shows the flow chart of this method with sample output illustrated in Fig. 5.
3.4 METHOD IV: DETECTION BASED ON BANANA WAVELETS AND CIRCULAR HOUGH TRANSFORM

Banana wavelets are shaped wavelets that respond to curved regions in an image. In [5] an ear detection approach based on banana wavelets is proposed. Convolution of this wavelets over an image responds to more than one regions in an image. Due to elliptical/circular shape of the ear, circular Hough transform detects ear shape more accurately. Fig. 6 illustrates the steps used in this method and the output using this method is shown in Fig. 7.

3.5 METHOD V: MODIFIED APPROACH BASED ON METHOD IV

In [18] we propose an approach by modifying method IV. The main change is the application of a pre-processing stage prior to banana wavelet application. Pre-processing includes mainly three steps. It involves skin segmentation, adaptive histogram equalization using CLAHE and morphological operation top hat.

Skin detection is an important pre-processing step in images having human face. Since ear is located in the skin region, the first step in ear detection is the segmentation of skin region from non-skin parts. Color spaces can be used for skin segmentation. YCbCr is the color space used in [19] for segmenting skin regions based on the threshold 77<=Cb<=127 and 133<=Cr<=173. The skin segmented image undergoes contrast enhancement through adaptive histogram equalization (AHE). This method divides image into different parts and histogram equalization is applied to each part to improve the contrast of the whole image. CLAHE (Contrast limited adaptive histogram equalization) is a variant of AHE that limits the over amplification of noise in an image while enhancing the contrast. Morphological operation is then applied to enhanced image using structuring element. Morphological operation used in our work is white top hat morphological operation that highlights bright pixels around dark backgrounds.

The image after pre-processing is convolved with banana wavelets to find curved regions. Banana wavelets are shaped or bent gabor wavelets and mainly they are used for identification of the curved regions in an image. Fig. 8 shows the structure of banana and Gabor wavelets.

A banana wavelet $B^v$ is represented by a vector $v$ of four variables, i.e. $v = (f, \theta, c, s)$ where $f$, $\theta$, $c$ and $s$ are the frequency, orientation, curvature, and size respectively [20].

The filter is made from a rotated and curved complex wave function $F^v(x, y)$ and a Gaussian $G^v(x, y)$ function rotated and curved in the same way as $F^v(x, y)$ [20] where
\[ B^v(x, y) = v \cdot G^v(x, y) \cdot (F^v(x, y) - DC^v) \]  

And

\[ G^v(x, y) = \exp \left[ -\frac{f^2}{2} \left( \frac{(x + cx^2)^2}{\sigma_x^2} + \frac{x^2}{\sigma_y^2} \right) \right] \]  

\[ F^v(x, y) = \exp \left( i \cdot f \left( x \cdot \cos \alpha + y \cdot \sin \alpha \right) \right) \]  

\[ x_c = x \cdot \cos \alpha + y \cdot \sin \alpha \]  

\[ y_c = -x \cdot \sin \alpha + y \cdot \cos \alpha \]  

\[ DC^v = \frac{\int G^v(x, y) \cdot F^v(x, y) \, dx \, dy}{\int G^v(x, y) \, dx \, dy} = \frac{-\sigma_y}{\ell} \]  

where \( v, \sigma_x \) and \( \sigma_y \) respectively are constant, Gaussian filter scales in \( x \) and \( y \) directions respectively.

Banana wavelet convolution is applied to preprocessed image by using five filters with different curvatures (-0.5, -0.1, 0, 0.1, 0.5) and orientations (0, \( \pi/4 \), \( \pi/2 \), \( 3\pi/4 \), \( \pi \)). Convolving the preprocessed image with these filters, positions where the magnitude has local maxima are found. Banana wavelets find multiple curvilinear regions in the image and since ear contains lots of concentric curved circles, circular Hough transform detects ear in the given image.

