A Survey on Optimization based Ear Recognition system

[1]Payal Verma, [2]Mr. Sandeep Patil

[1]M. Tech. Scholar, [2]Associate professor
Department of Electronics and Telecommunication,
F. E. T., S. S. G. I., Junwani, Bhilai (C.G.)

***Abstract*—Biometric are automated methods of recognizing a person based on a physiological or behavioral characteristic. Most of the well known ear biometric techniques have focused on recognition on manually cropped ears and have not used automatic ear detection and segmentation. This is due to the fact that detection of ears from an arbitrary profile face image is a challenging problem as ear images may vary in scale and pose (due to in-plane and out-of-plane rotations) under various viewing conditions. However, for an efficient ear recognition system, it is desired to detect the ear from the profile face image in an automatic manner. There exist few techniques in the literature which can be used to detect ear auto-matically. A detailed review of these techniques is as follows. The first well known technique for ear detection is due to Burge and Burger [1]. It has detected ears with the help of deformable contours. But contour initialization in this technique needs user interaction. As a result, ear localization is not fully automatic. Hurley et al. [2] have used force field technique to get the ear location. The technique claims that it does not require exact ear localization for ear recognition. However, it is only applicable when a small background is present in ear image. In [3], Yan and Bowyer have used manual technique based on two-line landmark to detect ear where one line is taken along the border between the ear and the face while other line is considered from the top of the ear to the bottom. The 2D ear localization technique proposed by Alvarez et al. [4] uses ovoid and active contour (snake) [5] models. Ear boundary is estimated by fitting the contour of an ear in the image by combining snake and ovoid models. This technique requires an initial approximated ear contour to execute and hence cannot be used in fully automated ear recognition system. There is no empirical evaluation of the technique.**

***Index Terms*—Ear recognition, Optimization, Automation, biometric, two dimensional analysis**

I. INTRODUCTION

All Yan and Bowyer [6] have proposed another technique by considering a predefined sector from the nose tip as the probable ear region. It first computes the ear pit using the curvature information obtained from 3D data and uses its boundary to initialize active contour which detects the ear boundary. It fails if the ear pit is occluded. It produces 78.79 % correct ear segmentation when only color information is used for active contour conversion. Ansari and Gupta [7] have presented an ear detection technique based on edges of outer ear helices. The accuracy of the this technique is reported to be 93.34 % on 700 sample images collected at IIT Kanpur. The technique solely relies on the parallelism between the outer helix curves and does not use any structural information present in inner part of the ear and hence, it may fail if the helix edges are poor. Yuan and Mu [8] have proposed a technique based on skin-color and contour information. It detects ear by roughly estimating the ear location and by improving the localization using contour information. It considers ear shape elliptical and fits an ellipse to the edges to get the accurate position of the ear. There is no quantitative evaluation reported for the technique.

Another ear localization technique which exploits the elliptical shape of the ear has been proposed in [9]. It has been tested on 252 images of 63 individuals selected from XM2VTS [10] and 942 image pairs of 302 subjects of UND database. For XM2VTS database which is relatively small and has less complex images, the tech-nique has achieved 100 % detection rate. However for UND database which contains complex images, it has offered only 91 % detection rate. Moreover, the assumption of considering ear shape elliptical for all subjects may not be true and hence, may not help in detecting the ear, in general. For example, as shown in Figure, assump-tion of elliptical boundary may correctly approximate the ear boundaries for round and oval shapes but may fail in case of triangular and rectangular shapes. Also, this assumption restricts the ear localization to a controlled environment as the presence of background objects may produce false positives.

