Automatic 2D Ear Detection: A Survey

Resmi K R, Raju G

Abstract: Human ear is a passive biometric and recognizing people by ear is popular in multimodal biometrics. Ear recognition is performed based on the features extracted from the detected ear structure. Detection in ear biometrics deals with localizing ear or region of interest (ROI) from the input image .Ear detection is the first and main step for an ear recognition system. Accuracy of the detection module affects the overall accuracy in an ear recognition system. Due to the complex structure of the ear, its detection is very challenging. The simplest approach for detection is manual cropping and automatic ear detection is a challenging task. In this paper, we present a survey on semi and fully automatic 2D ear detection approaches with detection accuracies. The detection accuracies found to be reduced in situations like occlusion by hair, specs, ear rings, illumination change, change in pose etc. Also a description of publicly available databases is presented for researchers.

Index Terms: Biometric, Deep learning, Ear databases, Ear detection, Morphological operators, Template matching, Wavelets

1. INTRODUCTION

Biometric authentication gained popularity due to security, access control and authentication. Biometric recognition is based on who you are instead of what you have or what you know. Ear is a physiological and passive biometric which uses the shape of ear structure. The usage of human ear for identifying a person was proposed by the famous criminologist Bertillon in 1890. Iannarelli[1] experiment showed that ear is a unique biometric even in the case of identical twins. Ear has certain unique features compared to other biometric traits like retina, iris, face etc. According to medical reports [1] ear structure is constant in the age group 8 to 70. However ear stretches in the age group 4 months to 8 years. Further, the ear structure tends to change when a person reaches 70 years. Recognition of ear is not affected by variation in pose and facial expression as it happens with other biometric like face [2].It is a passive biometric which does not need cooperation of subject and it is free from several other biometric problems like anxiety, privacy, and hygiene [2].More uniform color distribution [3] of ear, helps us to keep almost all information when converting the original image into grey scales. Smaller size and lower resolution [4] helps to work faster compared to other modalities. Unlike finger/palm print, ear is relatively free from wear and tear. Now a days multimodal biometric systems are common. Ear biometric can be used with other biometrics in multimodal systems [4]. The first step in ear biometrics is localization of ear from images for which different methods are proposed in literature. Automatic ear detection is a challenging task and many ear recognition system uses manually cropped ears. In this paper, we present a survey on popular ear detection methods in 2D. The paper is articulated as follows. Section 2 discusses important automatic and semi-automatic ear detection techniques followed by listing the standard ear databases in section 3.Discussion and conclusion are presented in section 4 and section 5 respectively.

2 AUTOMATIC 2D EAR DETECTION TECHNIQUES

The performance of ear recognition system mainly depends on the ear detection quality. Ear detection can be performed by various means like template matching, morphological operators, hybrid techniques etc. Manual cropping of ear from side face images is simple when compared to automatic ear detection. A detailed review of 2D automatic ear detection technique is presented as follows.Ear detection using contour based approach was used in many works.Burge and Burger [5] used deformable contours to develop the first well known ear detection system .Canny edge detection is applied to find the edges and edge relaxation method is applied to find larger curve segments. The detection is not fully automatic since user has to initialize the contour. Ansari and Gupta [6] used canny edge detector to find outer ear helix edges. The outer helix curves were localized using convex and concave edges from chain code. Their experiment was conducted on IITK database with detection accuracy of 93.34%.Yuan and Mu [7] proposed an ear detection based on two stages. The face profiles are tracked from the video using color based tracking method CAMSHIFT. A modified CAMSHIFT algorithm was used for profile tracking. In the modified approach skin color histograms are calculated from multiple face images and it was averaged for future calculation of probability distribution. During the second stage ear was localized using elliptical and coutour based fitting method.

