Human Ear Detection from Side Face Range Images

Hui Chen and Bir Bhanu
Center for Research in Intelligent Systems
University of California, Riverside, CA 92521, USA
{hchen, bhanu}@vislab.ucr.edu

Abstract

Ear detection is an important part of an ear recognition system. In this paper we address human ear detection from side face range images. We introduce a simple and effective method to detect ears, which has two stages: offline model template building and on-line detection. The model template is represented by an averaged histogram of shape index. The on-line detection is a four-step process: step edge detection and thresholding, image dilation, connect-component labeling and template matching. Experiment results with real ear images are presented to demonstrate the effectiveness of our approach.

1. Introduction

Ear is a viable new class of biometrics since the ear has desirable properties such as universality, uniqueness and permanence [2, 5, 6]. Although it has certain advantages over other biometrics, it has received little attention compared to other popular biometrics such as face, fingerprint and gait. The current research has used intensity images and, therefore, the performance of the systems is greatly affected by imaging problems such as lighting and shadows. Range sensors which are insensitive to above imaging problems can directly provide us 3D geometric information. Therefore it is desirable to design a human ear recognition system from 3D side face range images obtained at a distance. Human ear detection is the first task of a human ear recognition system and its performance significantly affects the overall quality of the system.

We propose a technique shown in Figure 1 for ear detection from side face range images. Our approach has two stages: off-line model template building and on-line ear detection. Ear can be thought of as a rigid object with much concave and convex areas. As compared to other approaches for object detection in range images [7, 8, 12], We use the averaged histogram of shape index to represent the ear model template since shape index is a quantitative measure of the shape of a surface [3]. During the on-line detection, we first perform the step edge computation and thresholding since there is a sharp step edge around the ear boundary, then we do image dilation and connected-component analysis to find the potential regions containing ears. Next
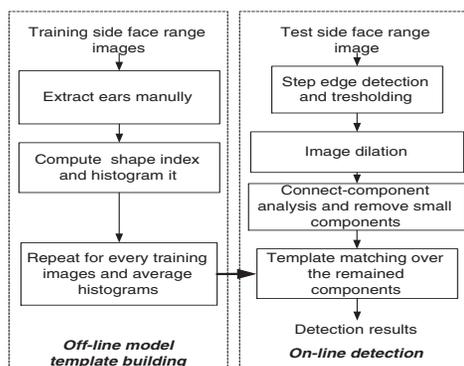
Figure 1. System diagram.

for every potential region, we grow the region and find the region with minimum dissimilarity between the histogram of shape indexes with the model template. Finally among all of the regions, we choose the one with the minimum dissimilarity as the detected region containing ear.

The contributions of our work are: (a) we develop a simple and effective approach to detect human ears from side face range images. Other researches have used Fisher linear discriminant [14], support vector machines [9, 13] and neural network [10] to do face detection, which are relatively complex. (b) we use the histogram of shape index to represent the model template since shape index captures the geometric information of the ear. (c) we only search the potential regions around the step edges, avoiding the exhaustive search over the entire test images.

The rest of paper is organized as follows. In Section 2, we describe our approach to build model template and detect human ears in side face range images in detail. Section 3 gives the experiment results to demonstrate the effectiveness of our approach. Conclusion is provided in Section 4.

2. Technical approach

2.1 Off-line model template building

• **Shape index:** Shape index S_i, a quantitative measure of the shape of a surface at a point p, is defined by (1),

$$S_i(p) = \frac{1}{2} - \frac{1}{\pi} tan^{-1} \frac{k_1(p) + k_2(p)}{k_1(p) - k_2(p)} \qquad (1)$$

where k_1 and k_2 are maximum and minimum principal curvatures respectively [3]. With this definition, all shapes can be mapped into the interval $S_i \in [0, 1]$.

Ear has significant convex and concave areas, which gives us a hint to use the shape index for ear detection. The original ear range image and its shape index image are shown in Figure 2. In Figure 2(b), the brighter points denote large shape index values which correspond to ridge and dome surfaces. We believe the ridge and valley areas form a pattern for ear detection. We use the distribution of shape index as a robust and compact descriptor since 2D shape index image is much too detailed. The histogram h can be calculated by $h(k) = \# \ of \ points \ with \ shape \ index \ \in \ bin(k)$. The histogram is normalized during the implementation.
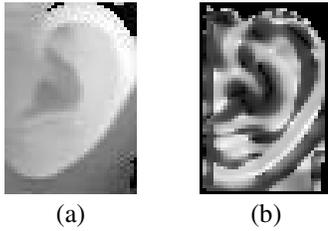
Figure 2. (a) Ear range image. (b) Its shape index image.

