

An Efficient Ear Localization Technique

Surya Prakash* and Phalguni Gupta

*Department of Computer Science & Engineering,
Indian Institute of Technology Kanpur, Kanpur-208016, India.
{psurya,pg}@cse.iitk.ac.in*

Abstract

This paper proposes an efficient technique for automatic localization of ear from side face images. The technique is rotation, scale and shape invariant and makes use of the connected components in a graph obtained from the edge map of the side face image. It has been evaluated on IIT Kanpur database consisting of 2672 side faces with variable sizes, rotations and shapes and University of Notre Dame database containing 2244 side faces with variable background and poor illumination. Experimental results reveal the efficiency and robustness of the technique.

Keywords: Skin Segmentation, Biometrics, Ear Localization, Ear Recognition, Connected Components.

1. INTRODUCTION

Ear biometrics has received much attention in recent years because of its consistent behaviour over the age. Unlike face, ear does not get changes in its shape with change in expression or age. It remains fixed at the middle of the side face with predictable background whereas face images need to

*Corresponding author, Tel.: +91-512-2597579, Fax: +91-512-2597647

be captured under controlled background to achieve good recognition performance. Moreover, size of the ear is larger than fingerprint, iris, retina etc and smaller than face. It can be acquired easily without the cooperation of the subject. For an efficient ear recognition system, it is required to detect the ear efficiently in an automatic manner from the captured image. Detection of ears from an arbitrary side face image is a challenging problem because ear images may vary in scale and pose (due to in-plane and out-of-plane rotations) under various viewing conditions.

Ear recognition consists of two major steps and they are (i) Ear detection and (ii) Recognition. Most of the well known ear recognition techniques have focussed on recognition on manually cropped ears. There exist a few techniques to detect ear automatically. Most of these techniques can detect the ear only when side face image contains a small background around the ear. But these techniques are not very efficient, particularly when side face images are affected by scaling and rotation (pose variations). Moreover, they are not fully automatic and fast enough to be deployed in realtime applications. However, it is often required, specially in non-intrusive applications, to detect the ear from a whole side face image which is affected due to scale and pose variations.

This paper proposes an efficient ear localization technique which overcomes these issues and is invariant to scale, rotation and shape. The technique makes use of connected components of a graph constructed with the help of edge map of the side 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.

Rest of the paper is organized as follows. Section 2 discusses some well known techniques for automatic ear detection. Section 3 briefly describes a skin color model which has been used for skin segmentation in the proposed technique. Next section presents the proposed technique. Rotation, scale and shape invariance of the proposed technique has been discussed in Section 5. Experimental results are analyzed in Section 6. Conclusions are given in the last section.

2. LITERATURE REVIEW

The first well known technique for detecting ear with the help of deformable contours is due to Burge and Burger [1]. Since contour initialization in this technique needs user interaction, ear localization is not fully automatic. Hurley et al. [2] have used force field technique to get the ear location. It is only applicable when a small background is present in ear image. A template based technique has been proposed by Chen and Bhanu [3]. The technique represents a model template using an averaged histogram of shape index and is applicable to 3D ear biometrics. In [4], Yan and Bowyer have used two-line landmark to detect ear where one line is taken along the border between the ear and the face and the other line from the top of the ear to the bottom. The 2D ear localization technique proposed by Alvarez et al. [5] uses ovoid and active contour (snake) [6] 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. Yan and Bowyer [7] have proposed another technique by considering

a predefined sector from the nose tip as the probable ear region. The technique 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. Ansari and Gupta [8] have presented an ear detection technique based on edges of outer ear helices. It 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 fails if the helix edges are poor. Yuan and Mu [9] 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. Another ear localization technique which exploits the elliptical shape of the ear is proposed in [10]. But the assumption of elliptical ear shape for all subjects may not be true and may not help in detecting the ear, in general. As shown in Fig. 1, these techniques may correctly approximate the ear boundaries for round and oval shapes but may fail in case of triangular and rectangular shapes. Moreover, assumption of elliptical ear shape restricts the use of these techniques to 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. To detect ears at different scales, ear templates of different sizes are maintained. However, in practice, any predefined set of templates may not be able to handle all situations. In [12, 13], there are two techniques for ear localization which are also based on template matching. In these techniques,

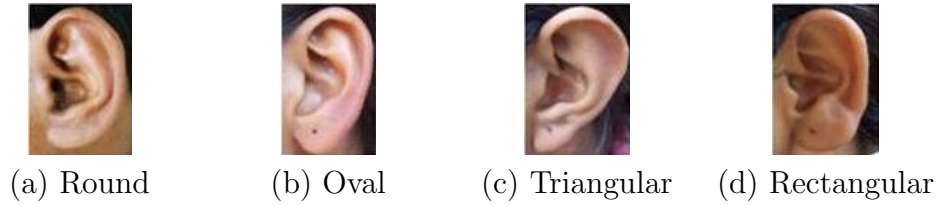


Figure 1: Ear Shapes

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 side face image which works well when side face includes only facial parts. But while capturing the side face, an image may include other skin parts such as neck. This makes the size of the skin area larger than the actual which leads to an incorrect resizing of the ear template and in turn, it produces an erroneous ear localization. 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 around the ear and fails if ear detection is carried out in whole side face image.

A cascaded AdaBoost based ear detection approach has been proposed in [15]. 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. In [16], ear localization has been proposed which is based on hierarchical 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 side face image does not

vary much. In presence of scale variation, cluster size of the ear needs to be adjusted which may not be possible without user intervention. In [17], 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, such an assumption of elliptical shape of the ear may be very rigid. Ibrahim et al. [18] have employed a bank of curved and stretched Gabor wavelets (popularly called banana wavelets) for ear detection. A technique which is based on skin segmentation and graph connected components has been proposed in [19]. The technique uses connected components of a graph constructed using the side face edge map for ear localization. It uses average vertex degree measure of a graph component for ear localization and assumes that the graph component with the largest average vertex degree represents the ear. However, often this criterion results into a false ear detection. It is due to the fact that in presence of noise or poor skin segmentation, a connected component not representing ear may also get the highest value of average vertex degree.

3. 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 [20] which has used 1976 CIE Lab color space for image representation. However, in this paper we have represented images in YCbCr space because it is perceptually uniform [21] and is widely used in video compression standards such as JPEG and MPEG [22].

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 characterizing 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) components 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 $N(\mu, \Sigma)$ with mean μ and covariance Σ . 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

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

Likelihood values obtained in Eqn. (1) can be used to segment skin and non-skin regions. An adaptive thresholding process [20] 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.

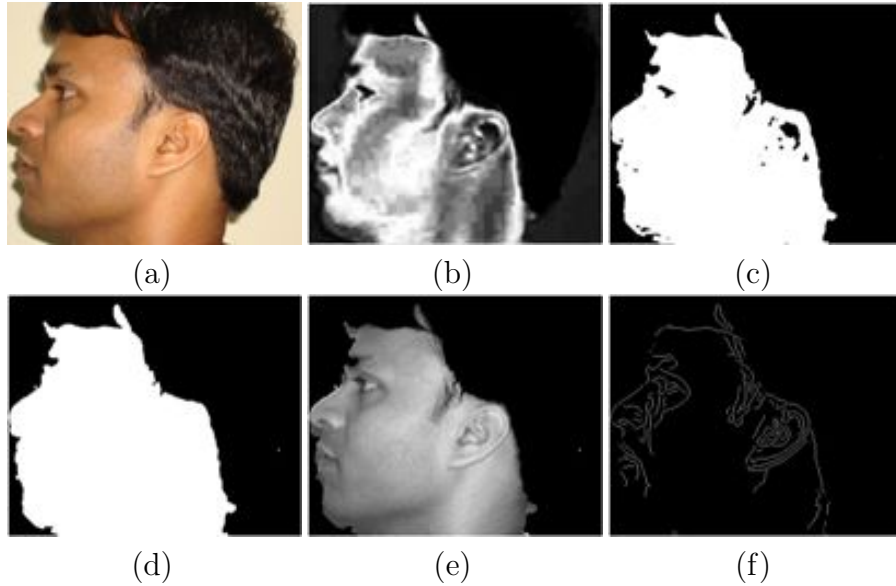


Figure 2: Skin Segmentation: (a) Input Image, (b) Skin-likelihood Image, (c) Binary Image (d) Dilated Binary Image, (e) Skin Segmented Image, (f) Edge Image

