Proceedings of International Conference on Advances in Pattern Recognition, ICAPR' 09,
pp. 213-216, Kolkata, India, Feruary 2009.

A Skin-Color and Template Based Technique for Automatic Ear Detection

Surya Prakash, Umarani Jayaraman and Phalguni Gupta
Department of Computer Science and Engineering,
Indian Institute of Technology Kanpur, Kanpur-208016, India.
E-mail: {psurya,umarani,pg}@cse.iitk.ac.in

Abstract

This paper proposes an efficient skin-color and template based technique for automatic ear detection in a side face image. The technique first separates skin regions from non-skin regions and then searches for the ear within skin regions. Ear detection process involves three major steps. First, Skin Segmentation to eliminate all non-skin pixels from the image, second Ear Localization to perform ear detection using template matching approach, and third Ear Verification to validate the ear detection using the Zernike moments based shape descriptor. To handle the detection of ears of various shapes and sizes, an ear template is created considering the ears of various shapes (triangular, round, oval and rectangular) and resized automatically to a size suitable for the detection. Proposed technique is tested on the IIT Kanpur ear database consisting of 150 *side face images and gives 94% accuracy.*

Keywords: Biometrics, skin-segmentation, ear detection, Zernike moments, shape descriptor.

1 INTRODUCTION

Recent research has shown ear as a good and reliable biometrics for human verification and identification [4]. To automate ear based recognition process, it is necessary to detect the ear automatically but at the same time detection of ear from an arbitrary side face image is one of the 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 [6, 9, 8, 1, 2, 14]. Burge and Burger [6] have used deformable contours for ear detection. Hurley et al. [9] uses force field approach but it is only applicable in presence of small background. Chen and Bhanu [8] have presented a template based method for ear detection from side face range images. This method works for 3D ear biometrics. Alvarez et al. [1] have proposed an ear localization method from 2D face image using ovoid and active contour model. This method requires an initial approximate ear contour as input and can't be used in fully automated ear recognition process. Ansari and Gupta [2] have presented an ear detection approach based on edges of outer ear helices. Since this technique solely relies on the parallelism between the outer helix curves, it fails if the helix edges are not proper. Yuan and Mu [14] have proposed an ear detection method based on skin-color and contour information. They assume the ear shape elliptical and fit ellipse to the edges to get the accurate position of the ear. The assumption of considering shape of the ear elliptical is not true in general and does not help is detecting the ear in all cases. In [11], Sana et al. have proposed an ear detection scheme based on wavelet based templates. In real scenario, ear occurs in various sizes and the preestimated templates are not sufficient to handle all the situations, and automatic resizing need to be done.

This paper presents an efficient novel technique for automatic ear detection in side face images using a skin-color model and template matching. In the proposed technique, an off-line created ear template is automatically resized to an appropriate size, to detect ears of different size and shape. Detection is validated by employing Zernike moment based shape descriptor.

2 PRELIMINARIES

Following subsections discuss some basic techniques which are required in developing the proposed ear detection model.

2.1 Color Based Skin Segmentation

This section briefly discusses a color based skin segmentation model [7]. Since chromatic color representation of color images is more suitable for characterizing skin-color, this technique first converts pixels from *RGB* color space to chromatic color space [13]. Chromatic color space is defined by a normalization process as $r = R/(R+G+B)$ and $b = B/(R+G+B)$ where green color is redundant as $r + g + b = 1$. 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 $\mu = E[x]$, covariance $C = E[(x-\mu)(x-\mu)^T]$, and $x = (r, b)^T$. $E[\phi]$ 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

(a) (b) (c)

Figure 1. Skin segmentation: (a) Input Image, (b) Skin-likelihood Values, (c) Skin-segmented Image

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

Figure 2. Shapes of the ear: (a) Triangular, (b) Round, (c) Oval and (d) Rectangular

RGB to chromatic color space, have a chromatic pair value of (r, b), the likelihood $P(r, b)$ of skin for this pixel can be computed by Eqn. 1. These likelihood values can be used for skin segmentation.

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

2.2 Zernike Moments

Among the family of moment invariants, Zernike moments [10] are one of the most commonly used feature extractor and have been used in variety of applications. In the proposed technique, Zernike moments based shape features are used for ear verification. Popularity of the Zernike moments stems from the fact that they are robust in the presence of noise and exhibit rotational invariant property inherited from the angular dependence of Zernike polynomials. Also, Zernike moments provide non-redundant shape representation because of their orthogonal basis.

3 PROPOSED TECHNIQUE

Proposed ear detection technique involves three steps namely skin segmentation, ear localization and ear verification. In the first step, skin segmentation is performed to eliminate all non-skin pixels from the image. Second step employs an off-line created template to detect ears. Third step is used to verify the detections. Following subsections present the details of these steps.

