==========
Engineering & Technology in India www.engineeringandtechnologyinindia.com
Vol. 1:2 March 2016
==========

Person Recognition Using Ear Images

R. Revathi and M. Bhavani

==========

Abstract

We present a complete three-dimensional (3-D) ear recognition system combining local and holistic features in a computationally efficient manner. The system is comprised of four primary components, namely: 1) ear image segmentation; 2) local feature extraction and matching; 3) holistic feature extraction and matching; and 4) a fusion framework combining local and holistic features at the match score level. For the segmentation component, we introduce a novel shape-based feature set, termed the Histograms of Indexed Shapes (HIS), to localize a rectangular region containing the ear. For the local feature extraction and representation component, we extend the HIS feature descriptor to an object-centered 3-D shape descriptor, the Surface Patch Histogram of Indexed Shapes (SPHIS), for local ear surface representation and matching. For the holistic matching component, we introduce a voxelization scheme for holistic ear representation from which an efficient, voxel-wise comparison of gallery-probe model pairs can be made. The match scores obtained from both the local and holistic matching components are fused to generate the final match scores.

Keywords: Histograms of Indexed Shapes, Surface Patch Histogram of Indexed Shapes

INTRODUCTION

BIOMETRICS deals with recognition of individuals based on their physiological or behavioral characteristics. Researchers have done extensive studies on biometrics such as fingerprint, face, palm print, iris, and gait. Ear, a viable new class of biometrics, has certain advantages over face and fingerprint, which are the two most common biometrics in both academic research and industrial applications. For example, the ear is rich in features; it is a stable structure that does not change much with age and it does not change its shape with facial expressions. Furthermore, ear is larger in size compared to fingerprints but smaller as compared to face and it can be easily captured from a distance without a fully cooperative

subject although it can sometimes be hidden with hair, cap, turban, muffler, scarf, and earrings. The anatomical structure of the human ear is shown in Fig. 1.

The ear is made up of standard features like the face. These include the outer rim (helix) and ridges (antihelix) parallel to the helix, the lobe, the concha (hollow part of ear), and the tragus (the small prominence of cartilage over the meatus). In this paper, we use the helix/antihelix for ear recognition.

In this paper, we present a fully automatic, 3-D ear recognition system that combines both local and holistic features in a computationally efficient manner.

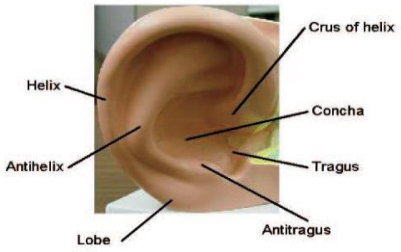

Fig. 1. The external ear and its anatomical parts

The motivation behind combining local and holistic surface features for 3-D ear recognition is that local representations have been found to be more robust to clutter and small amounts of noise, while the holistic representation captures information from the entire surface without excluding any information when describing the ear. When combined effectively, they can provide complementary information describing the 3-D ear shape and jointly enhance the matching performance.

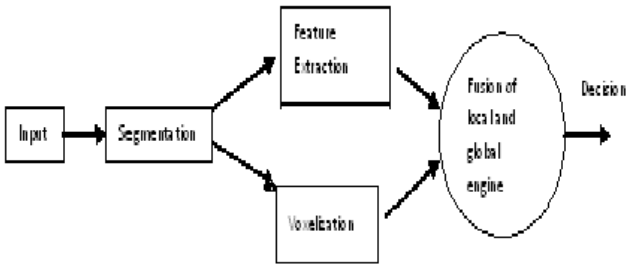

Fig. 2. System overview.

THREE-DIMENSIONAL EAR DETECTION

An overview of our ear detection procedure using the sliding window approach is shown in Fig. 3. To locate the ear in a range image, the image is scanned from the top left corner to the bottom right corner with a fixed-sized detection window. In each window position the HIS feature vector is extracted and employed to train a binary ear/non-ear classifier. The classifier is then used to determine whether the current window contains an ear (positive sample) or non-ear (negative sample) by evaluating the classification score based on the feature vector extracted from the window. SVM is one of the leading techniques used for binary classification due to its robust performance and efficiency in both the training and testing stages.