In [11], Sana et al. have given a template based ear detection technique where to detect ears at different scales, ear templates of different sizes are maintained. In practice, any predefined set of templates may not be able to handle all situa-tions. Experimental study in this technique has used 1800 images collected from 600 individuals. However ear detection accuracy is not reported explicitly in the paper. In [12, 13], there are two techniques for ear localization which are also based on template matching. In these techniques, an ear template which is created off-line is resized to obtain a template of suitable size. Resizing is done using the size of the skin part of profile face image which works well when profile face includes only facial parts. But while capturing the profile face, an image may include other skin parts such as neck. This makes the size of the skin area larger than the actual and leads to an incorrect resizing of the ear template and hence, it produces an erroneous ear

localization. Techniques in [12, 13] have been tested on part of IIT Kanpur ear data-base containing profile face images of 150 individuals and found to have accuracy of 94 % and 95.2 % respectively. Attarchi et al. [14] have proposed an ear detection technique based on the edge map. It relies on the hypothesis that the longest path in edge image is the outer boundary of the ear. It works well only when there is small background present around the ear and fails if ear detection is carried out in whole profile face image. Performance of ear detection of this technique has been reported on two databases, namely USTB database which contains

308 ear images from 77 persons [15] and Carreira-Perpinan database which includes 102 ear images from 17 persons [16]. Accuracy has been found to be 98.05 % for USTB database and 97.05 % for Carreira Perpinan database. A cascaded AdaBoost based ear detection approach has been proposed in [17]. The technique uses Haar-like rectangular features as the weak classifiers. AdaBoost is used to select good weak classifiers and then to combine them into strong classifiers. A cascade of classifiers is built which works as the final detector. The detection performance of the cascaded ear detector has been evaluated for 203 profile face images of UND database and is reported to have accuracy of 100 %. However, the technique needs huge amount of time for training and has been tested on relatively small set of images.

In [18], an ear localization technique has been proposed which is based on hierar-chical clustering of the edges. To identify the edge cluster related to ear, the technique assumes approximate size of the ear cluster. Because of this, it works well when scale of the profile face image does not vary much. The technique is rotation invariant. However to handle scale, cluster size of the ear needs to be adjusted which may not be possible without user intervention. The technique has been tested on a database consisting of 500 profile face images of human profile faces collected at IIT Kanpur and found to have an accuracy of 94.6 %.

In [19], an ear detection technique using the image ray transform has been presented. The transform is capable of highlighting the tubular structures of the ear such as helix. The technique exploits the elliptical shape of the helix to perform the ear localization. However, assumption of ear shape being elliptical may be very rigid. The technique has achieved 99.6 % ear detection on 252 images of the XM2VTS database [10]. Ibrahim et al. [20] have employed a bank of curved and stretched Gabor wavelets (popularly called banana wavelets) for ear detection. A 100 % detection rate is achieved by this technique on images of XM2VTS database. In [21], a technique for ear detection has been presented by Kumar et al. where skin-segmentation and edge detection has been used for initial rough ear region localization. Region based active contour technique [22] has been further applied to get exact location of ear contours. The technique has been tested on 700 ear images and has achieved 94.29 % correct ear detection. This technique is applicable only when small background is present in the ear images. It can be observed that most of the techniques discussed above which have achieved almost 100 % correct ear detection rate have been tested on small data sets (≤300 images).

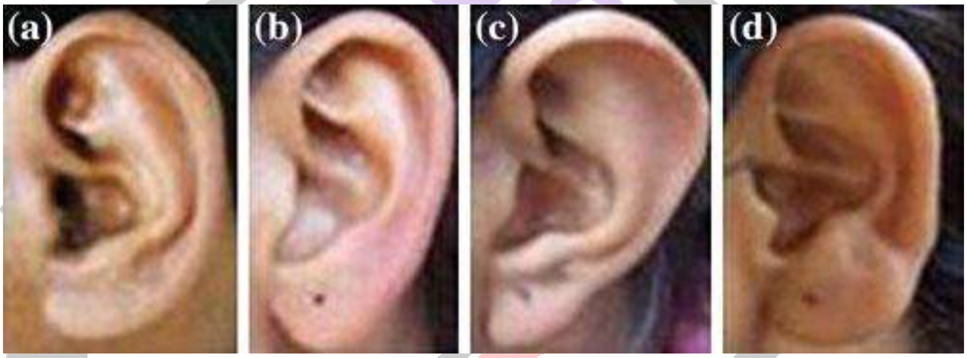

Figure 1. *Different ear shapes.* ***a*** *Round,* ***b*** *oval,* ***c*** *triangular,* ***d*** *rectangular*

Most of these techniques can detect the ear only when a profile face image con-tains a small background around the ear. These techniques are not very efficient,particularly when profile face images are affected by scaling and rotation (pose vari-ations). Moreover, they are not fully automatic and fast enough to be deployed in realtime applications. However, it is often required, specially in non-intrusive appli-cations, to detect the ear from a whole profile face image which may be affected by scale and pose variations.

