Ear Localization using Hierarchical Clustering

Surya Prakash, Umarani Jayaraman and Phalguni Gupta
Department of Computer Science and Engineering,
Indian Institute of Technology Kanpur,
Kanpur-208016, India.

ABSTRACT

Ear biometrics has been found to be a good and reliable technique for human recognition. With the initial doubts on uniqueness of the ear, ear biometrics could not attract much attention. But after it has been said that it is almost impossible to find two ears with all the parts identical, ear biometrics has gained its pace. To automate the ear based recognition process, ear in the image is required to be localized automatically. This paper presents a technique for the same. Ear localization in the proposed technique is carried out by using the hierarchical clustering of the edges obtained from the side face image. The technique is tested on a database consisting of 500 side face images of human faces collected at IIT Kanpur. It is found to be giving 94.6% accuracy.

Keywords: Biometrics, skin-segmentation, hierarchical clustering, ear localization.

1. INTRODUCTION

Since biometric systems uses human traits for recognition which can not be forgotten and stolen, they have become very essential components in almost all security aspects. These systems perform the recognition of a human being based on physiological and behavioral characteristics. Physiological characteristics are related to the shape of the body. Biometric traits such as face, fingerprint, iris, ear, hand geometry fall under this category. Behavioral characteristics are related to behavior of a person. Signature, voice, character strokes etc. are some of the biometric traits which fall under this category. Among the various physiological traits, ear has gained much attention in recent years as it has been found to be a good and reliable biometrics for human verification and identification.[1] Reason behind the ear biometrics gaining popularity is that ears are remarkably consistent. Unlike faces, they do not change shape with different expressions or age, and remain fixed in the middle of the side of the head against a predictable background. To automate ear based recognition process, automatic ear localization is necessary but at the same time detection of the ear from an arbitrary side face image is a challenging problem. This is because of the fact that ear image can vary in appearance under different viewing and illumination conditions. In literature, there are very few techniques available for automatic ear detection.[2–7] These techniques are either semi-automatic and need some user intervention or fail to detect ears in many situations. Hurley et al.[3] have used force field approach to get the ear location. Their approach is only applicable when a small background is present. Burge and Burger[2] have used deformable contours for ear detection. Chen and Bhanu[4] have presented a template based method for human ear detection from side face range images. The model template in this method is represented by an averaged histogram of shape index. This method works for 3D ear biometrics. Alvarez et al.[5] have proposed an ear localization method from 2D face image using ovoid and active contour model. Ear boundary is estimated by fitting the contour of an ear in an image by combining snake technique and an ovoid model. This method requires an initial approximate ear contour as input and cannot be used in fully automated ear recognition process.

Yuan and Mu[7] have proposed an ear detection method based on skin-color and contour information. This method detects the ear by first roughly estimating the ear location and then improving the localization using contour information. This technique assumes the ear shape elliptical and fits ellipse to the edges to get the accurate position of the ear. The assumption of considering shape of the ear elliptical may not be true for all the persons and not help is detecting the ear in general. For example, boundary of an ear of triangular shape cannot be detected in this approach (Figure 1(a)). Yan and Bowyer[8] has proposed skin segmentation and active contours based method for ear localization in 3D ear biometrics. Ear

Further author information: (Send correspondence to Surya Prakash)
Surya Prakash: E-mail: psurya@cse.iitk.ac.in,
Umarani Jayaraman: E-mail: umarani@cse.iitk.ac.in,
Phalguni Gupta: E-mail: pg@cse.iitk.ac.in

Proceedings of SPIE International Defence, Security and Sensing conference (Biometric Technology for Human Identification VI), Orlando, Florida, April 2009

(a) (b) (c) (d)

Figure 1. Ear shapes: (a) Triangular, (b) Round, (c) Oval, (d) Rectangular

localization in their approach is carried out by first detecting the ear pit and then further using its boundary as initialization for the active contour. Ear pit detection uses the location of nose tip which is detected by a heuristics. This heuristic may fail for the faces rotated around horizontal axis (specially persons looking down). Ansari and Gupta [6] have presented an ear detection approach based on edges of outer ear helices. Their method exploits the parallelism between the outer helix curves of the ear to localize the ear. This technique solely relies on the parallelism between the outer helix curves and does not use any structural information present in the inner part of the ear, and hence fails if the helix edges are not proper. Moreover, finding parallel edges in an image is a computation intensive process. In, [9] Sana et al. have proposed an ear detection scheme based on template matching. To detect ears of different size, they maintain templates of various sizes. Templates in their approach are created manually by cropping the ears of different size from the side face images and template matching is performed for the wavelet decomposed template and side face images. In real scenario, ear occurs in various sizes and the preestimated templates are not sufficient to handle all the situations. Further, detection of ear using templates of various sizes and then selecting best detection is a computationally expensive task.