4. PROPOSED TECHNIQUE

The proposed technique is based on the fact that in a side face image, ear is the only part which contains much variation in the pixel intensities, resulting this place rich in edges. This can be visualized from the image shown in Fig. 2(f) which displays the edge image for the skin segmented image of Fig. 2(e). It can be observed that the ear part has larger edge density

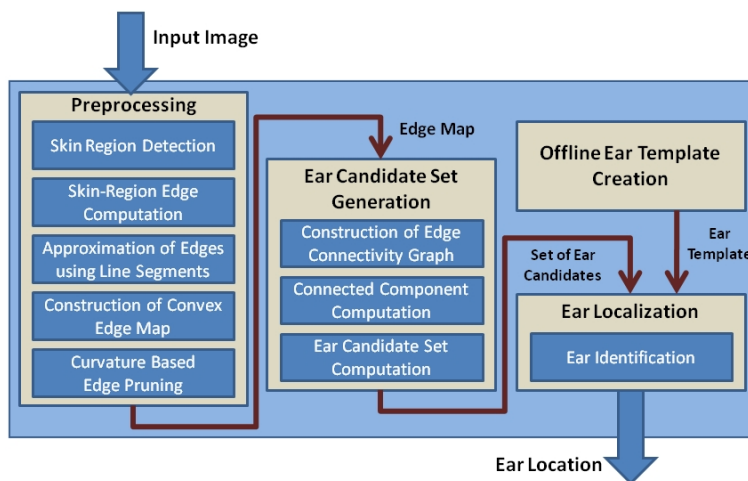


Figure 3: Flow Diagram of the Proposed Technique

compared to other parts. Further, all edges belonging to the ear part contain some curvature. These characteristics are exploited for ear localization. The proposed technique computes edge clusters in the edge map obtained from the side face image and examines them for ear localization. Flow diagram of the proposed technique is presented in Fig. 3.

4.1. Preprocessing

Side face image undergoes a preprocessing phase before ear localization. This involves skin segmentation, edge computation, edge approximation, convex edge map construction and edge pruning.

4.1.1. Skin Region Detection

This step isolates skin regions of the side face image from non-skin parts. Skin color model discussed in Section 3 is used for skin segmentation. This model transforms a color image into a gray scale image (called skin-likelihood image) using Eqn. (1) such that the gray value at each pixel shows the like-

likelihood of the pixel belonging to the skin. With an appropriate thresholding, the gray scale image is further transformed to a binary image depicting skin (white pixels) and non-skin (black-pixels) regions. Since people with different skins have different likelihood, an adaptive thresholding process [20] can be used to achieve the optimal threshold value for each image.

The binary image may contain some holes in it due to the presence of noise in side 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.

Fig. 2 considers an example of skin region detection with various intermediate steps. For a color image given in Fig. 2(a), corresponding skin-likelihood image is shown in Fig. 2(b). It can be noticed that skin regions in Fig. 2(b) are brighter than the non-skin regions. Fig. 2(c) shows the binary image obtained by thresholding the skin-likelihood image. Dilation is applied on this image to repair it by filling small holes present in it. Fig. 2(d) shows the repaired binary image. It is used for skin region detection where pixels of the side face image corresponding to white pixels of the binary image are considered as skin pixels. Fig. 2(e) shows the final skin segmented image. It can be observed from segmentation result that not all detected skin regions contain ear. Hence, ear localization can be used to locate the ear in all these skin like segments.

4.1.2. Skin-region Edge Computation

Edge detection is carried out on skin segmented image using Canny edge operator and a list of all edges is computed. An edge in the list is obtained

by connecting edge points together into a sequence of pixel coordinate pairs. Whenever an edge junction¹ is encountered, the sequence is terminated and a separate edge point sequence is generated for each of the branches and added to the list. This generates a set of edges containing two end points. Let χ be the set of all such edges.

4.1.3. Approximation of Edges using Line Segments

All pixels present in an edge (belonging to set χ) may not be equally important and may not be necessarily required to represent the edge. So to remove the redundant pixels from an edge and to get its compact representation, line segments are fitted to the edge. This keeps only those pixels which are important. It breaks every edge present in the image into a set of line segments. Line segments can be found by taking each array of edge points in the set χ and finding 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 process is repeated. In this manner each edge is broken into line segments, each of which adheres to the original data with the specified tolerance. Fig. 4(b) shows an example of edge approximation by line segments for the edge image in Fig. 4(a). Let χ_{ls} be the set containing all edges obtained after line segments fitting.

¹Edge junction is a pixel where an edge divides into two or more edges.



Figure 4: (a) Original Edge Image, (b) Edges Approximated by Line Segments

4.1.4. Construction of Convex Edge Map

It is observed that edges belonging to the ear have convex² nature. However, because of the presence of noise such as hair near the ear, often false edges due to noise join true ear edges and make them non-convex. It may lead to an improper ear localization. This usually happens with the outer helix edges of the ear. To avoid this, the derived edges with set χ_{l_s} are broken into a set of convex edges. Let χ_{convex} be the set of all convex edges. Identification of convex and non-convex edges and breaking the non-convex edges into convex can be done as follows.

Let there be an edge e obtained after approximation. Let e consist of k line segments with i^{th} line segment, l_i , having end points: t_i and t_{i+1} . Let the line segment l_i be represented by vector $\vec{v}_i = t_{i+1} - t_i$. Let $\vec{v}_{i,i+1}$ be the vector cross-product of \vec{v}_i and \vec{v}_{i+1} (vector representing line segment l_{i+1}). The edge e is convex if directions of $\vec{v}_{i,i+1}$, for all i , are found to be same. To test whether an edge e is convex or non-convex, a decision parameter ρ_e can be estimated as follows.

²Edges which have curvature throughout either positive or negative are considered convex.

Algorithm 1 Construction of Convex Edge Map

- **Input:** Set χ_{ls} of edges approximated with line segments.
- **Output:** Set χ_{convex} of convex edges.

```
1: Define a null set  $\chi_{convex}$ .
2: for  $\forall e \in \chi_{ls}$  do
3:   Compute  $\rho_e$  using Eqn. (2).
4:   if  $\rho_e == 0$  then
5:     Add  $e$  to  $\chi_{convex}$ .
6:   else
7:     Break  $e$  into a set of convex edges and add these edges to  $\chi_{convex}$ .
8:   end if
9: end for
```

$$\rho_e = \begin{cases} 0, & \text{if directions of vectors } \vec{v}_{(i,i+1)}, \forall i, \text{ are same} \\ 1, & \text{otherwise} \end{cases} \quad (2)$$

Thus the edge e is convex if ρ_e is 0. To break a non-convex edge into a set of convex edges, it is scanned from one side to another and direction of each cross-product is analyzed. When a cross-product is found to be of different direction with respect to the previous cross-product, the edge is broken at that point. This procedure is continued till whole edge is broken into convex edges. Steps for construction of convex edge map are given in Algorithm 1.