This paper discusses an efficient ear localization technique which attempts to address these issues. The technique is invariant to scale, rotation and shape. It makes use of connected components of a graph constructed with the help of edge map of the profile face image to generate a set of probable ear candidates. True ear is detected by performing ear identification using a rotation, scale and shape invariant ear template.

II. *COLOR BASED SKIN SEGMENTATION*

This section presents a color based technique to segment skin and non-skin regions. It is similar to the skin segmentation technique proposed in [23] which has used 1976 CIE Lab color space for image representation. However, we have represented images in YCbCr space because it is perceptually uniform [24] and is widely used in video compression standards such as JPEG and MPEG [25].

The technique is capable of adapting different skin colors and lighting conditions. It performs skin segmentation in *YCbCr* color space as it is more suitable for charac-terizing skin colors. It first converts an image from *RGB* color space to *YCbCr* color space and then uses *YCbCr* color information for further processing. In *RGB* color space, *(R, G, B)* components represent not only color information but also luminance which may vary across a face due to the ambient lighting. This makes *(R, G, B)* com-ponents an unreliable measure for separating skin from non-skin regions. *YCbCr* color space separates luminance from the color information and hence, provides a way to use only color information for segmenting skin and non-skin regions.

The distribution of skin colors of different people is found to be clustered in a small area in the *YCbCr* color space. Although skin colors of different people may vary over a wide range, they differ more in brightness than its color. Due to this fact, skin color model is developed in *YCbCr* color space and only chrominance components (*Cb* and *Cr*) are used for modeling the skin pixels. Since color histogram of skin color distribution of different people is clustered at one place in *Cb*, *Cr* plane, it can be represented by a Gaussian model $N(\mu, \sum)$ with mean μ and covariance $\sum$.

With the Gaussian fitted skin color model, likelihood of skin for each pixel can be computed. If a pixel, having transformed from *RGB* color space to *YCbCr*, has a chromatic color vector $x = (Cb, Cr)^T$, the likelihood $P(x)$ of skin for this pixel can then be obtained by

$$P(x) = \frac{1}{\sqrt{2\pi|\sum|}} exp\left[-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)^T\right] \quad (1)$$

Likelihood values obtained in Eq. 2.1 can be used to segment skin and non-skin regions. An adaptive thresholding process [23] is applied on likelihood image (obtained using skin likelihood values for all pixels) to compute an optimal threshold. Skin segmentation is obtained by thresholding the skin likelihood image using this threshold.

III. Ear Detection Technique

This section presents an efficient ear detection technique. This technique is based on the fact that in a profile face image, ear is the only part which contains large variation in the pixel intensities, resulting this part rich in edges. This can be visualized from the image shown in Fig. 2.3f which displays the edge image of the skin segmented image of Fig. 2.3e. It can be observed that the ear part has larger edge density as compared to other parts. Further, it can also be noticed that all edges belonging to the ear part contain some curvature. These characteristics are exploited for ear localization in the presented technique which computes edge clusters in the edge map obtained from the profile face image and examines them for ear localization. Flow diagram of the technique is presented in Fig. 2.4.

IV. *Preprocessing*

After Profile face image undergoes a preprocessing phase before ear localization. This invoves skin segmentation where skin areas of the image are segmented. Further, edge computation is carried out on skin segmented image. In the next step, obtained edges are approximated using line segments and subsequently used in the construc-tion of convex edge map. Erroneous edges are pruned out in the last step.