This paper presents a novel technique for automatic ear detection from side face images. The proposed technique is based on hierarchical clustering[10] and detects ear from the side face images efficiently using the edge information. To reduce the search space for the ear, skin segmentation of the image is performed and only skin regions are used for the edge computation. Obtained edges are clustered using the hierarchical clustering and obtained clusters are used for ear localization.

Rest of the paper is organized as follows. Section 2 briefly presents a skin-color model and hierarchical clustering which serve as the foundation for further discussion. In Section 3, the proposed ear detection technique is discussed. Experimental results are discussed in Section 4 and conclusion is given in the last Section.

2. PRELIMINARIES

This section discusses some of the basic techniques which are required in developing the proposed ear localization model. Section 2.1 describes a skin-color model which is used for skin segmentation. Section 2.2 discusses hierarchical clustering which is used for ear localization in the proposed technique.

2.1 Color Based Skin Segmentation

A color based skin segmentation technique[11,12] is discussed in this section, which is used for skin segmentation in the proposed ear localization approach. This technique takes a color image as input and segments it into skin and non-skin regions. The technique is adaptable to different skin colors and lighting conditions. It performs skin segmentation in chromatic color space as it is more suitable for the characterizing skin-color. First it converts the *RGB* color space to chromatic color space,[13] and then uses the chromatic color information for further processing. In *RGB* color space, the triple components *(R, G, B)* represent not only color but also luminance which may vary across a person's face due to the ambient lighting and is not a reliable measure in separating skin from non-skin regions. [11] Luminance can be removed from the color representation in the chromatic color space, which is defined by a normalization process as follows:

$$r = \frac{R}{(R+G+B)}, b = \frac{B}{(R+G+B)} \tag{1}$$

It can be observed from Eqn. 1 that green color is redundant after the normalization as $r + g + b = 1$. The color distribution of skin colors of different people is found to be clustered in a small area of the chromatic color space. Although skin colors of different people vary over a wide range, they differ much in brightness than color. Because of this fact, skin-color model is developed in chromatic color space. Since color histogram of skin-color distribution of different people is clustered at one place in the chromatic color space, it can be represented by a Gaussian model $N(\mu, C)$, where mean μ and covariance C can be defined as:

$$\mu = E[x] \quad (2)$$
$$C = E[(x - \mu)(x - \mu)T] \quad (3)$$

where $x = (r, b)^T$. $E[\phi]$ denotes the expectation of the predicate ϕ. With this Gaussian fitted skin-color model, likelihood of skin for any pixel of an image can be obtained. If a pixel, having transformed from RGB color space to chromatic color space, has a chromatic pair value of (r, b), the likelihood $P(r, b)$ of skin for this pixel can then be computed as follows:

$$P(r, b) = \frac{1}{\sqrt{2\pi|C|}} \exp[-\frac{1}{2}(x - \mu)C^{-1}(x - \mu)^T] \quad (4)$$

Likelihood values obtained using Eqn. 4 are used for the segmentation of skin and non-skin regions.

2.2 Hierarchical Clustering

Hierarchical clustering is a way to investigate grouping in the data, simultaneously over a variety of scales, by creating a hierarchical cluster tree. The tree is not a single set of clusters, but rather a multilevel hierarchy, where clusters at one level are joined as clusters at the next higher level. This gives us flexibility to decide what level or scale of clustering is most appropriate in our application. To perform hierarchical clustering on a data set, following steps are followed:

2.2.1 Similarity Estimation

This step finds the similarity (or dissimilarity) values between every pair of objects in the given data set. These values are obtained by calculating the distance between the objects by using a distance metric.

2.2.2 Forming Hierarchical Cluster Tree

In this step, objects are grouped into a binary hierarchical cluster tree. The pair of objects that are in close proximity to each other are linked together. Proximity is determined using the distance values obtained in previous section. As objects are paired into binary clusters, the newly formed clusters are grouped into larger clusters until a hierarchical tree is formed.