Circular Hough transform [21] is a specialized Hough transform that finds circle candidates by voting in Hough parameter space. An accumulator matrix is used to find circles from Hough parameter space. Voting is done for all the possible circles in the accumulator space and the local maximum voted circles of accumulator gives the circular Hough space. The circular Hough transform finds triplets \((x, y, r)\) in 3D parameter space

\[ x = c_1 + r \cdot \cos(\alpha) \]  

\[ y = c_2 + r \cdot \sin(\alpha) \]  

\[ r^2 = (x - c_1)^2 + (y - c_2)^2 \]  

where \( c_1 \) and \( c_2 \) represents the circle centers in \( x \) and \( y \) direction respectively and \( r \) is the radius. The value of \( \alpha \) ranges from 0 to 360°.

The workflow of our proposed method is shown in Fig. 9 and steps using a sample image is shown in Fig. 10. The proposed algorithm is summarized as follows.

1: Read input image I.
2: Apply YCbCr skin segmentation to input image.
3: Convert segmented image to grey scale.
4: Apply CLAHE to the grey scale image for contrast enhancement.
5: Apply top hat Operation.
6: Convolution of banana wavelets with transformed image and application of circular Hough transform.
7: Automatic ear detection

4. EXPERIMENTS AND ANALYSIS

The implementations of five methods are done using Matlab. The experiments were conducted on two databases GTAV [22] and RR database [23]. The five methods implemented in this work are semi-automatic and fully automatic methods. CNN or deep based methods are not considered for the comparison or analysis.

4.1 DESCRIPTION OF DATASETS

GTAV is a publicly available free face database with varying pose and illumination. The database consists of 44 subjects with 27 pictures per person. Each image is a color image in BMP format with resolution 240×320. The images are in different pose...
angles $0^\circ$, $\pm 30^\circ$, $\pm 45^\circ$, $\pm 60^\circ$ and $90^\circ$ [22]. Three different sources of light like natural light, strong light source at angle $45^\circ$ and a mid-strong light source at front is used for illumination. Additional frontal view pictures are also included for each person. For our experimental study, we used a part of this dataset. Ten selected side face images of each person are used for experimental purpose. A database called RR database with hundred subjects was created for experimental study. There are three images with different angles and rotation for each subject. This database includes color images with a size of $240 \times 320$ pixels in BMP format. For automatic ear detection verification the class ear and non ear contains cropped and normalized images of $50 \times 180$ pixels. 5-fold cross validation is used to evaluate the model on the sample data.

4.2 RESULTS AND DISCUSSION

Detection result using few samples from GTAV database are shown in Fig. 11. GTAV is a face database and only side face images which contain ear portions are used for ear detection works. Fig. 12. shows ear localization result using two image samples in RR database by using the five methods.

![Figure 11 – Ear detection result of five methods using two sample image from GTAV Database](image1)

![Figure 12 – Ear detection result of five methods using two sample image from RR Database](image2)

The experiment was also conducted on images captured in various background illumination conditions. Fig. 14. shows the ear detection from sample images of a single person with various lightning conditions. Ear ring, spex, scarf and hair are the important objects that occlude the ear. The proposed system is tested with images having different types of occlusion like hair, spex and ear ring. Fig. 15. shows ear detection in images with occlusion.

![Figure 14 – Ear detection results with various illumination](image3)

![Figure 15 – Ear detection results with occlusion](image4)

The five algorithms gave detection failures in some cases. Failures occur fully or partially in situations like occlusion, illumination change, varying pose and noise are heavier. Examples of false detection are shown in Fig. 16.

![Figure 16 – Sample images with false detection results](image5)

4.3 DETECTION ACCURACY

In this work ear detection accuracy is calculated by using manual as well as automatic methods.
4.3.1 MANUAL VERIFICATION

The five methods are implemented and tested in GTAV and RR database. Ear detection is performed using each method in both databases. Manual testing is performed to check whether the extracted portion is ear or not. This type of manual detection accuracy is person specific. Table 1 shows the detection accuracy of the five methods using manual verification.