• **Curvature estimation:** In order to estimate curvatures, we fit a quadratic surface $f(x, y) = ax^2 + by^2 + cxy + dx + ey + f$ to a local window centered at the surface point of interest and use the least square method to estimate the parameters of the quadratic surface. After we get the parameters, we use differential geometry to calculate the surface normal, Gaussian and mean curvatures and principal curvatures [1, 4].

• **Model template building:** Given a set of training side face range images, first we extract ears manually, then calculate its shape index image and histogram the shape index image. After we get the histograms for each training image, we average the histograms and use the averaged histogram as our model template.

2.2 Step edge detection and thresholding and dilation

The step edge magnitude, denoted by M_{step}, is calculated as the maximum distance in depth between the center point and its neighbors in a small window [15]. M_{step} can be formulated as follows:

$$M_{step}(i, j) = max|z(i, j) - z(i + k, j + l| \quad (2)$$
$$-(w - 1)/2 < k, l < (w - 1)/2$$

Where w is the width of the window and $z(i, j)$ is the z coordinate of the point(i, j). To get the step edge magnitude image, a $w \times w$ window is translated over the original side face range image and the maximum distance calculated from (2) replaces the pixel value of the pixel covered by

the center of the window. The original side face range image and its step edge magnitude image are shown in Fig. 3(a) and (b). From Fig. 3(b), we clearly see that there is a sharp step edge around the ear boundary since brighter points mean large step edge magnitude. The step edge im-
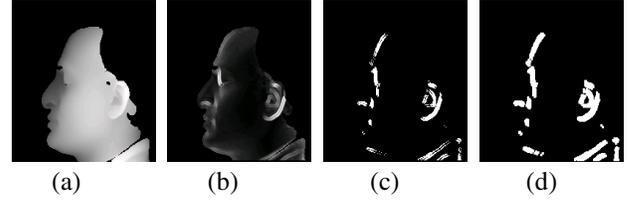
Figure 3. (a) Original side face range image. (b) Its step edge magnitude image. (c) Thresholded binary image. (d) Dilated image.

age is thresholded to get a binary image which is shown in Fig. 3(c). The threshold is set based on the maximum of M_{step}. Thus, we can get a binary image by using (3),

$$F_T(i, j) = \begin{cases} 1 & \text{if } M_{step}(i, j) \geq \alpha * max\{M_{step}\} \\ & 0 \leq \alpha \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

• **Image dilation:** There are some holes in the thresholded binary image and we want to get the potential regions containing ears. We dilate the binary image to fill the holes. The dilated image is shown in Fig. 3(d).

2.3 Connected Component labeling

Using the above result, we proceed to determine which regions can possibly contain human ears. To do so, we need to determine the number of potential regions in the image. By running connected component labeling algorithm, we can determine the number of regions. We used an 8-connected neighborhood to label a pixel. We remove smaller components whose area are less than β since the ear region is not small. The labeling result is shown in Fig. 4(a) and the result after removing smaller components is shown in Fig. 4(b).

Figure 4. (a) Labeled image. (b) Labeled image after removing smaller components.

After we get regions, we need to know the geometric properties such as the position and orientation. The position of a region may be defined using the center of the region.

The center of area in binary images is the same as the center of the mass and it is computed as below.

$$\bar{x} = \frac{1}{A} \sum_{i=1}^{n} \sum_{j=1}^{m} jB[i,j], \quad \bar{y} = \frac{1}{A} \sum_{i=1}^{n} \sum_{j=1}^{m} iB[i,j] \quad (4)$$

where B is $n \times m$ matrix representation of the region and A is the size of the region.

For the orientation, we find the axis of elongation of the region. Along this axis the moment of the inertia will be the minimum. The axis is computed by finding the line for which the sum of the squared distances between region points and the line is minimum. The angle of θ is given by (5):

$$\theta = \frac{1}{2} tan^{-1} \frac{b}{a-c} \quad (5)$$

The parameters a, b and c are given by (6), (7) and (8) respectively.

$$a = \sum_{i=1}^{n} \sum_{j=1}^{m} (x'_{ij})^2 B[i,j] \quad (6)$$

$$b = 2 \sum_{i=1}^{n} \sum_{j=1}^{m} x'_{ij} y'_{ij} B[i,j] \quad (7)$$

$$c = \sum_{i=1}^{n} \sum_{j=1}^{m} (y'_{ij})^2 B[i,j] \quad (8)$$

Where $x' = x - \bar{x}$ and $y' = y - \bar{y}$. θ gives us the hint about the region growing direction.