Morphological methods used for ear detection was performed in [2] and [8].Kumar and Wu [2] developed an automatic ear detection system which uses morphological operators along with fourier descriptors to extract ear shape. The exterior ear boundaries obtained after different morphological operations are not smooth and fourier descriptors are applied to obtain smooth ear boundaries. Their experiments were conducted on IIT Delhi database. Said et al. [8] employed mathematical morphology for automatic ear detection. They first applied filtering operation to the original image by using morphological top hat operation. K means clustering algorithm was then applied to filtered image to produce a binary image. To find the regions of interest they used 8 connected component labelling algorithm. Finally refinement stage was applied to remove smaller connected components to detect ear portion. Their method achieved more than 90% detection rate on three different dataset. Arbab-Zavar and Nixon [9] employed hough transform to find ear shape. Their experiment was conducted on UND database with 91% detection rate without occlusion. The detection rate of algorithm was reduced to 83% with 40% occlusion. Ear detection using wavelets was proposed in [10] and [11].Arbab-Zavar and Mark S Nixon [10] proposed a new approach based on Log gabor filters and wavelet transform. The algorithm offered a detection rate of 88.4% on XM2VTS

- *Resmi K R is currently pursuing PhD in computer science from Mahatma Gandhi University, Kerala, India E-mail: resmykr@gmail.com*
- *Raju G is currently working as professor, Department of computer science and Engineering , Christ, Bengaluru, India,*

database of 63 individuals. Ibrahim et al. [11] used banana wavelets which are derived from gabor wavelets to extract curvilinear structures in the ear. Their preprocessing step consists of skin detection based on color and texture analysis. The edges of the images are enhanced by applying adaptive histogram equalization. Their technique is fully automatic and their method gives good results even in case of images with high noise. Their experiment was conducted on XM2VTS and SOTON database. On XM2VTS database the detection accuracy was 100%.Cummings et al. [12] used the image ray transform, based on light rays for ear detection. This transform highlights tubular and curved structures in an image. Since ear image is elliptical, this method was used to segment the ear accurately. On XM2VTS database this technique reported a detection rate of 99.6%.Template based approach for automatic ear detection was used in [13],[14],[15] and[16].Prakash et al. [13] used distance transform and template based method for ear detection. First preprocessing is applied to the original image to segment skin region using color based segmentation. Canny edge detection is applied to obtain the edge map and the unwanted edges are removed using edge pruning based on length and curvature. Initially a proper ear template was created from a set of ear images taken from the database. Ear was localized by finding cross-correlation value at each pixel by moving the distance transform of the template over the distance transform of face edge image. The experiment was evaluated on IITK ear database of 150 images with detection accuracy of 95.2%. Failure in detection occurs with images having low quality and high occlusion with hair. Chen and Bhanu [14] used a two-phase system to extract ears from face images. First is offline model template building followed by on-line detection. Offline model template was created by averaging histograms of manually extracted ears. Their database consists of 30 subjects with two side face range images each. They reported a 91.5% detection rate. Prakash and Gupta [15] developed an automatic ear localization technique from the whole side face image which is scale and pose invariant. They first constructed edge map from the side face image and from the edge map connected components of the graph was formed to detect probable ear portion. Then true ear was identified among probable ear portions by using ear template matching. The experiment was conducted on IITK database with detection accuracy 99.25%.Sarangi et al. [16] used a technique by modifying the hausdorff distance for ear detection .Their method first segment skin region from non-skin region. Ear template was created to detect various ear shapes.Hausdorff distance measure the similarity likeness between the template and input image. Their experiments was conducted on CVL and UND-E database with detection accuracy of 91% and 94.54% respectively. Wahab et al. [17] used HEARD, an automatic ear localization technique based on three features in their work. The first one was the ratio between height and width of the ear and it should be greater than 0.5. Second one was the relationship between perimeter and number of pixels in boundary area. It should exceed 5. Last one was the fact that ear contains lots of semi circles and all semi circles are detected in the range 60% to 90%. The system was tested on 200 random images from UND database with detection accuracy of 98%.Recently ear detection techniques use deep learning instead of conventional machine learning algorithms. Emersic et al [18] developed an ear detection system based on convolutional Encoder-decoder networks that works well on even unconstrained environment. They used a two-step process for ear detection. In the first phase they used SegNet [19] , a convolutional encoder-decoder that allowed us to label each pixel as ear or non-ear. Instead of returning a bounding box around the detected ear, their method provided pixel wise information about ear location. The second step is post processing which discards additional pixels that are wrongly classified by CNN as ear portion. Their experiment was conducted on original uncropped images of Annoted web ears database [AWE]. Their method achieved an average detection accuracy of 99.21%. The method performed well compared to the results with haar based detector. Multiple Scale Faster Region-based Convolutional Neural Networks (Faster R-CNN) is another recent ear detection technique proposed in [20]. The detection system was trained to detect three regions head, pan ear region and ear. The detection accuracy was improved from the traditional faster R-CNN by combining morphological characteristics and location context of the ear. The experiment was conducted on three databases Webear, UND J2, UBEAR dataset with detection accuracy of 98%, 100% and 99.22% respectively. An ear detection system using Faster R-CNN was proposed in [21]. Their training consist of two stages. First classification was done by using an AlexNet model to classify ear and non-ear segment. Second the unified region proposal network with Alexnet was trained for ear detection. Their experiment was conducted on images from UND, FERET, WVU and AWE databases with maximum of 98% detection rate. Yuan li and lu fei [22] developed a real time detection of ear based on embedded system. The framework used for ear detection is YOLO. YOLO is an object detection system used for real time processing. They improved the detection accuracy by modifying the tiny yolov2 network which is an enhancement version of YOLO in terms of speed. An external camera was installed on Nvidia Jetson tx2 board for automatic ear detection in real time environment. Their method achieved a good detection accuracy in case of complex conditions such as poor illumination and partial occlusion. Their experiment was conducted on USTB Webear dataset and the USTB Helloear dataset with a detection accuracy of 96.17% and 98.67% respectively