Fig. 5 presents an example of testing and breaking of edges into convex type. Fig. 5(a) shows two edges, one convex (edge $ABCD$) and another non-convex (edge $PQRS$). Fig. 5(b) shows vector representation of the line segments used in these edges and the direction of the cross-products for adjacent vectors. Circle with a cross and circle with a dot at the joining points of two vectors represent the outward and the inward directions of the

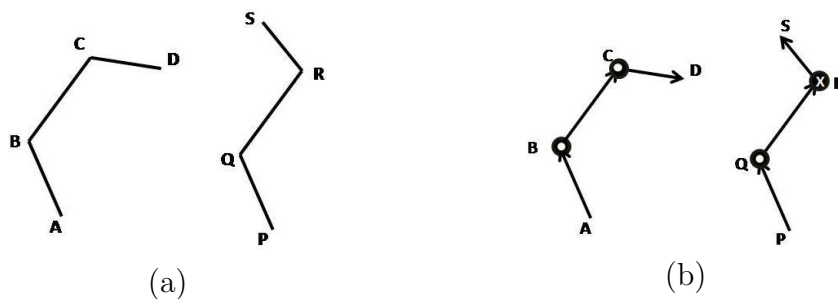


Figure 5: (a) Edge Map (b) Line Segments with Vector Representation

cross-product respectively. In edge $ABCD$ of Fig. 5(a), it can be observed that all cross-products are inward so this edge is marked as convex. In edge $PQRS$ of Fig. 5(b), one cross-product is inward while other is outward so the edge is marked as non-convex. While scanning the edge $PQRS$ from lower side, direction of the cross-product at point R is found to be different from the previous direction of the cross-product, so the edge $PQRS$ is broken at point R into two edges: PQR and RS .

Ear localization accuracy can be improved by converting all these edges to convex type. Breaking of non-convex edges into convex helps in removing the outlier edges (created due to noise). If the edges are converted to convex type, while constructing the edge connectivity graph, most of the outlier edges get isolated and do not appear in the connected component representing the ear and hence, do not affect the ear localization result. Fig. 6 shows one such example of ear detection. In Fig. 6(a), edge marked as A contains some erroneous part at its lower side arose due to the linking of true ear edge to a noisy edge present in the neck part. Due to this, when edge A participates in the connected component representing ear, localization result includes some skin portion from the neck which does not belong to the ear. Fig. 6(c) shows

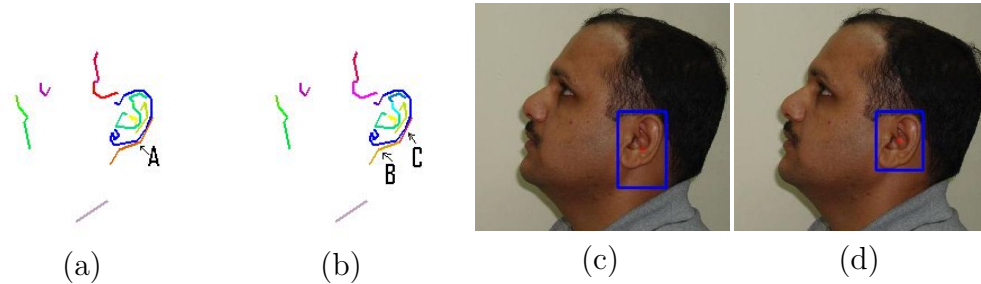


Figure 6: (a) Edges before Convex Type Segmentation, (b) Edges after Convex Type Segmentation, (c) Detected Ear when (a) is used, (d) Detected Ear when (b) is used

the localization result for this. When edge A is segmented into convex edges B and C (Fig. 6(b)), lower part of the edge A (*i.e.* B after breaking) gets isolated from the ear edge cluster and remaining ear edge cluster produces the correct localization result. Fig. 6(d) shows the localization result for this.

Any noise mainly affects the outer edge (helix) of the ear and hence, conversion of non-convex edges to convex primarily helps in removing noisy edges from the outer helix. Since detection of outer helix edge is difficult and computationally expensive, in the proposed technique all the edges are converted to convex type. However, conversion of non-convex edges present in the inner parts of the ear to convex type does not have any impact on the localization performance.

4.1.5. Curvature Based Edge Pruning:

All edges in set χ_{convex} are of convex nature and are represented by line segments. It can be seen that each edge in the set χ_{convex} represented by one line segment (or two points) depicts a linear edge in the original edge map (set χ). Since all edges belonging to the ear contain some curvature,

they need more than one line segment (or more than two points) for their representation. In other words, all edges having two points cannot be the part of ear edges and hence can be removed from the set χ_{convex} . This results a new edge set χ_c containing only the edges belonging to ear. Set χ_c can be formally defined as: $\chi_c = \{e \mid e \in \chi_{convex} \text{ and } \gamma(e) > 2\}$, where $\gamma(e)$ gives the number of points used in edge e to approximate it by line segments.

4.2. Ear Candidate Set Generation

This phase builds an edge connectivity graph which is used for finding the connected components in the graph to obtain ear candidate set.

4.2.1. Building Edge Connectivity Graph

The set χ_c can be used to define the edge map of the side face image. Let there be n edges in χ_c . The i^{th} edge e_i in χ_c is defined by a point p_i . Thus χ_c can be represented by a set P of points p_1, p_2, \dots, p_n where p_i refers to e_i for all i . Against each edge e_i , a convex hull³ $CH(e_i)$ is defined. If two convex hulls $CH(e_i)$ and $CH(e_j)$ intersect each other, then points p_i and p_j are connected through an edge of a newly defined graph $G = (V, E)$ with the set of vertices V and the set of edges E , where

$$\begin{aligned} V &= \{p_i \mid p_i \in P\} \\ E &= \{(p_i, p_j) \mid CH(e_i) \text{ intersects } CH(e_j)\} \end{aligned}$$

G is called edge connectivity graph. Algorithm 2 provides the steps invoked in building the graph G .

³Convex hull for an edge is a tightest convex polygon which includes all edge points.

Algorithm 2 Construction of Edge Connectivity Graph

- **Input:** Edge map χ_c of side face image I .
 - **Output:** Edge connectivity graph $G = (V, E)$.
- 1: Define a graph $G = (V, E)$ where V and E are initially null.
 - 2: Define a set $P = \{p_1, p_2, \dots, p_n\}$ for the n edges in set χ_c such that point p_i represents i^{th} edge in set χ_c .
 - 3: Define $V = \{p_i | p_i \in P\}$.
 - 4: Define convex hull CH_i for each edge e_i , $e_i \in \chi_c$.
 - 5: **for all** $i, j \in [1, n]$ **do**
 - 6: **if** $CH(e_i)$ intersects $CH(e_j)$ **then**
 - 7: Connect points p_i and p_j by an edge (p_i, p_j) in graph G and add it to E .
 - 8: **end if**
 - 9: **end for**
 - 10: Return G .
-

One can observe that the ear edges are mostly convex in nature and if one moves from outer part of the ear towards inside, most of the outer edges contain inner ear edges. Due to this nature of ear edges, convex hulls of the outer edges intersect the convex hulls of the inner edges. This guarantees that the convex hull of an edge belonging to the ear intersects at least another convex hull of the edge belonging to the ear. So this criterion to define connection between vertices (points) in a graph connects (directly or indirectly) all vertices belonging to the ear part with each other. Moreover, this criterion is able to define the connectivity irrespective of the scale; as a result, it makes the technique scale invariant. In general, property of one edge containing another is not true for the edges belonging to other parts of the side face image; so vertices corresponding to these edges remain mostly isolated in the edge connectivity graph.