I. *Skin Region Detection*

Since ear exist in skin region, non-skin regions of the profile face should be segmented and removed from further processing. The skin color model discussed in is used for skin segmentation. It transforms a color image into a gray scale image (called skin-likelihood image) using Eq. 1 such that the gray value at each pixel shows the likelihood of the pixel belonging to the skin. With an appropriate thresholding, the gray scale image is further transformed to a binary image segmenting skin (white pixels) and non-skin (black-pixels) regions. Since people with different skins have different likelihood, an adaptive thresholding process [23] is used to achieve the optimal threshold for each image.

The binary image showing skin and non-skin regions may contain some holes in it due to the presence of noise in the profile face image. Dilation is applied to fill these holes before using it for skin segmentation. The effect of this operation is to enlarge gradually the boundaries of regions of foreground pixels (i.e. white pixels). Thus the area of foreground pixels grows while filling holes within regions. Figure 3 considers an example of skin region detection with various intermediate steps. For a color image given in Fig. 3a, corresponding skin-likelihood image is shown in Fig. 3b. or posed ear may lead to false detections. The technique discussed in [17] achieves good detection rate, but the size of the test data set is very small (only 203 images). Also, if the test ear images are rotated or their appearances are changed with respect to training data, the presented technique may fail because the training images may not include such cases. Forming a database of ears with all possible rotation demands very large space and practically not feasible. Also to detect the ears of different scale, the technique should perform an exhaustive search with filters of various sizes which is computationally very expensive and makes the technique infeasible for real appli-cations. On the other hand, the technique discussed in this chapter can inherently handle rotation (pose) and scale changes and does not incur any extra computational overhead to achieve this. Also, it is tested on a very large data set of 4916 images comprising of rotated (in-plane and out-of-plane) and scaled images which dictates the stability and robustness of the technique. A detailed comparison is found to be less as compared to IITK database due to following reason. Hair color of many subjects in UND database is similar to their skin color. Since strength of the discussed technique is derived from the successful detection of skin regions, similarity of the hair color with skin reduces the performance of skin segmentation and in turn, affects the ear localization accuracy and increases false positives.