3.1 Skin Segmentation

The first step of the proposed technique is skin segmentation which aims to detect skin regions in an image to reduce the search space for the ear. Skin color model presented in Section 2.1 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 likelihood of the pixel belonging to the skin. With an appropriate thresholding, the grayscale image is further transformed to a binary image showing skin and non-skin regions. A sample color image and its resulting skin-likelihood image are shown in Figure 1(a) and Figure 1(b) respectively. All skin regions in Figure 1(b) are shown brighter than the non-skin region. Since people with different skins have different likelihood, an adaptive thresholding [7] process is used to achieve the optimal threshold value for each run. Figure 1(c) shows a skin segmented image (in grayscale) for the image shown in Figure 1(a). Since not all detected skin regions contain ear, a localization need to be performed to locate the ear.

3.2 Ear Localization

Three steps involve in ear localization process are discussed in following subsections.

3.2.1 Ear Template Creation

For any template based approach, it is very much necessary to obtain a template which is a good representative of the data. In the proposed technique, ear template is created by averaging the intensities of a set of ear images. Human ear shape can broadly be categorized into four classes: triangular, round, oval, and rectangular (see Figure 2). For creation of ear template, all types of the ear are considered to obtain a good representative template. The ear template T is formally defined as follows:

$$T(i,j) = \frac{1}{N}\sum_{k=1}^{N} E_k(i,j) \quad (2)$$

where N is the number of ear images used for ear template creation and E_k is the k^{th} ear image. $E_k(i,j)$ and $T(i,j)$ represent the pixel values of the $(i,j)^{th}$ pixel of E_k and T respectively. In our experiments, we considered $N = 20$.

3.2.2 Resizing Ear Template

Template resizing is an important step. We know that ear in a side face image may occur in different sizes depending

on the distance of the camera from the person. To handle the detection of ears of various sizes, ear template need to be resized to make it appropriate for the detection of ear in the image. In existing template based approaches [8, 11], predefined size of ear templates are used for localization which are very often not able to detect the ears of different size. An experiment has been performed by us and size of the side face image and the ear contained in it have been measured for various images taken from various distances. It is observed that the size of the ear is proportional to the size of the side face image. This observation is used for resizing the ear template in the proposed technique. Size of the side face is estimated using the bounding box of the side face skin regions. Resizing proceeds as follows. A reference side face image is considered whose ear size is same as the standard ear template size (120×80 pixels in our case). To get a suitable resized ear template for an input image, first its skin segmentation is done. Segmented image is then used to get the size of the side face. Let, w_i^f and w_r^f be the widths of the input face image and the reference face image respectively. If width of the ear of reference image is w_r^e, width of the resized ear template for given input image is given as:

$$w_i^e = \frac{w_r^e}{w_r^f} \times w_i^f \tag{3}$$

By keeping the aspect ratio of the ear template same, it is resized to the width obtained in Eqn. 3. Resized template is employed in searching of the ear in the input image. Resizing is done using the width of the side face image because it is less affected by the false face pixels (pixels which are segmented as face pixels but actually they are not, *e.g.* pixels of neck region etc.). Further, height of the side face is difficult to measure as it is often inaccurate because of the inclusion of the skin pixels of neck region in the side face image.

3.2.3 Localization

Once the ear template of suitable size is created, localization of the ear is carried out. To search an ear in the image I, ear template T is moved over the image and normalized cross-correlation coefficient (NCC)[3] is computed at every pixel. NCC at point (x, y) is defined as follows:

$$NCC(x,y) = \frac{\sum_{u,v}[I(u,v) - \bar{I}_{x,y}][T(u-x, v-y) - \bar{T}]}{\sqrt{\sum_{u,v}[I(u,v) - \bar{I}_{x,y}]^2 \sum_{u,v}[T(u-x, v-y) - \bar{T}]^2}} \tag{4}$$

where sum is performed over u, v under the window containing T positioned at (x, y). $\bar{I}_{x,y}$ and $\bar{T}$ are the average of brightness values of the portion of the target image under the template and template image respectively. Values of NCC lie between -1.0 and 1.0. When it is typically above a preestimated threshold, we accept the hypothesis that an ear exists in the region. Otherwise, it is reject. Value of NCC closer to 1 indicates a better match. We prefer NCC over cross-correlation in template search as NCC 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. This step accepts all the points having NCC value greater than the threshold as probable locations of the ear. These locations are sorted in a non-increasing order of their NCC values.

3.3 Ear Verification

To determine whether a detected ear is a true ear or not, shape based ear verification is performed. Since small set of lower order Zernike moments can characterize the global shape of an object effectively [12], magnitudes of these moments are used for ear shape representation. Similarity between the two sets of Zernike moments (one for template and another for detected ear) is estimated and used to validate the claim. To measure the similarity, Euclidian distance between the two sets of Zernike moments is used, and estimated as follows:

$$distance = \sqrt{\sum_{i=1}^{L}(|M_i^T| - |M_i^E|)^2} \tag{5}$$

where $\{M_i^T\}_{i=1}^L$ and $\{M_i^E\}_{i=1}^L$ are the L Zernike moments used to represent the shape of ear template and detected ear respectively.