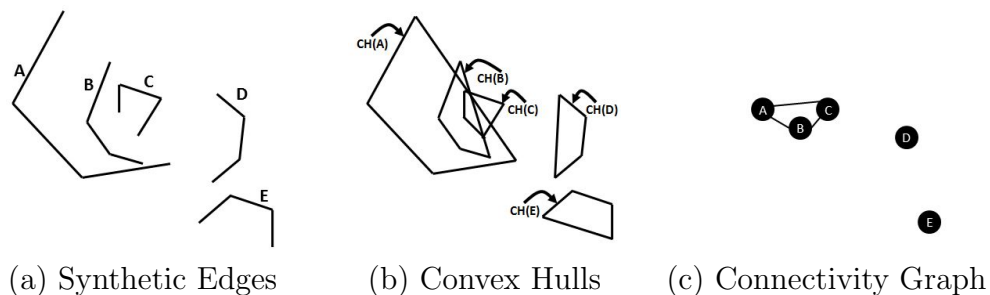


Figure 7: An Example of Construction of Edge Connectivity Graph

Fig. 7 shows an example of an edge map and convex hulls of edges. It is seen from Fig. 7(b) that convex hulls of edges A , B and C intersect with each other. So vertices corresponding to these edges are connected to each other in the graph as shown in Fig. 7(c). Points D and E are isolated in Fig. 7(c) since their respective convex hulls in Fig. 7(b) do not intersect to convex hull of any other edge.

It can be noted that there can be some simple criteria to define the connectivity among the vertices in edge connectivity graph. One such criterion may be based on some distance metrics between two edges. However, such choice makes ear detection scale dependent. This is due to the fact that the distance threshold required to define the connectivity among the vertices may vary for the images of different scales.

4.2.2. Connected Component Computation

Two vertices are in the same connected component of an undirected graph if there exists a path between them. After defining the graph for the edge map of side face image, its connected components are computed. These components are analyzed one by one to localize the ear. To compute the

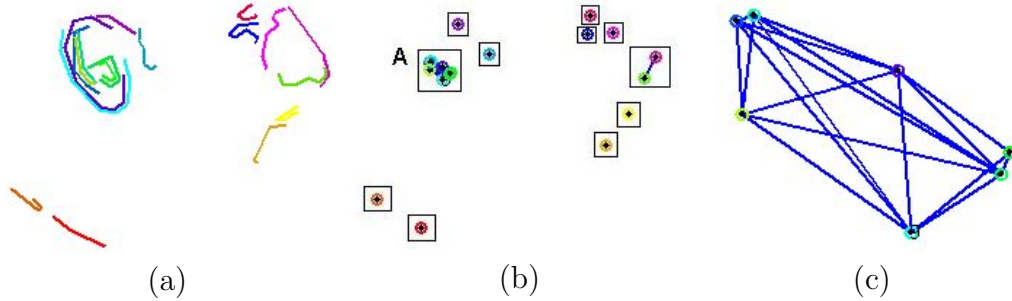


Figure 8: (a) Edge Map (Colors used to Differentiate Edges), (b) Graph for the Edge Map of (a) with Connected Components Labeling, (c) Magnified View of Component A of (b)

connected components in the graph $G = (V, E)$, we have used a breath first search based algorithm described in [23].

Fig. 8 presents a real example of edge connectivity graph and connected components labeling. Fig. 8(a) shows an edge image obtained from a side face image. A graph, shown in Fig. 8(b), is constructed for this edge image and connected components (enclosed inside rectangular boundaries) are computed. Magnified view of the component A present in Fig. 8(b) can be seen in Fig. 8(c).

4.2.3. Ear Candidate Set Computation

Ideally, it is believed that the vertices representing ear edges are connected to each other (directly or indirectly) and form one connected component while all other vertices representing non-ear edges remain isolated. Hence the criterion based on the size of the component can be used to find out the connected component representing ear. However, there may exist few more places in the side face where due to noise a convex hull of one edge may intersect that of other edges and give rise to a large connected com-

ponent. Hence, each connected component in the edge connectivity graph which have two or more vertices is considered as a probable candidate to represent the ear. Connected components which have single vertex can be straightaway removed from the graph as they cannot represent the ear. Let $K = \{K_1, K_2, \dots, K_m\}$ be the set of connected components of graph G where each component has two or more number of vertices. Average vertex degree of a connected component K_j is defined as:

$$d(K_j) = \frac{\sum_{i=1}^{n_j} d(p_i)}{n_j} \quad (3)$$

where $d(p_i)$ is the degree of vertex p_i and n_j is the total number of vertices present in component K_j . As stated earlier, ear part of the side face image is rich in edges due to large intensity variations present in this region; hence, it is less probable that a connected component representing an ear has only two vertices or average vertex degree one. Therefore, to further prune out the false connected components, only the components having average vertex degree greater than one can be considered to obtain probable ear candidates. A probable ear candidate in a side face image is defined as the image portion which is cropped using the bounding box of the edges participating in a connected component. A set of ear candidates is computed using all the connected components satisfying the criterion on the average vertex degree. Algorithm 3 presents the steps to generate ear candidate set using connected components.

Algorithm 3 Computation of Ear Candidate Set

- **Input:** Set $K = \{K_1, K_2, \dots, K_m\}$ containing m connected components of G .
 - **Output:** Set I_E containing the image portions cropped from the side face which are the probable candidates for ear.
- 1: **for** $j = 1$ to m **do**
 - 2: $d(K_j) = \frac{1}{n_j} \sum_{i=1}^{n_j} d(p_i)$, $p_i \in K_j$ and $d(p_i)$ is the degree of vertex p_i , n_j is the number of vertices in K_j .
 - 3: **end for**
 - 4: Define set $P = \{j | d(K_j) > 1\}$
 - 5: Define set $E = \{E_1, E_2, \dots, E_\eta\}$ where η is the cardinality of set P and $E_j = \{e_i | p_i \in K_j, j \in P\}$ contains edges with the edge e_i represented by point p_i in G as discussed in Section 4.2.1.
 - 6: Define $B = \{B_k | B_k \text{ is the bounding box of the edges present in } E_k\}$.
 - 7: Obtain $I_E = \{I_k | I_k \text{ is cropped image from side face using } B_k \in B\}$.
 - 8: Return I_E .
-

4.3. Ear Localization

It is carried out by identifying the true ear among the probable ear candidates with the help of an ear template which is created off-line. The template works as an ear representative which depicts the characteristics of ears of various scales, rotations and shapes. Identification is performed by comparing the probable ear candidates with the ear template.

4.3.1. Ear Template Creation

To identify true ear, the template used for ear identification should exhibit the characteristics of scale and rotation invariance. To compute such a template in the proposed technique, a shape descriptor which is invariant to rotation and scale, is used. Among several scale and rotation invariant

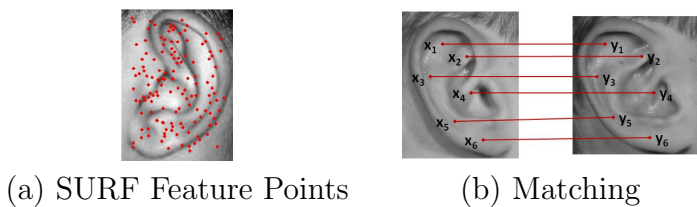


Figure 9: Example of SURF Features and Matching

shape descriptors, SURF [24] provides good distinctive features and at the same time it is robust to changes in viewing conditions, rotations and scales. SURF represents an image by first identifying some unique feature points in it and then by describing them with the help of a feature descriptor vector. Fig. 9(a) shows an example of SURF feature points. For the description of the feature points, SURF uses intensity content within the neighborhood of feature point and describes it by using the sum of approximated 2D Haar wavelet components. Typically, a SURF feature point is represented by a 64-dimensional descriptor vector.