2.2.3 Cutting Hierarchical Tree into Clusters

This is very crucial step. In this step, a cut point based on distance threshold is determined to cut the hierarchical tree into clusters. This step prunes the branches off the bottom of the hierarchical tree, and assign all the objects below each cut to a single cluster. This creates partition in the data.

3. PROPOSED TECHNIQUE

In this section, proposed ear localization technique is discussed in detail. First preprocessing step is discussed followed by ear localization and ear verification. In the preprocessing step, skin areas of the input side face image are segmented and processed for edge computation. Further, spurious edges are eliminated from the obtained edge map using edge length and curvature based criterion. Ear localization step computes the edge clusters using hierarchical clustering and uses them in ear localization. Localized ear is verified using normalized cross correlation technique. Figure 2 shows complete flow chart of the proposed technique.

3.1 Preprocessing

Preprocessing involves skin region segmentation, edge detection, edge approximation by line segments and non-ear edge pruning.

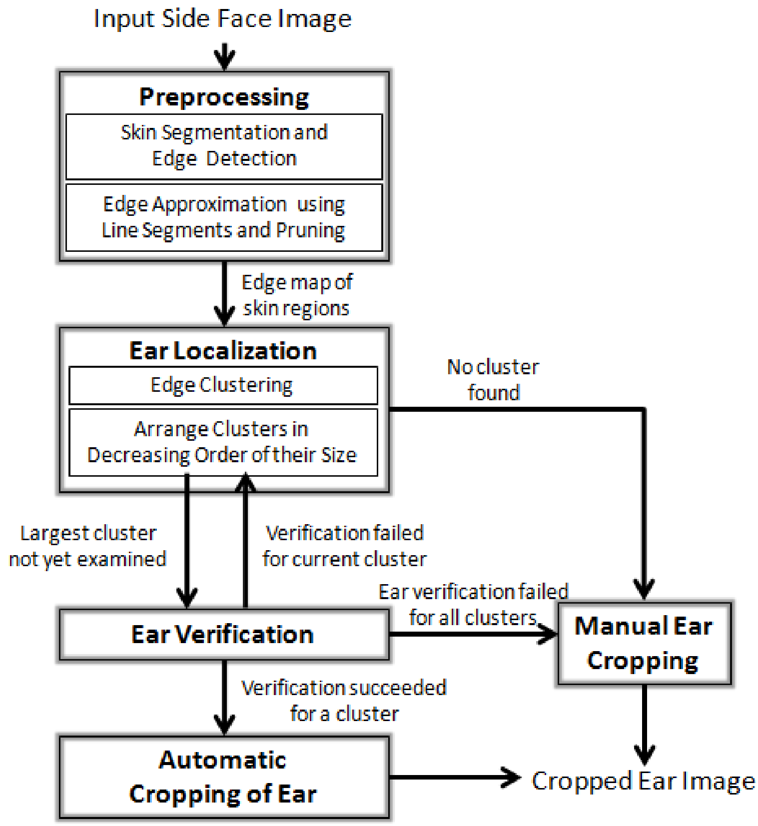

Figure 2. Flow chart of the proposed technique

3.1.1 Skin Region Detection

The first step of the proposed technique is skin region detection in the image. Main objective of this step is to segments the skin regions to reduce the search space for the ear. Naturally, ears are in skin region and there is no point looking for them in non-skin regions. For skin region detection, skin color model presented in Section 2.1 is used. This model first transforms the color image into a gray scale image (called skin-likelihood image) using Eqn. 4. Gray value at each pixel in the skin-likelihood image shows the likelihood of the pixel belonging to the skin. Using an appropriate thresholding, the grayscale image is further transformed to a binary image showing skin and non-skin regions. Since people with different skins have different likelihood, an adaptive thresholding [12] process is used to achieve the optimal threshold value for each run.

Repairing Binary Image: Usually binary image obtained after thresholding contains holes in it because of the presence of some noise in the image. These holes are required to be filled before using the image for segmentation purpose. In the proposed technique, hole filling in the binary image is done using dilation, which is a basic morphological operation. The effect of this operator on a binary image is to gradually enlarge the boundaries of regions of foreground pixels (i.e. white pixels, typically). Thus areas of foreground pixels grow in size while holes within those regions become smaller.