3 EAR DATABASES

Ear biometric databases helps researchers to carry out ear detection experiments and compare their results. The database is classified in to 2D or 3D depending on acquisition device. Table 1 gives a list of selected set of 2D databases. An overview of the databases are given below.

3.1 University of Science and Technology Beijing (USTB)

USTB database contains a large collection of ear images. It is a free database. USTB contains 4 ear image datasets.

3.1.1 Dataset 1

It contains 60 volunteers with 3 grey scale images per subject. The 3 images are normal, angle rotated and one with illumination variation.

3.1.2 Dataset 2

It contains 77 subjects with images of varying angle and illumination. Four color images of each person are captured to create the database. Among the four images, first and fourth are images with change in illumination. The second and third are profile images with +30 degree and -30 degree angle variation.

3.1.3 Dataset 3
It contains 79 subjects with images of partial occlusion, depth variation and full side face image. The database include left rotated, right rotated, and occluded images. This database can be used by fusing information from ear and face in multimodal biometric.

3.1.4 Dataset 4
This database include both color and greyscale images of 500 volunteers. The images include normal, up, down, left and right rotated images.

3.2 UND -University of Notre Dame
UND is a publicly available free database with 4 different collections. Collection E contains 464 light visible side face profile images of 114 subjects. Collection F includes 942 3D and its corresponding 2D profile ear images from 302 subjects. Collection G contains 235 subjects with 738 3D and its corresponding 2D ear images. Collection J2 is the highest ear database collection with 1800 3D and 2D ear images from 415 subjects.

3.3 IIT Delhi- Indian Institute of Tech
IIT Delhi is a free database with two collections. First ear database has 121 subjects. The database has 471 ear images which consists of at least 3 images per subject. Second is a preprocessed and cropped database of 754 ear images from 212 users.

3.4 IIT Kanpur
This database is a free database which include two subsets. Subset1 contains 801 illumination invariant side face images from 190 subjects. Subset 2 include images with pose variation. It consists of 9 images for 89 subjects.

3.5 UBEAR Dataset
Images in this database is captured from video in real world conditions. This is a free database which include grey scale images. The database include images with different illumination, occlusion and pose change. Subject pose change include watching to camera, watching down or up.

3.6 AMI- Mathematical Analysis of Images (AMI) Ear Database
The AMI is a free database which consists of 700 ear images of 100 subjects captured from the computer science department of ULPGC, Las palmas, Spain. The image is captured in an indoor environment. All the individuals were in the age 19-65 years. Seven images are captured for each subject with six right ear images and one left ear image.

3.7 XM2VTS Database
This database consists of 295 subjects which has multimodal data. The database consists of face images and voice input data. Each image consists of two head rotation and six speech shots. It is not a free database.

3.8 CVL Database
The database consists of 114 subjects with 7 images of each subject. Most of the subjects are men (90% men and 10% women).This database include both left and right ear of each persons. The images are captured under uniform illumination.

3.9 Annotated Web Ears (AWE) Database
The AWE dataset [29] contains 100 subjects. For each person there are 10 ear images. The images differ in terms of quality and size. The AWE was collected from web images. AWE is a publically available free database.