The scanning window approach employing a fixed-sized window, however, only allows for ear detection at a single scale. To achieve multi-scale detection, the image is gradually resized using a scale factor, and the detection window is then iteratively applied on each of the resized images. After scanning the detection window across the image at multiple scales, multiple detections usually occur around the target regions and it is useful to fuse overlapping detected windows into a single detection [11]. We select a nonmaximal suppression (NMS) method proposed by Dalal [12] as the solution to the fusion of overlapping detected windows, in which each detection is mapped to a respective point in 3-D space (position and scale of the scanning window) weighted by their confidence scores. A nonparametric density estimator is employed to estimate the corresponding density function, where the resulting peaks of the density function constitute the final detections, with positions, scales and confidence scores given by the positions of the peaks. After nonmaximal suppression, the detection system returns a bounding box with an associated detection score representing the ear region.

A. HIS Feature Descriptor

Objects can be characterized by their distinct 3-D surface shapes. The human ear, for instance, contains areas around the helix ring and antihelix that possess both prominent saddle and ridge shapes, while the inner ear regions have rut and trough shapes.

1) Shape Index: A quantitative measure of the shape of a surface at a vertex, called the shape index S_I , is defined as

$$S_I(\mathbf{p}) = \frac{1}{2} - \frac{1}{\pi} \arctan\left(\frac{k_{\max}(\mathbf{p}) + k_{\min}(\mathbf{p})}{k_{\max}(\mathbf{p}) - k_{\min}(\mathbf{p})}\right) \quad (1)$$

where K_{max},K_{min} are the principal curvatures of the surface at vertex , with $K_{max} > K_{min}$ defined as

$$k_{\max}(\mathbf{p}) = H(\mathbf{p}) + \sqrt{H^2(\mathbf{p}) - K(\mathbf{p})} \quad (2)$$
$$k_{\min}(\mathbf{p}) = H(\mathbf{p}) - \sqrt{H^2(\mathbf{p}) - K(\mathbf{p})}. \quad (3)$$

$$K = \frac{(eg - f^2)}{(EG - F^2)} \quad (4)$$
$$H = \frac{(eG - 2fF + gE)}{(2\{EG - F^2\})} \quad (5)$$

$$E = \|\mathbf{x}_u\|^2, \quad F = \mathbf{x}_u\mathbf{x}_v, \quad G = \|\mathbf{x}_v\|^2 \quad (6)$$
$$e = \frac{\det(\mathbf{x}_{uu}\mathbf{x}_u\mathbf{x}_v)}{\sqrt{EG - F^2}}, \quad f = \frac{\det(\mathbf{x}_{uv}\mathbf{x}_u\mathbf{x}_v)}{\sqrt{EG - F^2}}$$
$$g = \frac{\det(\mathbf{x}_{vv}\mathbf{x}_u\mathbf{x}_v)}{\sqrt{EG - F^2}}. \quad (7)$$

2) Curvedness: The shape index of a rigid object is not only independent of its position and orientation in space, but also independent of its scale. To encode the scale information, we utilize the curvedness, which is also known as the bending energy, to capture the scale differences [13]. Mathematically, the curvedness of a surface at a vertex is defined as

$$C_v(\mathbf{p}) = \sqrt{\frac{k_{\max}^2(\mathbf{p}) + k_{\min}^2(\mathbf{p})}{2}}. \quad (8)$$

3) HIS Descriptor: The HIS descriptor is defined using the shape index and curvedness values calculated from the vertices contained within the surface region to be encoded. To build the histogram descriptor, first the curvedness and shape index values are collected at each vertex over the surface region. Each vertex contributes a weighted vote for a histogram bin based on its shape index value, with a strength that depends on its curvedness. The votes

of all vertices are then accumulated into the evenly spaced shape index bins forming the HIS descriptor encoding the shape information over the surface region. To avoid boundary effects, linear interpolation is used to distribute each curvedness value into adjacent shape index histogram bins. Let x and c be the shape index and curvedness values of a 3-D vertex on the surface region which contribute a weighted vote to the HIS histogram,x_1 and $x_2$be the centers of the two nearest neighboring bins of such that ,$x_1 \leq x \leq x_2$ and be a HIS histogram with bins. The linear interpolation method that distributes the vertex's curvedness into the two nearest neighboring bins is defined as follows:

$$h(x_1) = h(x_1) + c\left(1 - \frac{x - x_1}{b}\right) \qquad (9)$$

$$h(x_2) = h(x_2) + c\left(\frac{x - x_1}{b}\right) \qquad (10)$$

B. Feature Encoding for Detection Window

1) Block-Structured HIS: In our detection system using the sliding window approach, the detection window is tiled with blocks of various sizes from which the HIS feature descriptors are built and combined. Fig. 3 explains our feature encoding procedure for the detection window with tiled blocks. For an image window of size 96x64, we construct a feature descriptor based on the histograms of blocks of size 32x32, 16x16, 16x32, and 32x16 overlayed on the image window with an overlap of half a block size. Feature descriptors derived from larger blocks capture holistic details of the input image while features derived from smaller blocks capture finer shape information.

2) Integral HIS*:* In order to efficiently compute the HIS features used by our detector, we make use of the integral histogram method suggested by Porikli [16] for computing histograms over arbitrary rectangular image regions and devise a way for the fast evaluation of HIS features on the blocks. In order to extend the framework of the integral histogram to our detector, we generate for each vertex an 8-D HIS calculated from its own shape index and curvedness values using the method described in Section III-A3. The histogram will consist of only two bins with nonzero values; these are the two adjacent bins that are closest in distance to the shape index of the vertex. The spatial positions of the vertices are then used to propagate an aggregated function of the integral histogram, starting from a point of origin, e.g., top left corner, and traverse through the remaining points. We iterate the integral HIS at

the current vertex using the histograms of the previously processed neighboring points. At each iteration, the values of the integral HIS bins are incremented by the values of the corresponding HIS bins of the current vertex. After the integral HIS is obtained for each vertex, the HIS histogram of a rectangular block can be computed in a constant computational time of an 8-D histogram vector addition and two 8-D histogram vector subtractions, accessing only the integral histogram values at the corner points of those blocks without reconstructing a separate histogram for every block.

LOCAL FEATURE EXTRACTION

A. 3-D Keypoint Detection

To generate the set of local features, the input image is initially searched to identify potential keypoints that are both robust to the presence of image variations and highly distinctive, allowing for correct matching. The keypoint detection method proposed here is inspired by the 3-D face matching approach presented by Mian, *et al.* in [9], but with major enhancements tailored towards improved robustness and applicability to objects with salient curvature, such as the ear. In [9], the input point cloud of the range image is sampled at uniform intervals. By observing 3-D ear images, we found that the majority of these salient points are located in surface regions containing large curvedness values. This signifies that sampling in regions containing large curvedness values will result in a higher probability of obtaining repeatable keypoints. Once a candidate keypoint has been located, a local surface patch surrounding the candidate keypoint is cropped from the ear image using a sphere centered at the candidate keypoint. The purpose of examining its nearby surface data is to further reject candidate keypoints that are less discriminative or less stable due to their location in noisy data or along the image boundary.

B. Local Surface Matching Engine

In our local feature representation, a 3-D ear surface is described by a sparse set of keypoints, and associated with each keypoint is a descriptive SPHIS feature descriptor that encodes the local surface information in an object-centered coordinate system. The objective of the local feature matching engine is to match these individual keypoints in order to match the entire surface. To allow for efficient matching between gallery and probe models, all

gallery images are first processed. The extracted keypoints and their respective SPHIS feature descriptors are stored in the gallery. Each feature represents the local surface information in a manner that is invariant to surface transformation. A typical 3-D ear image will produce approximately 100 overlapping features at a wide range of positions that form a redundant representation of the original surface. In the local feature matching stage, given a probe image, a set of keypoints and their respective SPHIS descriptors are extracted using the same parameters as those used in the feature extraction of the gallery images. If the cropped surface data contains boundary points, the candidate keypoint is rejected automatically as being close to the image boundary. Otherwise, PCA is applied to the cropped surface data, and the eigenvalues and eigenvectors are computed to evaluate its discriminative potential.