Probable ear locations found in the previous section are considered one by one in the non-increasing order of their NCC values. Edge images of the ear template and the detected ear image are obtained using Canny edge detector. Lower order Zernike moments of the Distance transforms [5] of the edge images of template and the detected ear are estimated and the similarity distance between them is calculated using Eqn. 5. To get the detected ear image at a point (x, y), template sized image is cropped from the input side face image keeping the point (x, y) at the center of the template. If the value of $distance$ is less than a preestimated threshold, detection is accepted otherwise it is rejected.

4 EXPERIMENTAL RESULTS

Experimentation is performed on IIT Kanpur ear database which contains total 150 side images of human face with resolution 640×480 pixels. These images are captured using a digital camera from a distance of 0.5 to 1 meter. To create ear template, a set of side face images of 20 people is considered. For ear detection, first skin segmentation is performed to separates skin regions from non-skin

Figure 3. Detected ears

regions and then ear is localized using template matching approach. After ear localization, a verification procedure based on Zernike moment based shape descriptor is called to check whether detected ear is a true ear or not. NCC is used to localize the ear. Points having NCC values above a preestimated threshold (in our experiment 0.5) are declared as the probable locations of the ear. These claims are verified using the Zernike moments based shape descriptor. To measure the similarity between the detected ear and the template, Euclidian distance between the two sets of Zernike moments (one for template and one for detected ear) is calculated. If similarity value lies below a preestimated threshold (in our experiment 0.3), the claim is declared valid.

Figure 3 shows some of the results of the proposed technique where ears of different size and shape are accurately detected from the side face images. The proposed technique is also able to detect ear in presence of little occlusion due to hair. Last two images in the last row of Figure 3 show such examples. Accuracy of the localization is defined by: $(genuine_localization/total_sample) \times 100$. It is found to be 94% for the proposed technique. Localization method has failed in some cases, especially for the images which are of poor quality or heavily occluded due to hair.

5 CONCLUSION

This paper first draws attention towards the inability of the existing ear localization techniques for automatic ear detection in 2D side face images and then proposes an efficient skin-segmentation and template matching based technique for the same. To detect ears, proposed technique follows three steps. (1) Skin segmentation to detect skin pixels, (2) ear localization to detect ears, and (3) ear verification to validate the detections. Since proposed techniques employs ears of various shapes for template estimation and performs automatic template resizing, it is able to detect ears of different shapes and sizes automatically without any user interaction and can be employed in an automatic ear based biometric systems. Ear detection in the proposed technique is fast as it prunes almost 60% area of the side face image and searches for ear only in the skin regions. The Proposed technique is tested on IIT Kanpur ear database containing 150 side face images and found to be giving 94% accuracy. Possible extensions of the presented work could be the detection of ear in noisy images and in the cases where ear is immensely occluded by the hair.

References

[1] L. Alvarez, E. Gonzalez, and L. Mazorra. Fitting ear contour using an ovoid model. In *Proc. of ICCST, 2005*, pages 145–148, 2005.

[2] S. Ansari and P. Gupta. Localization of ear using outer helix curve of the ear. In *Proc. of ICCTA, 2007*, pages 688–692, 2007.

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

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

[5] H. Breu, J. Gil, D. Kirkpatrick, and M. Werman. Linear time euclidean distance algorithms. *IEEE Trans. on PAMI*, 17(5):529–533, 1995.

[6] M. Burge and W. Burger. Ear biometrics in computer vision. In *Proc. of ICPR, 2000*, volume 2, pages 822–826, 2000.

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

[8] H. Chen and B. Bhanu. Human ear detection from side face range images. In *Proc. of ICPR, 2004*, volume 3, pages 574–577, 2004.

[9] D. J. Hurley, M. S. Nixon, and J. N. Carter. Automatic ear recognition by force field transformations. In *IEE Colloquium: Visual Biometrics*, pages 8/1–8/5, 2000.

[10] R. J. Prokop and A. P. Reeves. A survey of moment-based techniques for unoccluded object representation and recognition. *CVGIP*, 54(5):438–460, 1992.

[11] A. Sana, P. Gupta, and R. Purkait. Ear biometric: A new approach. In *Proc. of ICAPR, 2007*, pages 46–50, 2007.

[12] C.-H. Teh and R. T. Chin. On image analysis by the methods of moments. *IEEE Trans. on PAMI*, 10(4):496–513, 1988.

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

[14] L. Yuan and Z.-C. Mu. Ear detection based on skin-color and contour information. In *Proc. of ICMLC, 2007*, volume 4, pages 2213–2217, 2007.