2.4 Template matching

As mentioned in Section 2.1, the model template is represented by an averaged histogram of shape index. Since histogram can be thought of as an approximation of probability distribution function, it is natural to use the $\chi^2 - divergence$ function (9) [11].

$$\chi^2(Q,V) = \sum_i \frac{(q_i - v_i)^2}{q_i + v_i} \quad (9)$$

Where Q and V are normalized histograms. From (9), we know the dissimilarity is between 0 and 2. If the two histograms are exactly the same, the dissimilarity will be zero. If the two histograms do not overlap with each other, it will achieve the maximum value 2.

From the Section 2.3, we get the potential regions which may contain the ears. For each region, we can find a minimum rectangular box to include the region, then we grow the region based on the angle θ. If $0 \leq \theta \leq \pi/2$, we grow the rectangle by moving the top-right vertex right, up and anti-diagonal and moving the bottom-left vertex left, down and anti-diagonal. If $\pi/2 \leq \theta \leq \pi$, we grow the rectangle by moving the top-left vertex left, up and diagonal
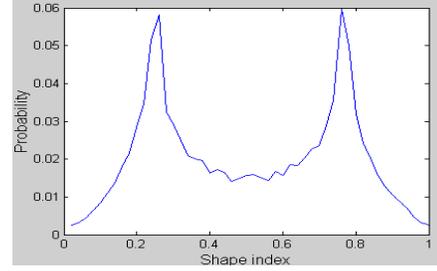
Figure 5. Model template (discretized 50 bins).

and moving the bottom-right vertex right, down and diagonal. For every region, we choose the grown rectangular box with the minimum dissimilarity as the candidate ear region. Finally over all of the candidate regions, we select the one with the minimum dissimilarity as the detected region. We set a threshold γ for region growing, which controls the size of the region.

3. Experiments

• **Data and parameters:** We use real range data acquired using Minolta Vivid 300. [1] There are 30 subjects in our database and every subject has two side face range images taken at different viewpoints. The data is split into training and test sets. Each set has 30 subjects. When we built the model template, we only use the first 20 subjects in the training set. The manually extracted ears are shown in Figure 6. The model template built using these 20 training images is shown in Fig. 5. The training set includes 30 images and there is no overlap between training and testing images.

The parameters of our approach are $\alpha = 0.35$, $w = 5$ pixels, $\gamma = 35$ pixels and $\beta = 99$ pixels. The bin size of the histogram is 0.02.

• **Results:** The detection results on real range data are shown in Fig 8. In Fig 8, We draw a rectangle for the detected region. We evaluate our detection results in terms of correct detection rate and false alarm rate. Fig. 7 shows how to calculate the correct detection rate and false alarm rate given the ground truth rectangle and detected rectangle. We manually get ground truth rectangle which is the minimum rectangle bounding the ear. We get 91.5% correction rate with 2.52% false alarm rate.

4. Conclusions

In this paper, we present a technique for ear detection from side face range images. The success of our approach relies on two facts: 1) there is a sharp step edge around the ear boundary which can be easily detected; 2) shape index is a good measurement to capture the geometric information of ears since the ear has much ridge and valley areas. Furthermore, our approach is simple, effective and