From Table 1, the detection accuracy of the proposed automatic ear detection method is higher compared to the other three methods M I, M III and MIV in RR database and GTAV database. Method II gave high detection accuracy due to snake model. Here the detection is done manually by using control points. Method 1 is applied to cropped side face ear images for localization. It cannot be applied to the whole face image. For method III the detection accuracy depends on the initially created proper ear templates.

Table 1. Detection accuracy of the five methods using manual verification

| Method | GTAV Detection rate % | RR Detection rate % |
|--------|-----------------------|---------------------|
| M I (Automatic detection of ear using morphological operators) | 86.25% | 85.36% |
| M II (Ear localization using snake model) | 95.43% | 93.32% |
| M III Ear Detection using Template Matching) | 86.45% | 81.66% |
| M IV (Ear Detection using Banana Wavelets and Circular Hough Transform) | 89.25% | 88.56% |
| M V (Proposed Method) | 92.27% | 91.24% |

4.3.2 AUTOMATIC VERIFICATION

An automatic verification technique was also employed in this work to find the ear detection accuracy. Two classes are formed for classification from the segmented ear images obtained by the five methods in both RR and GTAV databases. Class Ear and class Non Ear. Correctly segmented ear portions are placed in one class called class ear and incorrectly segmented parts are placed in another class called non ear. LBP [24] and gabor features [25] are extracted and a trained model was created by using SVM and KNN. The accuracy of KNN model was found to give good result compared to SVM. This model was used to test the images from each method. Table 2 summarizes the test result of each method. Automatic verification result along with manual verification are given in table. It follows that LBP feature and KNN classification gives result closer to manual verification.

Table 2. Detection rates of five methods using different classifiers

| Method | Features Extracted | SVM Accuracy | Average detection Accuracy (Manual verification) |
|--------|--------------------|-------------|-----------------------------------------------|
| M I    | LBP Gabor          | 76.67%      | 85.81%                                        |
| M II   | LBP Gabor          | 82.89%      | 94.38%                                        |
| M III  | LBP Gabor          | 75.78%      | 84.05%                                        |
| M IV   | LBP Gabor          | 80.12%      | 88.90%                                        |
| M V    | LBP Gabor          | 81.24%      | 91.75%                                        |

5. CONCLUSION

This paper implements four ear localization techniques used in 2D ear images and then proposes a modification to one of them. All the five algorithms were applied to two datasets and a comparative analysis of ear detection using manual method is presented. Further an automatic evaluation method for ear detection is also experimented with.

6. REFERENCES

[1] A. V. Iannarelli, *Ear Identification*, Paramount Publishing Company, Freemont, California, 1989, 213 p.
[2] A. Kumar and C. Wu, “Automated human identification using ear imaging,” *Pattern Recognition*, vol. 45, issue 3, pp. 956-968, 2012.
[3] B. Moreno, A. Sanchez, and J. F. Velez, “On the use of outer ear images for personal identification in security applications,” *Proceedings of the IEEE 33rd Annual 1999 International Carnahan Conference on Security Technology*, Madrid, 1999, pp. 469-476. doi: 10.1109/CCST.1999.797956.
[4] M. Choras, “Ear biometrics on geometrical feature extraction,” *Articulated Motion and Deformable Objects*, vol. 3179, pp. 51-61, 2004.
[5] L. Jacob, and G. Raju, “Automatic ear segmentation using banana wavelets and Hough transform,” *International Journal of Applied Engineering Research*, vol. 10, pp. 24979-24990, 2015.
[6] A. Abaza, A. Ross, C. Hebert, M.A.F. Harrison, and M. S. Nixon, “A survey on ear biometrics,” *ACM Computing Surveys*, vol. 45, issue 2, pp. 1-35, 2013.
[7] M. Burge and W. Burger, “Ear biometrics in computer vision,” Proceedings of the 15th International Conference on Pattern Recognition ICPR-2000, Barcelona, Spain, 2000, pp. 822-826.

[8] S. Prakash and P. Gupta, “An efficient ear localization technique,” Image Vision Comput., vol. 30, pp. 38-50, 2012. Doi: 10.1016/j.imavis.2011.11.005.