The ear template is computed by fusing the SURF feature descriptors obtained from various ear images together considering the redundant features only once. Let n be the number of ear images used for template creation. Let T_1, T_2, \dots, T_n be the SURF feature descriptor sets obtained from these images. A fused ear template T is obtained by

$$T = \bigcup_{i=1}^n T_i \quad (4)$$

Let the set T_i contains c_i feature descriptor vectors. The total number of descriptor vectors c present in T satisfies the inequality $c \leq \sum_{i=1}^n c_i$. Fusion

of the SURF feature descriptor sets is done incrementally where first two sets T_1 and T_2 are fused to generate a new intermediate feature descriptor set which is further fused with feature descriptor set T_3 . This process is continued till all sets are fused together. While fusing two SURF feature descriptor sets T_i and T_{i+1} , SURF matching is performed between the two sets to find out the redundant feature descriptor vectors. If a descriptor vector in a template matches to a descriptor vector in another template, it is considered as common to both and is used only once in fusion. Fig. 9(b) shows an example where matching points between two ear images are shown. If a feature point x_i from the first ear image matches to a feature point y_i in the second ear image, either descriptor for x_i or for y_i is used in generation of fused feature descriptor set.

It can be noted here that attempts have been made to utilize the power of invariant feature points in other ear biometric systems as well. For example, Bustard and Nixon [25] have used Scale Invariant Feature Transform (SIFT) [26] feature points for registration of probe and gallery image before matching to perform ear recognition.

4.3.2. Ear Identification

Let the ear candidate set be $I_E = \{I_1, I_2, \dots, I_\eta\}$ where η is the cardinality of set I_E and I_k is the image portion of the side face image representing k^{th} probable ear candidate, $k = 1, 2, \dots, \eta$. For identification purpose, SURF feature descriptor set is computed for all the ear candidates present in I_E . Identification of true ear is performed by comparing the ear template with the SURF descriptor sets of the ear candidates present in I_E . Comparison between two SURF descriptor sets is performed using SURF matching which

uses the ratio-matching scheme [26] to find out the number of descriptor vectors matching between the two sets. Let D_i and D_j be two descriptor vectors from sets S_1 and S_2 respectively. Let $d(D_i, D_j)$ be a distance metric between the descriptor vectors D_i and D_j . The descriptor vector D_i is said to be matched with D_j if

$$d(D_i, D_j) < \rho \times d(D_i, D_k), \quad D_k \in S_2, k \neq j, \forall k \quad (5)$$

where ρ is a constant lying between 0 and 1. Very small value of ρ gives a tighter matching while a large value of ρ provides a relaxed matching.

Let $T_E = \{T_{I_1}, T_{I_2}, \dots, T_{I_\eta}\}$ be the SURF feature descriptor sets for the ear candidate images present in I_E . To obtain the true ear, SURF matching is performed between ear template (T) and all elements present in T_E and a match score vector *MatchScore* is generated. SURF matching between two descriptor sets returns the number of matched points between them. The true ear candidate I_ξ is obtained such that

$$\xi = \arg \max_i \{MatchScore[i]\}$$

That means, the ear candidate from I_E for which SURF match score is maximum, is declared as the true ear candidate. Algorithm 4 provides steps involved in ear identification process.

5. SCALE, ROTATION AND SHAPE INVARIANCE

In the proposed technique, there are two major steps which play key roles in ear localization. First step is the construction of edge connectivity

Algorithm 4 Ear Identification using SURF Descriptive Ear Template

- **Input:** Set $I_E = \{I_1, I_2, \dots, I_\eta\}$ containing η probable ear candidates and off-line created ear template T .
 - **Output:** I_ξ which is the true ear.
- 1: Define set $T_E = \{T_{I_1}, T_{I_2}, \dots, T_{I_\eta}\}$ where T_{I_i} represents SURF feature descriptor set for the ear candidate $I_i \in I_E$.
 - 2: **for** $i = 1$ to η **do**
 - 3: $MatchScore[i] = SURFmatch(T, T_{I_i})$.
 - 4: **end for**
 - 5: $\xi = \arg \max_i \{MatchScore[i]\}$.
 - 6: Return I_ξ .
-

graph which is used to detect probable ear candidates while second one is the identification of true ear among probable ear candidates using ear template. Construction of edge connectivity graph is made scale invariant by defining the connectivity among the vertices in the graph using intersection of convex hulls of corresponding edges. Such criterion to define the connectivity is unaffected by scale changes. Also, intersection of two convex hulls is unaffected if both are rotated; hence rotation also does not influence the process of defining the connectivity of two vertices in the graph. It can be observed that there is no significance of shape invariance at this step.

Rotation, scale and shape invariance at ear identification step is obtained by defining an ear template which exhibits these properties. It is achieved by using SURF feature descriptor for ear template creation which provides rotation and scale invariant description of ear feature points. An ear template is defined as a collection of rotation and scale invariant descriptor vectors obtained from multiple training ear images. Shape invariance is achieved

by choosing the ears of different shapes from the database to define the ear template.

6. EXPERIMENTAL RESULTS

The technique has been tested on two databases, namely IIT Kanpur (IITK) database and University of Notre Dame database (Collections E and J2) [27]. IITK database consists of three data sets. Data Set 1 contains 801 side face images collected from 168 subjects, 2 or more images per subject. These images include frontal view of the ears. Few sample images from Data Set 1 are shown in Fig. 10(a). Data Set 2, whose acquisition setup is shown in Fig. 12(a), contains 801 side face images collected from 89 subjects, 9 images per subject for various in-plane rotations and scales. Images contains frontal view of the ear taken at three different positions, a person looking straight, the person looking at 20^0 (approx) down and looking at 20^0 (approx) up. At all these 3 positions, images are captured at 3 different scales by setting the digital zoom of the camera at 1.7x, 2.6x and 3.3x and positioning the camera at a distance of about 1 meter. Fig. 10(b) shows 9 images from Data Set 2 captured for an individual. Data Set 3, whose acquisition setup is shown in Fig. 12(b), contains complex images captured for various out-of-plane rotations from 107 subjects. The camera is moved on a circle with the subject assumed to be at the center of the circle. Camera facing the frontal of the ear is considered as 0^0 . Side face images are captured at -40^0 , -20^0 , 0^0 , $+20^0$ and $+40^0$ placing the camera tripod at fixed landmark positions. Two images for each pose (angle) are obtained, producing 10 images per subject. Database contains 1070 images from 107 subjects. Fig. 11 shows a sample

Table 1: Gaussian Parameters used for Skin Segmentation in IITK and UND Databases

Data Set	Mean (Cb, Cr)	Covariance (Cb, Cr)
IITK Data Sets 1, 2 and 3	$\begin{pmatrix} 102.35 \\ 154.52 \end{pmatrix}$	$\begin{pmatrix} 71.76 & 9.95 \\ 9.95 & 111.77 \end{pmatrix}$
UND-E Data Set	$\begin{pmatrix} 90.65 \\ 170.23 \end{pmatrix}$	$\begin{pmatrix} 55.55 & -4.79 \\ -4.79 & 107.19 \end{pmatrix}$
UND-J2 Data Set	$\begin{pmatrix} 109.48 \\ 148.31 \end{pmatrix}$	$\begin{pmatrix} 55.74 & 41.76 \\ 41.76 & 93.62 \end{pmatrix}$

snapshot of angular posed side face images for a subject.

Further, Collection E (UND-E) and Collection J2 (UND-J2) of University of Notre Dame database are used for the experiment. UND-E consists of 464 side face images collected from 114 subjects, 3 to 9 samples per subject. Images are collected on different days with different conditions of pose and illumination. UND-J2 consists of 2414 side face images collected from 415 subjects. For the experiment, 1780 2D side face images are considered, after removing all duplicate images. Few sample images from UND database are shown in Fig. 13.