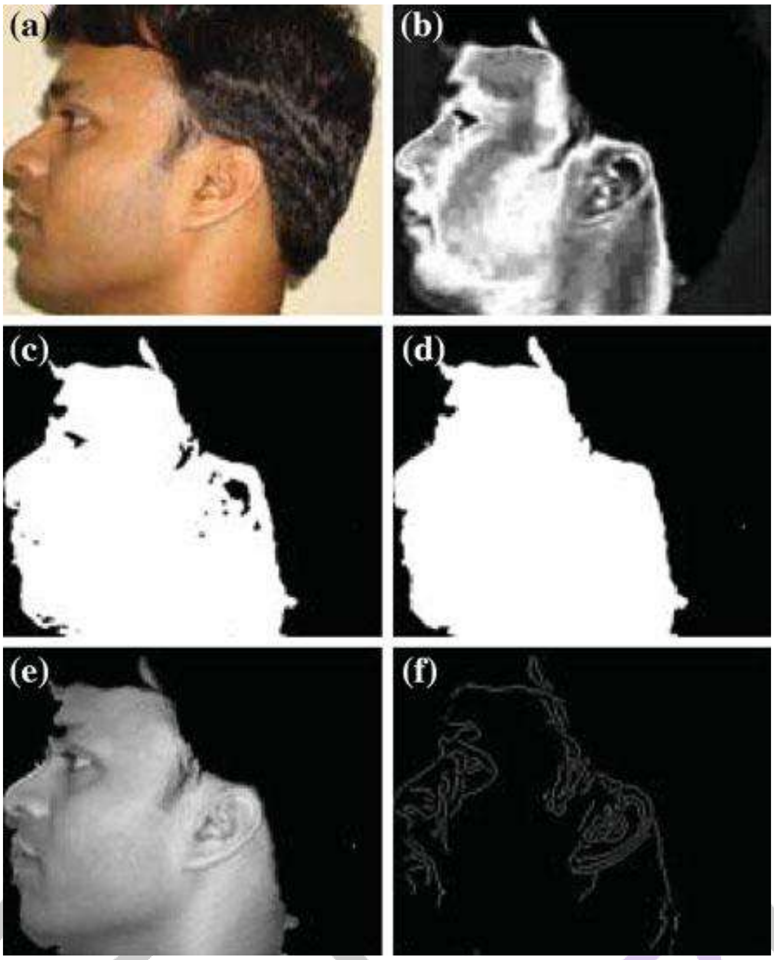

Figure 2: (a) Original Image (b) binary image, ***(c)*** *dilate*

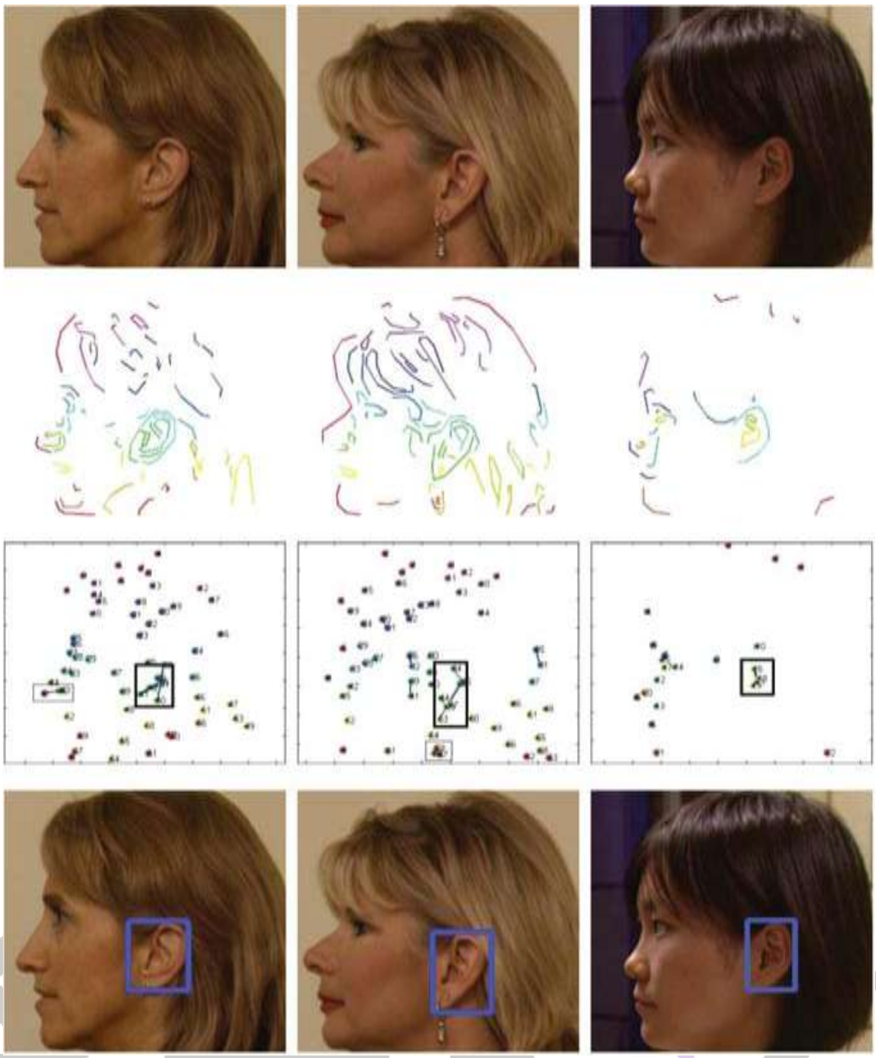

Figure 2 *Ear detection: (row-1) original input images, (row-2) edge maps approximated with lines (colors used to distinguish edges), (row-3) edge connectivity graphs (graph components having average vertex degree >1 enclos*