Figure 3 shows an example of skin region detection process with various intermediate steps involved in it. Figure 3(a) shows a color image where skin region detection need to be performed. Figure 3(b) shows skin likelihood image obtained for it. All skin regions in Figure 3(b) are shown brighter than the non-skin region. Figure 3(c) shows the binary image after applying threshold on skin likelihood image. We can observe that this image contains some holes in it. Dilation is applied on this image to fill these small holes. Figure 3(d) shows the repaired binary image after hole filling. This image is used for skin region detection. Figure 3(e) shows final skin segmented image (in grayscale) for the image shown in Figure 3(a). It can be observed from the result of skin segmentation that not all detected skin regions contain ear. Hence, ear localization

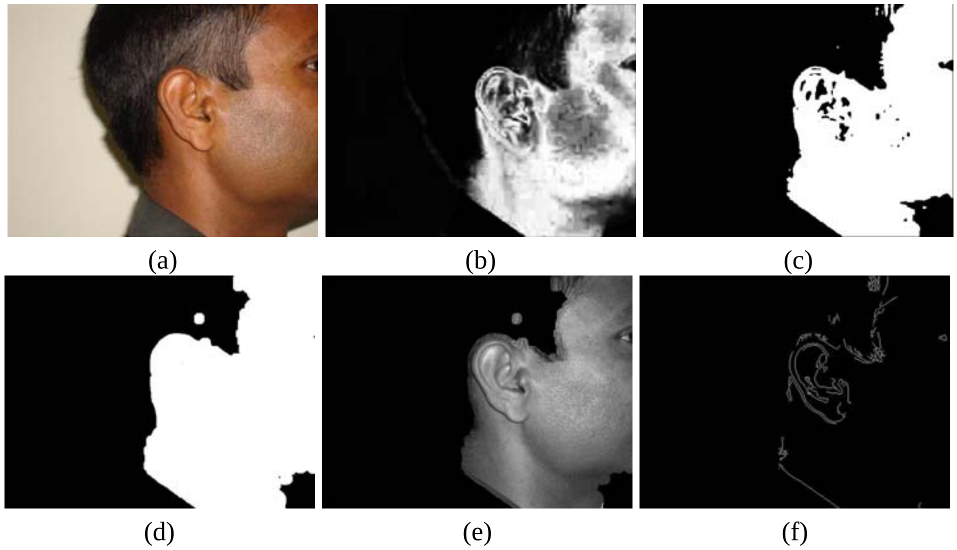

Figure 3. Skin segmentation: (a) Input image, (b) Skin-likelihood values, (c) Binary image (d) Binary image after dilation, (e) Skin-segmented image, (f) Edge image

can be performed to locate the ear in all these skin-like segments.

3.1.2 Edge Computation of Skin Regions

After detecting the skin regions, edges are computed for these regions and used for further processing. Canny edge operator[14] is employed for this purpose. After edge detection, a list of all edges present in the image is obtained. This is done by connecting points together into a list of coordinate pairs. Wherever an edge junction is found, the list is terminated and a separate list is generated for each of the branches. This edge list includes many spurious edges of short length which arise due to the noise in the image. These edges can be easily pruned out using a length based criterion. If set X contains all the edges present in the list, new edge set X_l can be obtained using edge length based pruning as follows:

$$X_l = \{e \mid e \in X \text{ and length(e)} > T_l\} \quad (5)$$

where length(e) returns the length of edge e and T_l is the threshold for edge length. Length based edge pruning removes all the edges which are of shorter length and arose due to the presence of noise such as hair on the side face image.

3.1.3 Approximating Edges using Line Segments

Approximation of edges using line segments is an important step in the proposed ear localization technique. This step not only helps in simplifying the edge representation, but also assists in removing the non-ear edges from the the edge map of side face image. As we can easily see, all of the pixels present in an edge obtained in previous step may not be equally important and may not be necessary to represent the edge. So to remove the redundant pixels and to get a compact representation of the edge, line segments are fitted to it. This eliminates all the pixels which are not necessary to represent the edge and preserves only the pixels which are important. This breaks the edges present in the image into line segments. Line segment fitting proceeds as follows. Procedure takes each array of edge points in edge list and finds the size and position of the maximum deviation from the line that joins the endpoints. If the maximum deviation exceeds the allowable tolerance, the edge is shortened to the point of maximum deviation and the test is repeated. In this manner each edge is broken down to line segments, each of which adhere to the original data with the specified tolerance.