6.1. Estimation of Parameters

Skin parameters used for skin segmentation are computed for each data set separately by collecting skin samples from them. Table 1 summarizes skin parameters for various data sets. Minimum and maximum thresholds used in Canny edge detector are 0.0 and 0.1 respectively while σ is set to 1.0 for IITK database and 2.0 for UND database. Value of σ is kept little high for UND database as images in it are noisy. Distance tolerance for edge approximation is set to 20 for both the databases.

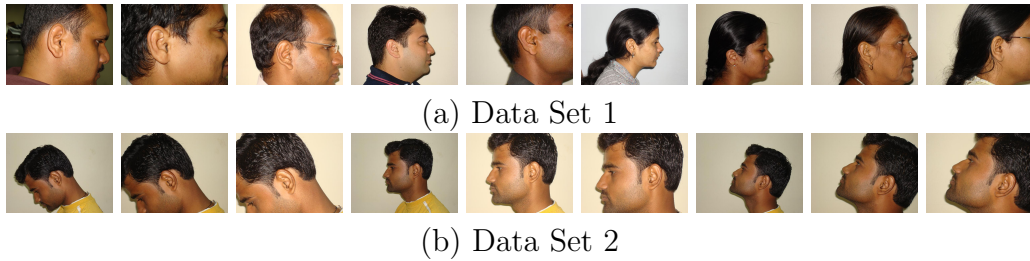


Figure 10: Sample Images from IITK Database

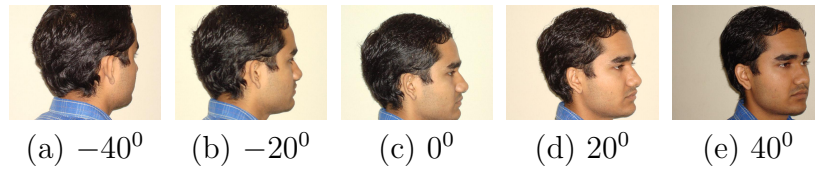


Figure 11: Sample Posed Images for an Individual from IITK Data Set 3

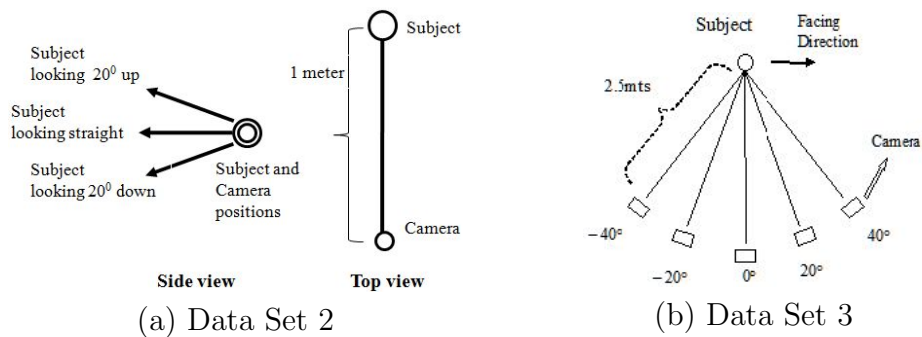


Figure 12: Data Acquisition Setup used in IITK Database

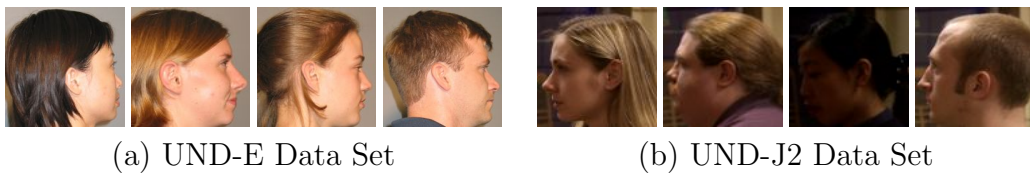


Figure 13: Sample Images from UND Database

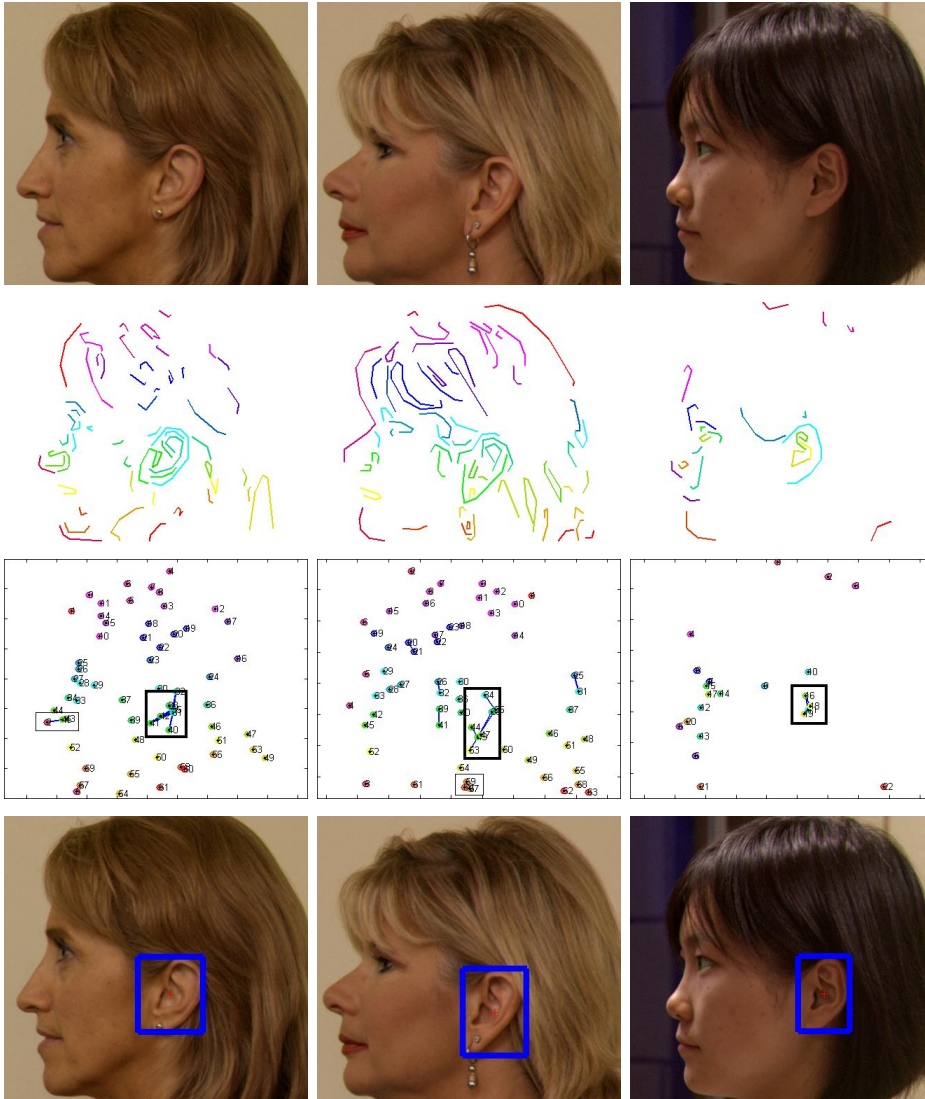


Figure 14: Ear Detection: Row-1) Original Input Images, Row-2) Edge Maps Approximated with Lines (Colors used to Distinguish Edges), Row-3) Edge Connectivity Graphs (Graph Components having Average Vertex Degree > 1 Enclosed in Rectangles), Row-4) Ear Detection Results