[9] B. Arbab-Zavar, and M. S. Nixon, “Robust log-Gabor filter for ear biometrics,” Proceedings of the International Conference on Pattern Recognition, Tampa, FL, 2008, pp. 1-4.

[10] J. Liu, Y. Gao, Y. Li, “Few-example affine invariant ear detection in the wild,” Structural, Syntactic, and Statistical Pattern Recognition, vol. 11004, pp. 248-257, 2018.

[11] Y. Li and Z.-C. Mu, “Ear detection based on skin-color and contour information,” Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, ICMLC, 2007, pp. 2213-2217. Doi: 10.1109/ICMLC.2007.4370513.

[12] P. Peer, V. Struc, Ž. Emersiš, and G. L. Lan, “Convolutional encoder–decoder networks for pixel-wise ear detection and segmentation,” IET Biometrics, vol. 7, issue 3, pp. 175-184, 2018.

[13] V. Badrinarayanan, A. Kendall, R. Cipolla, “SegNet: A deep convolutional encoder-decoder architecture for image segmentation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, issue 12, pp. 2481-2495, 2017. https://doi.org/10.1109/tpami.2016.2644615

[14] Y. Zhang and Z. Mu, “Ear detection under uncontrolled conditions with multiple scale faster region-based convolutional neural networks,” Symmetry, vol. 9, issue 4, p. 53, 2017.

[15] S. El-Naggar, A. Abaza and T. Bourlai, “Ear detection in the wild using faster R-CNN deep learning,” Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Barcelona, 2018, pp. 1124-1130.

[16] C. Cintas et al., “Automatic ear detection and feature extraction using geometric morphometrics and convolutional neural networks,” IET Biometrics, vol. 6, issue 3, pp. 211-223, 2017.

[17] A. S. Anwar, K. K. A. Ghany, and H. Elmahdy, “Human ear recognition using geometrical features extraction,” Procedia Computer Science, vol. 65, pp. 529-537, 2015, Doi: 10.1016/j.procs.2015.09.126.

[18] K. R. Resmi and G. Raju, “A novel approach to automatic ear detection using banana wavelets and circular Hough transform,” Proceedings of the 2019 International Conference on Data Science and Communication (IconDSC), Bangalore, India, 2019, pp. 1-5.

[19] M. C. Shin, K. I. Chang, and L. V. Tsap., “Does colorspace transformation make any change on skin detection?” Proceedings of the IEEE Workshop Applications of Computer Vision, 2002, pp. 275-279.

[20] M. I. Ibrahim, M. S. Nixon, and S. Mahmoodi, “The effect of time on ear biometrics,” Proceedings of the International Joint Conference on Biometrics (IJCB), Washington, DC, 2011, pp. 1-6.

[21] D. H. Ballard, “Generalizing the Hough transform to detect arbitrary shapes,” Pattern Recognition, vol. 13, issue 2, pp. 111-122, 1981.

[22] https://gtav.upc.edu/en/research-areas/facedatabase

[23] https://eardatabase.blogspot.com/2019/12/rr-database.html

[24] T. Ojala, M. Pietikäinen, and D. Harwood, “Performance evaluation of texture measures with classification based on Kullback discrimination of distributions,” Proceedings of the 12th IAPR International Conference on Pattern Recognition (ICPR), 1994, vol. 1, pp. 582-585.

[25] T. Aach, A. Kaup and R. Mester, “On texture analysis: Local energy transforms versus quadrature filters,” Signal Processing, vol. 45, issue 2, pp. 173-181, 1995.

Resmi K. R., is a Ph.D student at school of computer science, Mahatma Gandhi university, Kerala, India. She received her master’s in computer applications from Mahatma Gandhi university in 2008. Her research interests include biometric, pattern recognition, image processing and computer vision.

Raju G., received the Ph.D. degree in computer science from University of Kerala, India, in 2013. He is currently working as professor in the department of data science, Christ deemed to be university, Lavasa Campus Pune, India. He is serving as an editorial member and reviewer of several international reputed journals. He has authored many research articles/chapters in books. His areas of interest include document image processing, data mining, medical image processing and biometric.