3.1.4 Curvature based Edge Pruning

Approximation of the edges using line segments results in a new set of edges where each edge is represented using very less number of points. These points define the end points of the line segments used in the approximation of an edge. It is easy to observe that the edges which are linear (or close to linear) will contain only two points for their representation after

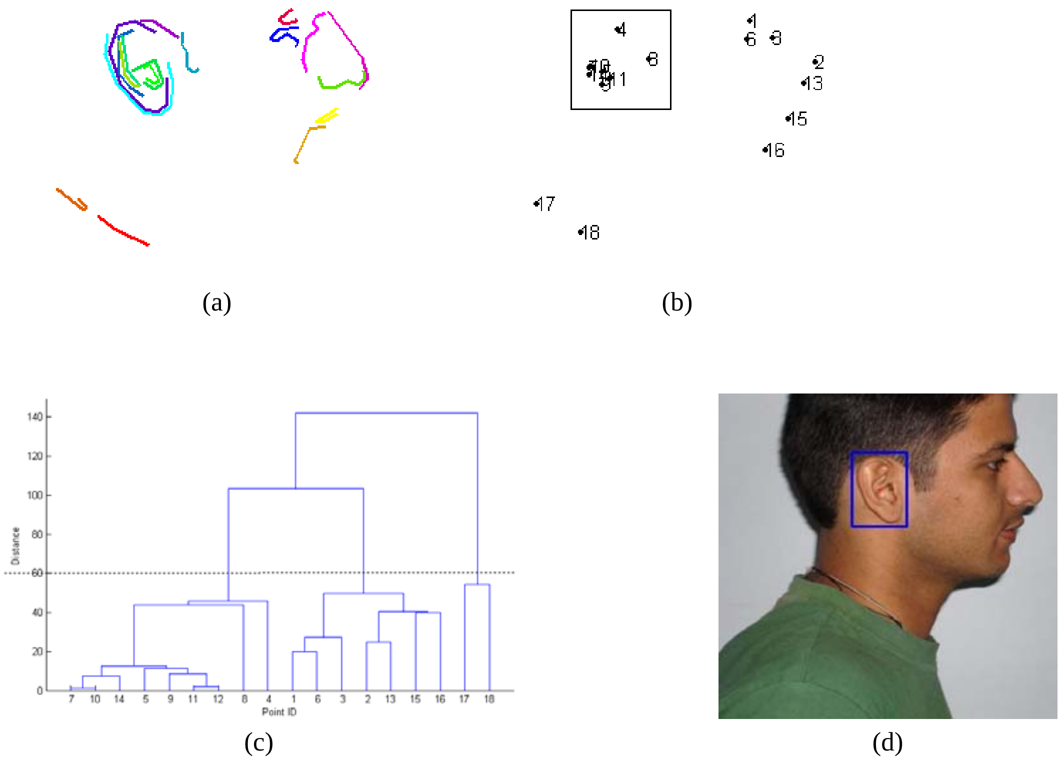

Figure 4. (a) Edge map (colors are used to distinguish the edges), (b) Center of mass points for edges, (c) Cluster tree for the edge points shown in (a), (d) Ear localization result

line segment fitting. Since the edges belonging to the ear contain some curvature, they require more than two points for their representation using line segments. In other words, we can say that all the edges having two points cannot be a part of the ear and hence can be removed from the edge list. This results a new edge set X_c containing all the edges which may belong to the ear. Set X_c is formally defined as follows:

$$X_c = \{e \mid e \in X_l \text{ and } \gamma(e) > 2\} \tag{6}$$

where $\gamma(e)$ is a function which returns the number of points used to approximate edge e using line segments. Clearly, set X_c contains all the edges which are represented using two or more line segments. Curvature based edge pruning removes all the edges which may not be ear edges.

3.2 Ear Localization

Ear localization technique presented in this section is motivated from the fact that in the side face image, ear is the place where much variation in the intensity happens resulting this place rich in edges. This makes the density of the edges more at the ear part compare to other locations. So if we cluster the edges, normally we will find that the largest edge cluster represents the ear. Since number of clusters present in the edge map are not know a priori, we employ hierarchical clustering to for grouping the edges into clusters. Obtained edge clusters are analyzed subsequently for ear localization purpose.