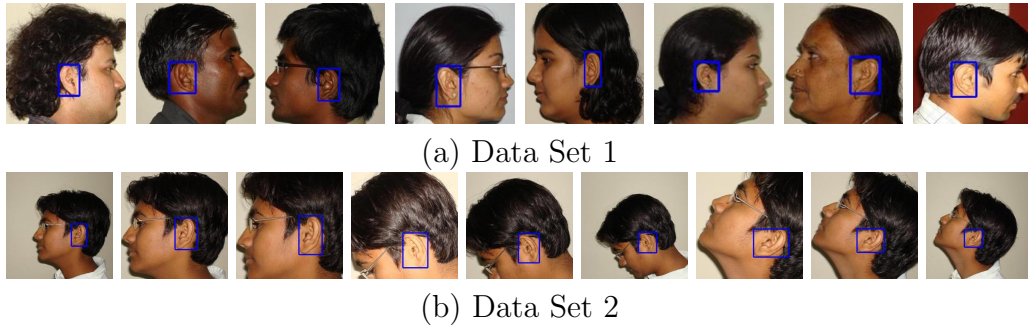


Figure 15: Ear Detection Results for IITK Database

Ear template for each data set of IITK and UND databases has been created separately as the nature of the data present in each set is entirely different. Few images are randomly selected from each data set to compute ear templates. It is found that 50 images from a data set are sufficient to capture the properties of the ears for creating a good ear template. The ratio value ρ used in SURF matching for template creation is taken as 0.5 whereas for true ear identification, it is set to 0.7. Since for template creation, SURF matching is performed between the ear images, a lower value of ρ (which gives tighter matching) helps in capturing the distinct features of the ears. Ear identification is used to discriminate ear and non-ear candidates and hence matching is relaxed and little higher value of ρ is used.

6.2. Results

Fig. 14 provides the results obtained at various steps of the ear detection. It shows the original input images, side face edge maps approximated with lines, edge connectivity graph and ear detection results. Accuracy of ear

Table 2: Percentage Accuracy for IITK Database

Data Set	# of Test Images	Ear Localization Accuracy (%)	
		Reported in [19]	Proposed Method
Data Set 1	801	95.88	99.25
Data Set 2	801	94.73	98.50
Data Set 3	1070	91.11	95.61

localization is defined by

$$\text{Accuracy} = \frac{\text{Number of Correct Localizations} \times 100}{\text{Total Test Samples}}\% \quad (6)$$

Fig. 15(a) shows the ear detection results for Data Set 1 of IITK database which contains normal frontal ear images. To show the rotation (pose) and scale invariance of the proposed technique, Data Set 2 of IITK database is used. Fig. 15(b) gives few results from Data Set 2 where ears of different sizes and rotations are efficiently detected without any user intervention and change of parameters. The proposed technique has also detected ears successfully for the images of Data Set 3 of IITK database (where images contain out-of-plane rotations) even for the extreme poses (-40° and $+40^\circ$). Fig. 16(a) shows detection results for few images taken from Data Set 3. Further, few ear localization results for extreme poses (-40°) where ear localization is found to be very challenging are shown in Fig. 16(b). The technique has localized ears precisely for almost all extreme cases. Since it uses edge property of the ear and not the shape, it exhibits shape invariance. It has detected ears of all shapes (viz. round, oval, triangular, rectangular) successfully.

Table 2 summarizes ear detection results for IITK database. It is seen that



(a) Detection Results for two Subjects



(b) Detection in Extreme Views

Figure 16: Ear Detection Results for IITK Database (Data Set 3)

accuracy for Data Set 1 is the highest as it contains frontal ear images. In such images, full ear structure is visible and good amount of edges are obtained which help in achieving strong connectivity among the edges representing ear. Accuracy for Data Set 2 is comparable with that of Data Set 1, in spite of images having variations in scale and rotation. This is due to the fact that the proposed technique exploits the structural details of the ear which do not change with scale and rotation. Data Set 3 shows the least accuracy among all data sets of IITK database. This is because in the presence of out-of-plane rotation, the availability of the structural details of the ear decreases as camera moves away from the frontal position. Ear localization results for IITK database are compared in Table 2 with the results reported in [19]. It is evident that the proposed technique performs much better than the technique discussed in [19]. This improvement is achieved due to following

reasons:

1. The proposed technique breaks the derived edges of the side face into a set of convex edges to reduce the participation of noisy edges in the cluster of true ear edges.
2. The proposed technique has used a rotation, scale and shape invariant ear template which depicts the characteristics of ears of various scales, rotations and shapes. Identification of the true ear is performed by comparing the probable ear candidates with the ear template. Use of rotation, scale and shape invariant ear template greatly helps in localization of ears of various poses, scales and shapes.
3. Identification of true ear among the probable ear candidates with the help of an ear template results into much better and robust ear localization and reduces false positives. But, the technique in [19] performs ear localization merely based on the size of the connected components which often leads to wrong ear localization as there may exist a cluster of the largest size of non-ear edges.
4. The performance obtained in the proposed technique is found to be robust and stable on a larger data set as compared to [19].

Results for University of Notre Dame (UND) database are shown in Fig. 17 while their localization accuracies are given in Table 3. Accuracy 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 proposed technique is derived from the successful detection of skin regions, similarity of the hair color with skin reduces the performance of skin

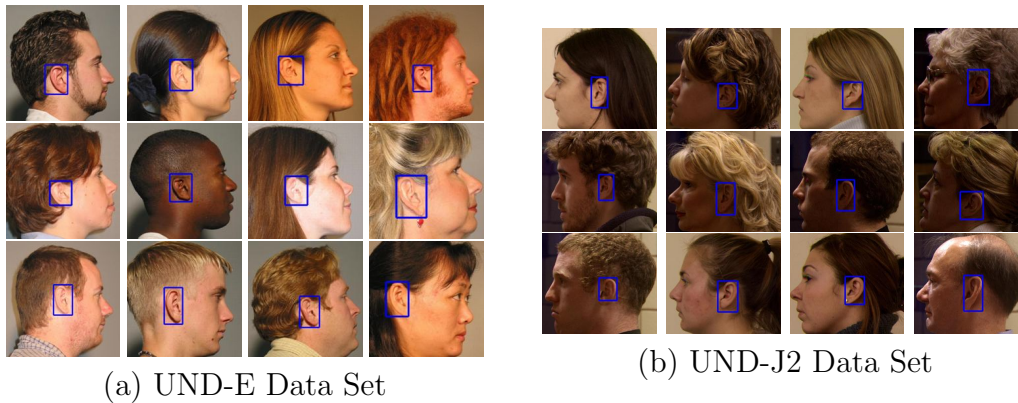


Figure 17: Ear Detection Results for UND Database

segmentation and in turn, affects the ear localization accuracy and increases false positives.

Table 3 also shows comparative performance of some well known techniques on UND database. It is seen from the table that [10] produces low detection rate as compared to the proposed technique. Moreover, it makes the assumption 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 or posed ear may lead to false detections.

Table 3: Percentage Accuracy for UND Database

Technique	Data Set	# of Test Images	Localization Accuracy (%)
[10]	Part of UND-J2	942	91%
[15]	Part of UND-J2	203	100%
Proposed	UND-J2	1780	96.63%
	UND-E	464	96.34%

The technique discussed in [15] 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, this 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 applications. On the other hand, the technique proposed in this paper 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 of [15] with the proposed technique is given in Table 4.

Performance of the proposed technique could not be compared with [5] because of the non-availability of the test results. Also comparisons could not be made with [17, 18] as these techniques have used XM2VTS database [28] 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.

The proposed technique has failed to detect ears fully or partially in some cases of IITK and UND databases. Failure has occurred when ears are

Table 4: Comparison with the technique discussed in [15]

Parameters	Techniques	
	[15]	Proposed Technique
Time per detection (same configuration)	26.40 Seconds	7.95 Seconds
Training Overhead	More. To train classifiers with 1000s of positive and negative samples	Very Less. Only required to learn skin parameters and ear 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 variation	Good amount of scaling and rotation (IITK Data Sets 2 and 3)

occluded by hair or affected by noise and poor illumination. Few examples of failure in detecting ears due to these reasons are shown in Figure 18.