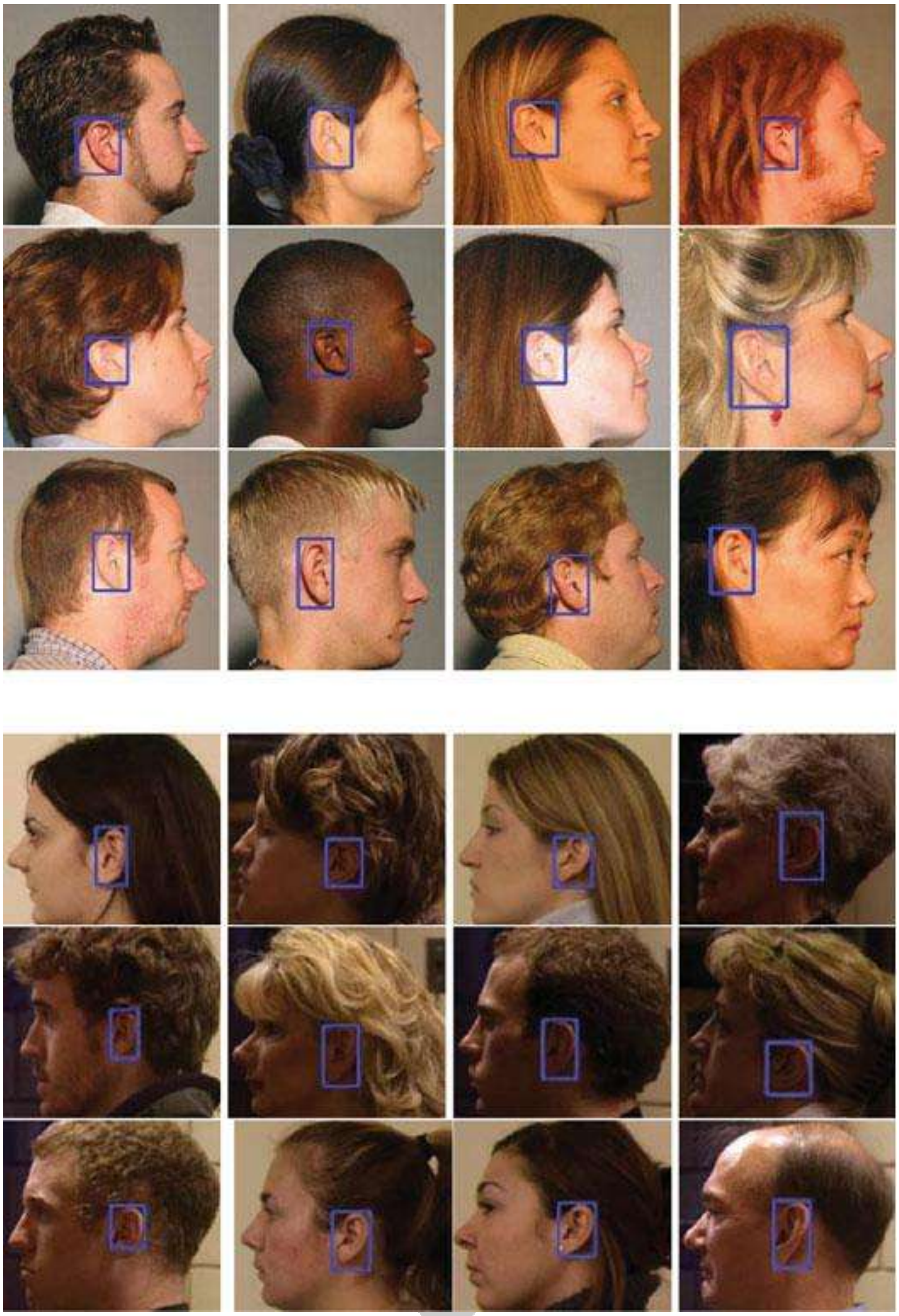

Figure 3 *Ear detection results for UND database.* ***a*** *UND-E data set,* ***b*** *UND-J2 data set*