3.3 Clustering of Edges

The set X_c can be used to define the edge map of the side face image. Let there be n edges in X_c. Define the i^{th} edge e_i in X_c by a point p_i. Thus X_c can be represented by a set P of points $p_1, p_2, .., p_n$. Actual position of point p_i on the

2D plane is given by the center of mass of the edge e_i. These points are clustered using hierarchical clustering technique which gives the multi-scale hierarchical representation of the data.

Figure 4(a) shows an example of edge map obtained for a side face image shown in Figure 4(d). This edge map is obtained after applying edge length and curvature based edge pruning. These edges are represented by their center of mass points, as shown in Figure 4(b). These points are taken as a representative of the edges and used for clustering. Figure 4(c) shows a hierarchical cluster tree for obtained using these points. This tree is used for computing the clusters in the data. A cut in the cluster tree is carried out using a distance threshold which prunes branches off the bottom of the hierarchical tree, and assigns all the points bellow each cut to a single cluster. In Figure 4(c), a distance threshold used to cut the tree into cluster is shown by a horizontal dotted line.

3.4 Localization of Ear using Cluster Information

Hierarchical clustering gives a multi-scale representation of the points. This representation is broken down to clusters using a distance threshold which can be learnt from the data we are handling. In the proposed technique, the points obtained for the edges are clustered using the hierarchical clustering and obtained clusters are analyzed for ear localization. These clusters are arranged in decreasing order of their size where cluster size is estimated by counting the number of points present in it. Cluster of the largest size is considered as the ear. This is motivated from the fact that in side face image, ear location contains the largest number of edges. Bounding box of the edges corresponding to the points used in the cluster is considered as the ear location. This claim is verified using the normalized cross-correlation based criterion. If the ear verification fails, next largest component is considered as the ear and claim is verified again. This continues till ear is detected or all the components are examined. Success of the ear localization depends on breaking of the hierarchical cluster tree into appropriate size of clusters. By estimating the appropriate threshold for the distance value, we can localize the ears of various size (scale). Figure 4(d) shows an example of the localized ear using the proposed approach. Cluster of points which results the ear location is shown inside a rectangle in Figure 4(b).

3.5 Ear Verification

To decide whether a localized ear is actually an ear or not, normalized correlation based method is used. For the verification of the localization, normalized cross-correlation coefficient (NCC)[15,16] between the localized ear and an off-line created ear template is computed. NCC is defined as follows:

$$NCC = \frac{\sum_{u,v}[I(u,v) - \bar{I}][T(u - x_c, v - y_c) - \bar{T}]}{\sqrt{\sum_{u,v}[I(u,v) - \bar{I}]^2 \sum_{u,v}[T(u - x_c, v - y_c) - \bar{T}]^2}}$$

where I is grayscale image and the sum is performed over u, v running for the image I. (x_c, y_c) is the center point of the template image. $\bar{I}$ and $\bar{T}$ are the average of brightness values of the localized ear and template image respectively. Normalized cross-correlation coefficient estimates the degree of linear dependence between the corresponding pixel brightness values being compared. Since the cross-correlation coefficients lie between -1.0 and 1.0, match ratings also lie between -1.0 and 1.0. When the match rating is typically above a preestimated threshold we consider the localization as genuine. Otherwise, we reject the localization. Value of normalized cross-correlation coefficient closer to 1 indicates a better match.

We prefer normalized cross-correlation coefficient over cross-correlation in ear verification as normalized cross-correlation is more suited for the image-processing applications in which the brightness of the image and template can vary due to lighting and exposure conditions.

4. EXPERIMENTAL RESULTS

Database consisting of 500 side images of human face with resolution 640×480, collected at IIT Kanpur is used for experimentation. These images are captured using a digital camera from a distance of 0.5 to 1 meter. To create an ear template in our experiment, a set of side face images of 20 people is considered. These images include side face images of both men and women. Ears are cropped from these images and resized to 120×80 pixels. Ear images of all shapes viz. triangular, round, oval and rectangular are considered for template creation. Figure 1 shows various shapes of the human ear. Ear template is created by averaging the intensity values of the pixels.

Figure 5. Ear detection results where detected ears are marked with rectangular boundary.