1) Surface Patch Histogram of Indexed Shape (SPHIS) Descriptor:

As mentioned in Section III-A3, the HIS descriptor can be used to encode shape information of any surface region. In addition, we can form an HIS of arbitrary size by uniformly spacing the shape index values over the range [0, 1]. The larger the dimensionality of the HIS, the more descriptive it is. However, too large of a descriptor may be sensitive to noise. Based on the HIS descriptor, the SPHIS descriptor is employed to represent the keypoint and is built from the surface patch surrounding it.

IV. HOLISTIC FEATURE EXTRACTION

A. Preprocessing

For a gallery model, the ear surface output from the detection component is normalized to a standard pose. The centroid of the surface is firstly mapped to the origin of the coordinate system. Then, the principal components corresponding to the two largest eigenvalues of the surface are calculated. The surface is then rotated such that the two principal components are aligned with the axes of the coordinate system. The preceding section described the method by which to establish correspondences between a probe-gallery pair. The probe model is then registered onto the gallery model by applying the transformation obtained by the local matching stage for each point on the probe model. In the event that the number of established correspondences is below three, we rely on the pose normalization scheme for the model registration.

The local and holistic matching components result in independent similarity matrices S_i each of size P x G , where I Є {1,2} denotes the matching engine and represents the number of probe and gallery models, respectively. We fuse the local and holistic match scores using the weighted sum technique. This approach is in the category of transform-based techniques (i.e., based on the classification presented in [18]. In practical multimatcher biometric systems, a common fusion method is to directly combine the match scores provided by different matchers without converting them into posteriori probabilities. However, the combination of the match scores is meaningful only when the scores of the individual matchers are comparable. This requires a change of the location and scale parameters of the match score distributions at the outputs of the individual matchers. Hence, the *sigmoid function* score normalization [19], which is proven to be an efficient and robust technique in [18], is used to transform the match scores obtained from the different matchers into a common domain. It is defined as follows:

$$s_j^n = \begin{cases} \frac{1}{1+exp\left(-2\left(\frac{s_j-\tau}{\alpha_1}\right)\right)} & s_j < \tau, \\ \frac{1}{1+exp\left(-2\left(\frac{s_j-\tau}{\alpha_2}\right)\right)} & \text{otherwise} \end{cases} \qquad (11)$$

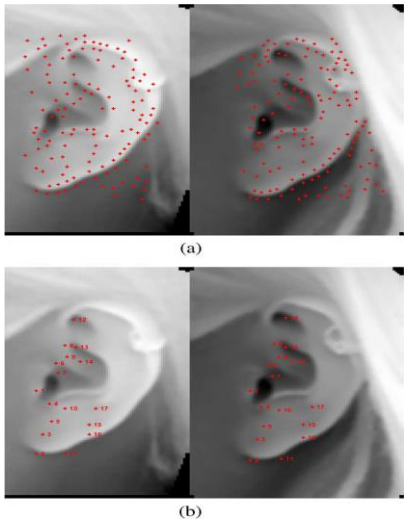

Fig. 3. Example of finding feature correspondences for a pair of gallery and probe ears from same subject. (a) Key points detected on ears. (b) True feature correspondences recovered by local surface matching engine.

V. CONCLUSION

We have presented a complete, automatic 3-D ear biometric system using range images. Within the system, a novel 3-D shape descriptor, the Histogram of Indexed Shapes, is proposed to robustly encode 3-D ear shape and is used in 3-D ear detection and recognition tasks. The proposed 3-D ear surface matching approach employs both local and holistic 3-D ear shape features. The experimental results demonstrate the accuracy and efficiency of our novel 3-D ear shape matching approach. The proposed system achieves a rank-one recognition rate of 98.3% and an EER of 1.7% on a 415-subject dataset with an average time lapse of 17.7 weeks between successive acquisitions of a subject. Moreover, the proposed approach achieves a significantly higher computational efficiency than the SOA systems.

==========

References

[1] H. Chen and B. Bhanu, "Human ear recognition from side face range images," in *Proc. Int. Conf. Pattern Recognition*, Aug. 2004, pp.574–577.