Table 1 *Summary of ear image dataset*

Database	Number of subjects	Total Images
USTB[23]		
Dataset 1	60	180 images
Dataset 2	77	308 images
Dataset 3	79	1738 images
Dataset 4	500	8500 images
UND[24]		
Collection E	114	464images
Collection F	302	942 images
Collection G	235	738 images
Collection J2	415	1800 images
IITD[2]		
Subset1	121	471 images
Subset 2	212	754 images
IITK[9]		
Subset 1	190	801 images
Subset 2	89	801 images
UBEAR[25]	126	4330 images
AMI[26]	100	700 images
XM2VTS[27]	295	4 recordings
CVL[28]	114	798 images
AWE[29]	100	1000 images

4 DISCUSSION
A summary of different detection methods reviewed is given in Table 2. The observations based on Table 2 are given below.

Table 2 *Summary of Detection Methods*

Paper reference	Ear Detection Method	Database	Detection Accuracy
Detection Methods			
Burge and Burger [5]	Deformable contours	NA	NA
Ansari and Gupta [6]	Canny edge detector	IITK database	93.34%.
Yuan and Mu [7]	Modified CAMSHIFT algorithm		
Kumar and Wu[2]	Morphological operators	IIT Delhi	NA
Said et al. [8]	Mathematical morphology	Various Databases	>90%
Arbab-Zavar and Nixon[9]	Hough transform	UND database	91%
Arbab-Zavar and Nixon[10]	Log gabor filters	XM2VTS database	88.4%
Ibrahim et al. [11]	Banana wavelets	XM2VTS	100%
Cummings et al. [12]	Ray transform	XM2VTS	99.6%.
Prakash et al. [13]	Distance transform and template	IITK database	95.2%.
Chen and Bhanu [14]	Template matching		91.5%
Prakash and Gupta [15]	Edge map and connected components	IITK	99.25%.
Sarangi et al [16]	Modified hausdorff distance	CVL UND-E	91% 94.54%
Wahab et al. [17]	HEARD	UND	98%
Detection Using Deep Learning			
Emersic et al [18]	Convolutional Encoder-decoder networks	AWE	99.21%
Zhang and Mu [20]	Multiple Scale Faster Region-based Convolutional Neural Networks	AWE UND J2 UBEAR	98% 100% 99.22%
Naggar et al. [21]	Faster R-CNN	Various Databases	98%
Yuan li and lu fei [22]	YOLO	USTB Web ear dataset USTB Helloear dataset	96.17% 98.67%

Among different detection methods discussed, XM2VTS and IITK database gave high detection accuracy compared to other databases in traditional ear detection techniques. With Deep learning the detection rate is very high when compared to older methods. Annotated web ear database (AWE) was used mainly for deep learning works due to its larger size. AWE is also one of the most important database because it uses images in unconstrained settings. From Table 2 the comparison of different detection methods is not possible since it uses different databases and different methods. The presence of occlusion (like ear ring, hair, glasses etc.), illumination change, lightening conditions, pose etc. affects the accuracy of detection. This survey presents a detailed study of significant works in ear detection, giving more importance to automatic detection techniques.

5 CONCLUSION

In this paper, we have reviewed some of the prominent works in 2D ear detection. This review also provide information about various factors regarding ear databases such as number of images, cost etc. This is a detailed survey of detection works including deep learning. With deep learning techniques, the detection accuracy is very high when compared to traditional detection methods.

REFERENCES

[1] Iannarelli, A. V, “Ear identification,” in Proceedings of International Workshop Frontiers in Handwriting Recognition, Paramont Publishing Company,Freemont,California ,1989.
[2] Ajay Kumar and Chenye Wu, “Automated human identification using ear imaging,” Pattern Recognition, vol. 45, Issue 3, pp. 956-968, 2012.
[3] Moreno, B., Sanchez, A., and Velez, J. F, “On the use of outer ear images for personal identification in security applications,”in Proceedings IEEE 33rd Annual 1999 International Carnahan Conference on Security Technology (Cat. No.99CH36303) , Madrid: doi: 10.1109/CCST.1999.797956,1999,pp. 469-476.
[4] Choras, M, “Ear Biometrics on Geometrical Feature Extraction,” Articulated Motion and Deformable Objects, 3179, 51-61 ,2004.
[5] M. Burge and W. Burger, "Ear biometrics in computer vision," in Proceedings 15th International Conference on Pattern Recognition. ICPR-2000, Barcelona, Spain, 2000, pp. 822-826.