Figure 5 shows some of the results obtained using proposed technique. Ears of different shape are efficiently localized from the side face images. Usually, localizing the ear in case of ear images of women is a challenging problem. This is because of the fact that in case of women, ear images are very much surrounded by hair and create more confusion for the algorithm by adding the false edges around the ear. The proposed technique performs the detection for these images also satisfactorily. It is also able to detect ear in presence of little occlusion due to hair. First three images in the last row of Figure 5 show such examples. The accuracy of the localization is defined by:

$$\text{Accuracy} = \frac{\text{Number of Genuine Localization} \times 100}{\text{Total Samples}} \% \tag{7}$$

The accuracy of the proposed technique is found to be 94.6%. A localization is considered as correct only when complete ear is enclosed inside the rectangular boundary with at most 10% non-ear pixels. Localization method has failed in some cases to localize ear correctly, especially for the images which are of poor quality or heavily occluded by the hair.

5. CONCLUSION

This paper first presents a brief review of the existing ear localization techniques in the literature for automatic ear detection in 2D side face images. It then proposes a novel technique for the same. The proposed technique localizes ear automatically in the side image of human face without any user interaction and is appropriate to be employed in an automatic ear based biometric system. It effectively detects ears of different shape without any user intervention. Ear detection in the proposed technique is fast as it prunes out almost 60% area of the side face image and searches for ear only in the skin regions. Proposed technique follows three steps viz. preprocessing, ear localization and ear verification. Preprocessing step segments skin and non-skin regions in the image and computes the edge map of the skin regions. It employs edge length and curvature based criterion to eliminate spurious edges from the edge map. Ear localization phase uses skin regions edge map for ear localization by clustering them using hierarchical clustering. Once ear is localized, a cross correlation based approach is used to verify the ear localization. The proposed technique is tested on a database containing side face images of 500 individuals collected at IIT Kanpur and found to be giving 94.6% accuracy. The proposed technique can be extended for the detection of ear in noisy images and in the cases where ear is immensely occluded by the hair. It can also be easily extended to multiple ears detection.

ACKNOWLEDGMENTS

Authors would like to thank Pradeep Nayak and Gaurav Ghadoke for helping in collecting the data needed for experiment. This work was partly supported by Ministry of Communication and Information Technology, Government of India.

REFERENCES

[1] Bhanu, B. and Chen, H., [*Human Ear Recognition by Computer*], Springer (2008).

[2] Burge, M. and Burger, W., "Ear biometrics in computer vision," in [*Proceedings of International Conference on Pattern Recognition, ICPR' 00*], **02**, 822–826, IEEE Computer Society (2000).

[3] Hurley, D. J., Nixon, M. S., and Carter, J. N., "Automatic ear recognition by force field transformations," in [*IEE Colloquium: Visual Biometrics*], 7/1–7/5 (2000).

[4] Chen, H. and Bhanu, B., "Human ear detection from side face range images," in [*Proceedings of International Conference on Pattern Recognition, ICPR' 04*], **3**, 574–577, IEEE Computer Society (2004).

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

[6] Ansari, S. and Gupta, P., "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, IEEE Computer Society (2007).

[7] Yuan, L. and Mu, Z.-C., "Ear detection based on skin-color and contour information," in [*International Conference on Machine Learning and Cybernetics, ICMLC' 07*], **4**, 2213–2217 (2007).

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

[9] Sana, A., Gupta, P., and Purkait, R., "Ear biometric: A new approach," in [*Proceedings of International Conference on Advances in Pattern Recognition, ICAPR' 07*], 46–50, World Scientific, Singapore (2007).

[10] Duda, R. O., Hart, P. E., and Stork, D. G., [*Pattern Classification, 2nd Edition*], Wiley (2000).

[11] Cai, J., Goshtasby, A., and Yu, C., "Detecting human faces in color images," in [*International Workshop on Multi-Media Database Management Systems, IWMDMS' 98*], 124–131 (1998).

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

[13] Wyszecki, G. and Stiles, W., [*Color Science: Concepts and Methods, Quantitative Data and Formulas*], Wiley (1982).

[14] Canny, J., "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **8**(6), 679–698 (1986).

[15] Aggarwal, J. K., Davis, L. S., and Martin, W. N., "Correspondence processes in dynamic scene analysis," in [*Proceedings of the IEEE*], **69(5)**, 562–571 (1981).

[16] Barnea, D. I. and Silverman, H. F., "A class of algorithms of fast digital image registration," *IEEE Transactions on Computers* **21**(2), 179–186 (1972).