[2] H. Chen and B. Bhanu, "Contour matching for 3-D ear recognition," in *Proc. IEEE Workshop Applications Computer Vision*, Jan. 2005, pp.123–128.

[3] H. Chen and B. Bhanu, "Human ear recognition in 3D," *IEEE Trans.Pattern Anal. Machine Intell.*, vol. 29, no. 4, pp. 718–737, Apr. 2007.

[4] P. Yan and K. Bowyer, "Empirical evaluation of advanced ear biometrics," in *Proc. CVPR 2005—Workshops*, Jan. 2005, pp. 41–41.

[5] P. Yan and K. Bowyer, "Biometric recognition using 3D ear shape," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 29, no. 8, pp.1297–1308, Aug. 2007.

[6] T. Theoharis, G. Passalis, G. Toderici, and I. Kakadiaris, "Unified 3Dface and ear recognition using wavelets on geometry images," *Pattern Recognition*, vol. 41, no. 3, pp. 796–804, Mar. 2008.

[7] I. Kakadiaris,G.Passalis, G. Toderici, N.Murtuza, Y. Lu, K. N., and T. Theoharis, "3D face recognition in the presence of facial expressions:An annotated deformable model approach," *IEEE Trans. Pattern Anal.Machine Intell.*, vol. 29, no. 4, pp. 640–649, Apr. 2007.

[8] S. Islam, R. Davies, M. Bennamoun, and A. Mian, "Efficient detection and recognition of 3D ears," *Int. J. Computer Vision*, vol. 95, no. 1, pp.52–73, 2011.

[9] A. Mian, M. Bennamoun, and R. Owens, "Keypoint detection and local feature matching for textured 3D face recognition," *Int. J. Computer Vision*, vol. 79, no. 1, pp. 1–12, 2008.

[10] S. Cadavid and M. Abdel-Mottaleb, "3-D ear modeling and recognition from video sequences using shape from shading," *IEEE Trans. Inform.Forensics Security*, vol. 3, no. 4, pp. 709–718, Dec. 2008.

[11] P. Viola and M. Jones, "Robust real-time object detection," *Int. J. Computer Vision*, vol. 57, no. 2, pp. 137–154, 2001.

[12] N. Dalal, "Finding People in Images and Videos," Ph.D. dissertation, Institut National Polytechnique de Grenoble, Grenoble, France, Jul.2006.

[13] C. Dorai and A. Jain, "Cosmos—A representation scheme for 3D freeform objects," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no.10, pp. 1115–1130, Oct. 1997.

[14] A. Gray, *Modern Differential Geometry of Curves and Surfaces With Mathematica*, 2nd ed. Boca Raton, FL: CRC, 1997.

[15] P. Besl, "Surface in range image understanding," in *Springer Series in Perception Engineering*. New York: Springer-Verlag, 1988.

[16] F. Porikli, "A fast way to extract histograms in Cartesian spaces," in *Computer Vision Pattern Recognition (CVPR)*, 2005.

[17] S. Wang and A. Kaufman, "Volume sampled voxelization of geometric primitives," in *Proc. 4th Conf. Visualization*, 1993, pp. 78–84.

[18] A. Ross, K. Nandakumar, and A. Jain, *Handbook of Multibiometrics (International Series on Biometrics)*. Secaucus, NJ: Springer-Verlag, 2006.

[19] R. Cappelli, D. Maio, and D. Maltoni, "Combining fingerprint classifiers," in *Proc. Fist Int. Workshop Multiple Classifier Systems*, 2000,pp. 351–361.

[20] J. Zhou, S. Cadavid, and M. Abdel-Mottaleb, "Histograms of categorized shapes for 3D ear detection," in *Proc. IEEE Fourth Int. Conf.sep2010.*

==========

R. Revathi
Assistant Professor in Computer Science & Engineering
Theni Kammavar Sangam College of Technology
Theni 625 534

Tamilnadu
India
revavyas@gmail.com

M. Bhavani
Assistant Professor in Computer Science & Engineering
Theni Kammavar Sangam College of Technology
Theni 625 534
Tamilnadu
India
gmbhavani1990@gmail.com\