[6] Ansari, S., and Gupta, P. “Localization of ear using outer helix curve of the ear,” In International Conference on Computing: Theory and Applications ,March 2007,pp. 688-692.
[7] Yuan, Li and Mu, Zhi-Chun, “Ear Detection Based on Skin-Color and Contour Information,” in Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, ICMLC. 4. 10.1109/ICMLC.2007.4370513, 2007, pp. 2213 - 2217.
[8] Said, E. H., Abaza, A., and Ammar, H, “Ear segmentation in color facial images using mathematical morphology,” Biometrics Symposium,doi:10.1109/bsym.2008.4655519,2008.
[9] Banafshe Arbab-Zavar and Mark S. Nixon, “On shape-mediated enrolment in ear biometrics,” in Proceedings of the 3rd international conference on Advances in visual computing - Volume Part II (ISVC'07), Vol. Part II, Springer-Verlag, Berlin, Heidelberg, 2007, pp. 549-558.
[10] Arbab-Zavar, Banafshe and S. Nixon, Mark, “Robust Log-Gabor Filter for Ear Biometrics,” in Proceedings International Conference on Pattern Recognition, 2009, pp. 1-4.
[11] Mina Ibrahim, Mark S. Nixon and Sasan Mahmoodi “Shaped Wavelets for Curvilinear Structures for Ear Biometrics,” ISVC, 6453. pp. 499-508. 10.1007/978-3-642-17289-2_48, 2010.
[12] Cummings, A., Nixon, M., and Carter, J, “ A novel ray analogy for enrolment of ear biometrics,” Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS),2010,pp. 38,39,41.
[13] Surya Prakash and Phalguni Gupta, “An efficient ear localization technique,” Image Vision Comput. 30, 38-50,10.1016/j.imavis.2011.11.005, 2012.
[14] Chen, H., and Bhanu, B, “ Human ear detection from side face range images,”Proc. International Conference on Pattern Recognition, 3, 2004, pp. 574-577.
[15] Surya Prakash and Phalguni Gupta, “Ear Detection in 2D,” Ear Biometrics in 2D and 3D, Vol. 10. pp. 21-49, Springer, 2015.
[16] Sarangi P.P., Panda M., Mishra B.S.P., and Dehuri S. “An Automated Ear Localization Technique Based on Modified Hausdorff Distance,” in Proceedings of International Conference on Computer Vision and Image Processing (CVIP), 2017, pp. 229-240.
[17] Wahab, N.A., Hemayed, E.E., and Fayek, M.B, “HEARD: An automatic human EAR detection technique,” in International Conference on Engineering and Technology (ICET), 2012, pp 1-7.
[18] Peer Peter, Štruc Vitomir , Emeršič, Žiga and Lan Gabriel Luka, “ Convolutional encoder–decoder networks for pixel-wise ear detection and segmentation,” IET Biometrics, vol. 7, Issue 3, pp. 175-184, 2018.
[19] Badrinarayanan, V., Handa, A., and Cipolla, R, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling,” CoRR, Vol. abs/1505.07293, 2015.
[20] Zhang Yi and Mu Zhichun, “Ear Detection under Uncontrolled Conditions with Multiple Scale Faster Region-Based Convolutional Neural Networks,” Symmetry, vol. 9, Issue 4, p.53, 2017
[21] El-Naggar, Susan Abaza Ayman and Bourlai Thirimachos, “Ear Detection in the Wild Using Faster R-CNN Deep Learning,” 1124-1130. 10.1109/ASONAM.2018.8508487, 2018.
[22] Yuan L., and Lu F, “Real-Time Ear Detection Based On Embedded Systems,” in International Conference on Machine Learning and Cybernetics (ICMLC), Vol. 1, 2018, pp.115-120 .
[23] http://www1.ustb.edu.cn/resb/en/index.htm
[24] https://cvrl.nd.edu/projects/data/
[25] R. Raposo, E. Hoyle, A. Peixinho and H. Proenca, "UBEAR: A dataset of ear images captured on-the-move in uncontrolled conditions," in IEEE Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM), Paris, 2011, pp. 84-90.
[26] http://ctim.ulpgc.es/research_works/ami_ear_database/
[27] http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/
[28] Peer, Peter. CVL Face Database (2010).
[29] Ž. Emeršič, V. Štruc, and P. Peer, “Ear Recognition: More than a Survey”, Neurocomputing, Vol. 255, pp. 26-39, 2017.