---

[1] http://www.minoltausa.com/vivid/
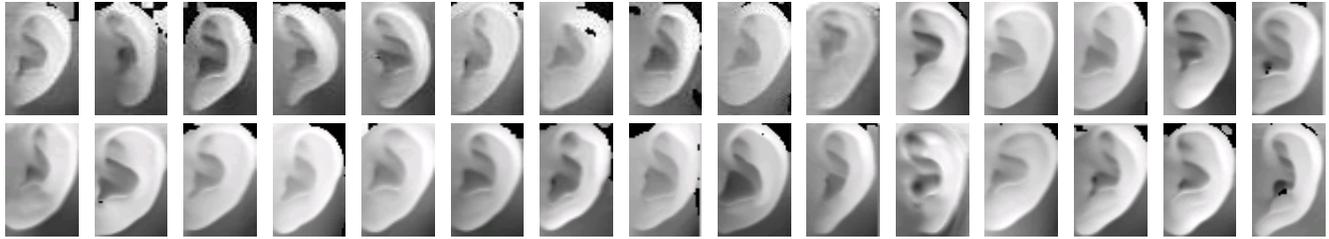
Figure 6. Manually extracted ear images for subjects 0-29.
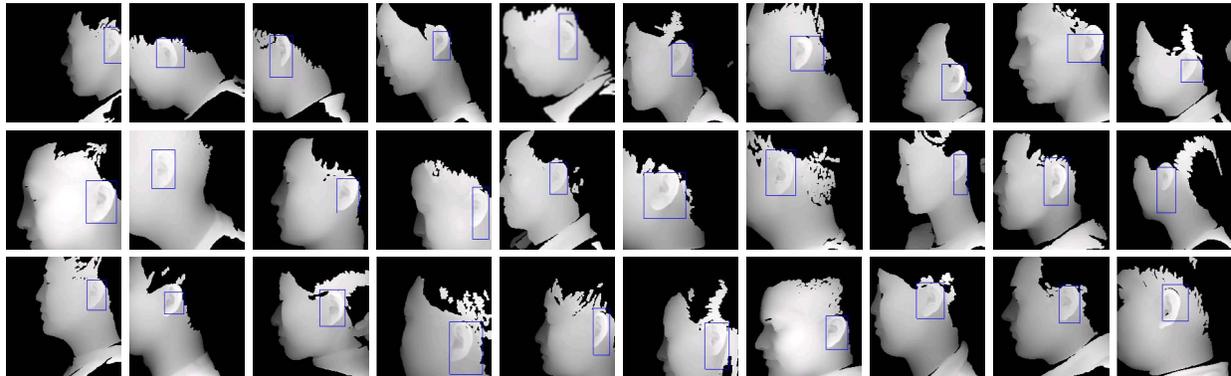
Figure 8. Ear detection results for subjects 0-29.
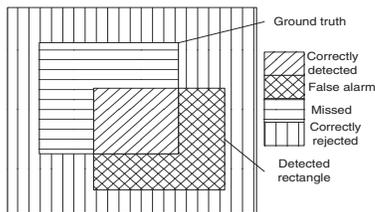
Figure 7. Illustration of calculation of correct detection rate and false alarm rate.

easily implemented. Experiment results with real ear images demonstrate the effectiveness of our approach.

References

[1] B. Bhanu and H. Chen. Human ear recognition in 3D. *Workshop on Multimodal User Authentication*, pages 91–98, 2003.

[2] K. Chang, K. Bowyer, S. Sarkar, and B. Victor. Comparison and combination of ear and face images in appearance-based biometrics. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(9):1160–1165, 2003.

[3] C. Dorai and A. Jain. COSMOS-A representation scheme for free-form surfaces. *Proc. Int. Conf. on Computer Vision*, pages 1024–1029, 1995.

[4] P. Flynn and A. Jain. On reliable curvature estimation. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 110–116, 1989.

[5] A. Iannarelli. *Ear Identification*. Forensic Identification Series. Paramont Publishing Company, 1989.

[6] A. Jain. *BIOMETRICS: Personal Identification in Network Society*. Kluwer Academic, 1999.

[7] J. M. Keller, P. Gader, R. Krishnapuram, and X. Wang. A fuzzy logic automatic target detection system for LADAR range images. *IEEE International Conference on computatinoal intelligence*, 1:71–76, 1998.

[8] E. B. Meier and F. Ade. Object detection and tracking in range images sequences by separation of image features. *IEEE Internation conference on Intelligent Vehicles*, pages 176–181, 1998.

[9] E. Osuna, R. Freund, and F. Girosi. Training support vector machine: An application to face detection. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pages 130–136, 1997.

[10] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(1):23–38, 1998.

[11] B. Schiele and J. Crowley. Recognition without correspondence using multidimensional receptive field histograms. *International Journal of Computer Vision*, 36(1):31–50, 2000.

[12] J. Sparbert, K. Dietmayer, and D. Streller. Lane detection and street type classification using laser range images. *IEEE Intelligent Transportation Systems conference proceedings*, pages 454–459, 2001.

[13] D. Xi and S. W. Lee. Face detection and facial component extraction by wavelet decomposition and support vector machine. *4th international conference, AVBPA*, pages 199–207, 2003.

[14] M. H. Yang, N. Ahuja, and D. Kriegman. Mixtures of linear subspaces for face detection. *Proc. Fourth Int'l Conf. Automatic face and gesture recognition*, pages 70–76, 2000.

[15] N. Yokoya and M. D. Levine. Range image segmentation based on differential geometry: A hybrid approach. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(6):643–649, 1989.