Table below also shows comparative performance of some well known techniques on UND database. It is seen from the table that [9] produces low detection rate as compared to the technique discussed in this chapter. Moreover, it makes the assump-tion that the ear is the principal elliptical shape in the image which limits its use to the controlled environment and frontal ears, as the presence of background objects

Parameters	Techniques	
	[17]	Presented technique
Time per detection (same	26.40 s	7.95 s
configuration)		
Training overhead	More. To train classifiers with	Very less. Only required to
	1000s of positive and negative	learn skin parameters and ear
	samples	template using few 100
		samples
Invariant to		
(i) Rotation	No	Yes
(ii) Scale	No	Yes
(iii) Occlusion	Up to some extent	No
Total test data size	Very small (307 images)	Large (4916 images)
Test data	No scaling, minor pose	Good amount of scaling and
	variation	rotation (IITK data sets 2
		and 3)

Performance of the presented technique could not be compared with [4] because of the non-availability of the test results. Also comparisons could not be made with [19, 20] as these techniques have used XM2VTS database [10] which is not available. However, it can be noted that XM2VTS database is relatively easy to work because it contains images captured in plane background with controlled illumination and comprises of good quality images whereas UND images contain non-uniform cluttered background, poor illumination and pose variations.

II. Conclusion

The presented technique has been successful to detects ears fully or partially in some cases of IITK and UND databases. Failure has occurred when ears are occluded by hair or affected by noise and poor illumination. Few examples of failure in detecting ears due to these reasons are shown in Fig. 4.

References

[1] Burge, Mark and Burger, Wilhelm 2000. Ear biometrics in computer vision. In *Proceedings of International Conference on Pattern Recognition (ICPR'00)*, vol. 2, 822–826.

[2] Hurley, David J., Mark S. Nixon, and John N. Carter. 2005. Force field feature extraction for ear biometrics. *Computer Vision and Image Understanding* 98(3): 491–512.

[3] Yan, Ping, Kelvin W. Bowyer. 2005. Empirical evaluation of advanced ear biometrics. In *Proceedings of International Conference on Computer Vision and Pattern Recognition-Workshop*, vol. 3, 41–48.

[4] Alvarez, L., E. Gonzalez and L. Mazorra. 2005. Fitting ear contour using an ovoid model. In *Proceedings of IEEE International Carnahan Conference on Security Technology (ICCST'05)*, 145–148.

[5] Kass, M., A. Witkin, and D. Terzopoulos. 1988. Snakes: Active contour models. *International Journal of Computer Vision* 1(4): 321–331.

[6] Yan, Ping, and K.W. Bowyer. 2007. Biometric recognition using 3D ear shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(8): 1297–1308.

[7] Ansari, Saeeduddin and Phalguni Gupta. 2007. Localization of ear using outer helix curve of the ear. In *Proceedings of the International Conference on Computing: Theory and Applications (ICCTA'07)*, 688–692.

[8] Yuan, Li, Zhi-Chun Mu. 2007. Ear detection based on skin-color and contour information. In *Proceedings of International Conference on Machine Learning and Cybernetics (ICMLC'07)*, vol. 4, 2213–2217.

[9] Arbab-Zavar, Banafshe and Mark S. Nixon. 2007. On shape-mediated enrolment in ear biomet-rics. In *Proceedings of the 3rd International Conference on Advances in Visual Computing— Volume Part II,*

[10] Messer, K., J. Matas, J. Kittler, J. Lttin and G. Maitre. 1999. XM2VTSDB: The extended M2VTS database. In *Proceedings of 2nd International Conference on Audio and Video-based Biometric Person Authentication*, 72–77.

[11] Sana, Anupam, Phalguni Gupta, Ruma Purkait. 2007. Ear biometric: A new approach. In *Proceedings of International Conference on Advances in Pattern Recognition (ICAPR'07)*, 46–50.

[12] Prakash, Surya, Umarani Jayaraman and Phalguni Gupta. 2008. Ear localization from side face images using distance transform and template matching. In *Proceedings of IEEE International Workshop on Image Proceedings Theory, Tools and Applications (IPTA'08)*, 1–8.