7. CONCLUSIONS

This paper has proposed an efficient technique for automatic ear localization from the side face which can be employed to automate an ear biometric

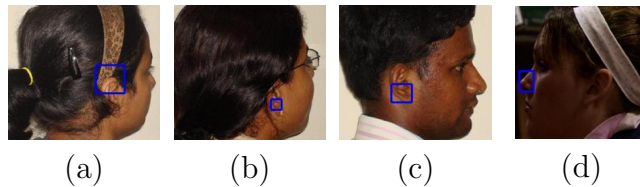


Figure 18: Few Failure Cases from IITK and UND Databases

system. The technique can detect ears of different rotations (poses), scales (sizes) and shapes efficiently and does not require any prior knowledge of rotation, shape or size of the ear for localization. It has been tested on 4916 side face images of IITK and UND databases. IITK database includes images of various rotations (in-plane and out-of-plane), scales and shapes while UND database consists of frontal ear images with variable background, varied contrast, pose variations and illumination changes. It has been found to provide very good ear detection accuracy for both the databases.

Comparing to the well known ear detection techniques, the proposed technique is found to be unique and significant in many aspects. It is the first technique proposed for ear localization which is rotation (in-plane and out-of-plane), scale and ear shape invariant. It can detect both left and right ears without any prior information. It does not require any kind of user intervention for ear localization. It has addressed the issues that has plagued earlier attempts of ear detection, specifically for rotated and scaled ear images.

ACKNOWLEDGMENT

Authors are thankful to the anonymous reviewers and the Editor for their valuable suggestions which have helped us to improve the quality of the manuscript. Authors also like to acknowledge the support provided by the Department of Information Technology, Government of India to carry out this work.

References

- [1] M. Burge, W. Burger, Ear biometrics in computer vision, in: Proc. of International Conference on Pattern Recognition (ICPR' 00), Vol. 02, 2000, pp. 822–826.
- [2] D. J. Hurley, M. S. Nixon, J. N. Carter, Force field feature extraction for ear biometrics, *Computer Vision and Image Understanding* 98 (3) (2005) 491–512.
- [3] H. Chen, B. Bhanu, Human ear detection from side face range images, in: Proc. of International Conference on Pattern Recognition (ICPR' 04), Vol. 3, 2004, pp. 574–577.
- [4] P. Yan, K. Bowyer, Empirical evaluation of advanced ear biometrics, in: Proc. of International Conference on Computer Vision and Pattern Recognition-Workshops, Vol. 3, 2005, pp. 41–48.
- [5] L. Alvarez, E. Gonzalez, L. Mazon, Fitting ear contour using an ovoid model, in: Proc. of IEEE International Carnahan Conference on Security Technology (ICCST' 05), 2005, pp. 145–148.
- [6] M. Kass, A. Witkin, D. Terzopoulos, Snakes: Active contour models, *International Journal of Computer Vision* 1 (4) (1988) 321–331.
- [7] P. Yan, K. Bowyer, Biometric recognition using 3D ear shape, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29 (8) (2007) 1297–1308.
- [8] S. Ansari, P. Gupta, Localization of ear using outer helix curve of the ear, in: Proc. of the International Conference on Computing: Theory and Applications (ICCTA' 07), 2007, pp. 688–692.
- [9] L. Yuan, Z.-C. Mu, Ear detection based on skin-color and contour information, in: Proc. of International Conference on Machine Learning and Cybernetics (ICMLC' 07), Vol. 4, Hong Kong, 2007, pp. 2213–2217.
- [10] B. Arbab-Zavar, M. S. Nixon, On shape-mediated enrolment in ear biometrics, in: Proc. of the 3rd International Conference on Advances in Visual Computing - Volume Part II, 2007, pp. 549–558.

- [11] A. Sana, P. Gupta, R. Purkait, Ear biometric: A new approach, in: Proc. of International Conference on Advances in Pattern Recognition (ICAPR' 07), World Scientific, Singapore, 2007, pp. 46–50.
- [12] S. Prakash, U. Jayaraman, P. Gupta, Ear localization from side face images using distance transform and template matching, in: Proc. of IEEE International Workshop on Image Proc. Theory, Tools and Applications (IPTA' 08), IEEE Xplore, Sousse, Tunisia, 2008, pp. 1–8.
- [13] S. Prakash, U. Jayaraman, P. Gupta, A skin-color and template based technique for automatic ear detection, in: Proc. of International Conference on Advances in Pattern Recognition (ICAPR' 09), IEEE Computer Society, ISI Kolkata, India, 2009, pp. 213–216.
- [14] S. Attarchi, K. Faez, A. Rafiei, A new segmentation approach for ear recognition, in: Proc. of International Conference on Advanced Concepts for Intelligent Vision Systems, 2008, pp. 1030–1037.
- [15] S. Islam, M. Bennamoun, R. Davies, Fast and fully automatic ear detection using cascaded adaboost, in: Proc. of IEEE Workshop on Applications of Computer Vision (WACV' 08), 2008, pp. 1–6.
- [16] S. Prakash, U. Jayaraman, P. Gupta, Ear localization using hierarchical clustering, in: Proc. of SPIE International Defence Security and Sensing Conference, Biometric Technology for Human Identification VI, 730620, Orlando, FL, 2009.
- [17] A. Cummings, M. Nixon, J. Carter, A novel ray analogy for enrolment of ear biometrics, in: Proc. of International Conference on Biometrics: Theory, Applications and Systems (BTAS' 10), 2010, pp. 1–6.
- [18] M. I. S. Ibrahim, M. S. Nixon, S. Mahmoodi, Shaped wavelets for curvilinear structures for ear biometrics, in: Proc. of 6th International Conference on Advances in Visual Computing (ISVC'10)- Part I, 2010, pp. 499–508.
- [19] S. Prakash, U. Jayaraman, P. Gupta, Connected component based technique for automatic ear detection, in: Proc. of 16th IEEE International Conference on Image Processing (ICIP' 09), Cairo, Egypt, 2009, pp. 2741–2744.

- [20] J. Cai, A. Goshtasby, Detecting human faces in color images, *Image and Vision Computing* 18 (1) (1999) 63–75.
- [21] C. A. Poynton, *A technical introduction to digital video*, John Wiley & Sons, Inc., New York, NY, USA, 1996.
- [22] C. Garcia, G. Tziritas, Face detection using quantized skin color regions merging and wavelet packet analysis, *IEEE Transactions on Multimedia* 1 (3) (1999) 264–277.
- [23] J. Hopcroft, R. Tarjan, Algorithm 447: efficient algorithms for graph manipulation, *Communication of ACM* 16 (6) (1973) 372–378.
- [24] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, Speeded-up robust features (SURF), *Computer Vision and Image Understanding* 110 (3) (2008) 346–359.
- [25] J. Bustard, M. Nixon, Robust 2D ear registration and recognition based on SIFT point matching, in: *Proc. of International Conference on Biometrics: Theory, Applications and Systems (BTAS' 08)*, 2008, pp. 1–6.
- [26] D. G. Lowe, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision* 60 (2) (2004) 91–110.
- [27] University of Notre Dame Profile Face Database, Collections E and J2, <http://www.nd.edu/~cvr1/CVRL/DataSets.html>.
- [28] K. Messer, J. Matas, J. Kittler, J. Lttin, G. Maitre, XM2VTSDB: The extended M2VTS database, in: *Proc. of 2nd International Conference on Audio and Video-based Biometric Person Authentication*, 1999, pp. 72–77.