[13] Prakash, Surya, Umarani Jayaraman and Phalguni Gupta. 2009. A skin-color and template based technique for automatic ear detection. In *Proceedings of International Conference on Advances in Pattern Recognition (ICAPR'09)*, 213–216.

[14] Attarchi, S., K. Faez and A. Rafiei. 2008. A new segmentation approach for ear recognition. In *Proceedings of International Conference on Advanced Concepts for Intelligent Vision Systems*, 1030–1037. USTB Database, University of Science and Technology Beijing. http://www.ustb.edu.cn/resb/.

[15] Carreira-Perpinan. 1995. Compression neural networks for feature extraction: Application to human recognition from ear images. Master's thesis, Faculty of Informatics, Technical Uni-versity of Madrid, Spain.

[16] Islam, S.M.S., M. Bennamoun and R. Davies. 2008. Fast and fully automatic ear detection using cascaded adaboost. In *Proceedings of IEEE Workshop on Applications of Computer Vision (WACV'08)*, 1–6.

[17] Prakash, Surya, Umarani Jayaraman, Phalguni Gupta. 2009. Ear localization using hierarchical clustering. In *Proceedings of SPIE International Defence Security and Sensing Conference, Biometric Technology for Human Identification VI, 730620*, vol. 7306, 730620–730620-9. Cummings, A., M. Nixon and J. Carter. 2010. A novel ray analogy for enrolment of ear bio-metrics. In *Proceedings of International Conference on Biometrics: Theory, Applications and Systems (BTAS'10)*, 1–6.

[18] Ibrahim, Mina I.S., Mark S. Nixon and Sasan Mahmoodi. 2010. Shaped wavelets for curvilinear structures for ear biometrics. In *Proceedings of 6th International Conference on Advances in Visual Computing (ISVC'10)—Part I*, 499–508.

[19] Kumar, Amioy, Madasu Hanmandlu, Mohit Kuldeep and H.M. Gupta. 2011. Automatic ear detection for online biometric applications. In *Proceedings of National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, NCVPRIPG 2011*, 146–149.

[20] Lankton, S., and A. Tannenbaum. 2008. Localizing region-based active contours. *IEEE Trans-actions on Image Processing* 17(11): 2029–2039.

[21] Cai, J., and A. Goshtasby. 1999. Detecting human faces in color images. *Image and Vision Computing* 18(1): 63–75.

[22] Poynton, C.A. 1996. *A Technical Introduction to Digital Video*. New York: Wiley.

[23] Garcia, C., and G. Tziritas. 1999. Face detection using quantized skin color regions merging and wavelet packet analysis. *IEEE Transactions on Multimedia* 1(3): 264–277.

[24] Bay, Herbert, Tinne Tuytelaars and Luc Van Gool. 2006. SURF: Speeded up robust features. In *Proceedings of 9th European Conference on Computer Vision (ECCV'06)*, 404–417.

[25] Bay, Herbert, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. 2008. Speeded-up robust features (SURF). *Computer Vision and Image Understanding* 110(3): 346–359.

[26] Lowe, David G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2): 91–110.

[27] Mikolajczyk, Krystian, and Cordelia Schmid. 2005. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(10): 1615–1630.

[28] Hopcroft, John, and Robert Tarjan. 1973. Algorithm 447: Efficient algorithms for graph manipulation. *Communications of the ACM* 16(6): 372–378.

[29] Bustard, J.D., and M.S. Nixon 2008. Robust 2D ear registration and recognition based on SIFT point matching. In *Proceedings of International Conference on Biometrics: Theory, Applications and Systems (BTAS'08)*, 1–6. University of Notre Dame Profile Face Database, Collections E and J2. http://www.nd.edu/ cvrl/CVRL/DataSets.html.

[30] Prakash, Surya, Umarani Jayaraman, Phalguni Gupta. 2009. Connected component based technique for automatic ear detection. In: *Proceedings of 16th IEEE International Conference on Image Processing (ICIP'09)*, 2741–2744