A NOVEL 3D EAR IDENTIFICATION APPROACH BASED ON SPARSE REPRESENTATION

Zhixuan Ding, *Lin Zhang**[1], *and Hongyu Li*
School of Software Engineering, Tongji University, Shanghai, China

ABSTRACT

Recently, ear shape has attracted tremendous interests in biometric research due to its richness of feature and ease of acquisition. In this paper, we present a novel 3D ear identification approach based on the sparse representation framework. To this end, at first, we propose a template-based ear detection method. By utilizing such a method, the extracted ear regions are represented in a common standard coordinate system determined by the template, which facilitates the following feature extraction and classification. For each 3D ear, a feature vector can be generated as its representation. With respect to the ear identification, we resort to the l_1-minimization based sparse representation. Experiments conducted on a benchmark dataset corroborate the effectiveness and efficacy of the proposed approach. The associated Matlab source code and the evaluation results have been made online available at *http://sse.tongji.edu.cn/linzhang/ear/srcear/srcear.htm*.

Index Terms—Biometrics, 3D ear recognition, sparse representation, Iterative Closest Point

1. INTRODUCTION

Among all the biometric identifiers, ear is a relative new member and it has been proved viable for its desirable properties such as universality, uniqueness and permanence [1, 2]. Besides the traditional 2D ear recognition [2, 3, 4, 5], there also exists a technology to acquire ear data by using a 3D sensor which provides both 2D and 3D data for an ear. Compared with 2D data, 3D ear data contains more information about ear shape and is not sensitive to illumination and occlusion.

Recently, several researches for 3D ear recognition have been conducted. In [6], Chen and Bhanu developed a 3D ear recognition system. The algorithm they suggested is based on a 2-step Iterative Closet Points (ICP) [7] framework and all the ear regions are extracted from profile images manually. To make their system automatic, they presented an ear detection algorithm by using an ear shape model in their later work [8], in which the first coarse ICP step is performed on two extracted ear helixes and the second fine ICP step is further applied on two corresponding ear point clouds by setting the result gained from last step as initial translation. In 2007, Yan and Bowyer also proposed an automatic ear recognition approach by applying a 3D ICP algorithm [9]. In their work, they tried to locate the ear pit and then used active contour algorithm [10] to extract the ear contour. Their recognition process is no different from any other ICP based approaches. At the same year, Chen and Bhanu improved their work by introducing a Local Surface Patch Representation [11]. 3D ear recognition was also investigated by Islam and Mian [12]. For ear detection, they adapted the Viola-Jones object detection algorithm [13] and for feature extraction, they adopted the Local 3D Feature (L3DF) scheme proposed in [14].

In another aspect, as an effective classification tool [15, 16], sparse representation has also been introduced to the 3D biometrics field. For instance, in [17], Li and Jia proposed a 3D face recognition approach based on sparse representation and promising results were reported.

From the aforementioned introduction, it can be seen that most existing 3D ear or 3D face recognition methods are based on ICP. While ICP is a feasible 3D matching model for the one-to-one verification, it is not quite appropriate for the one-to-many identification case. If there are multiple samples for each subject in the gallery set, the recognition based on ICP usually would have to match the test sample to all the gallery samples. With the number of gallery samples rising, the performance of ICP-based methods will markedly slow down. Since the task of recognition is essentially to find a single individual out of the entire dataset, Wright *et al.* [16] proved that the recognition based on sparse representation framework is more suitable to solve the multiple samples case efficiently. Li and Jia have also shown the feasibility of applying the sparse representation framework to 3D face recognition [17]. However, to the best of knowledge, so far there is no work reported to apply sparse representation for 3D ear recognition.

Based on these considerations, in this paper, we propose a novel 3D ear identification approach based on sparse representation. Our approach consists of three components, ear detection, feature extraction, and classification. For ear detection, we propose a template based scheme which is robust to ear pose change. For feature extraction, we adopt an effective PCA-based descriptor proposed in [20]. Our approach takes 3D point cloud as input and no extra color image is required. The performance of the proposed approach is examined on the benchmark dataset and is compared with the ICP based

[1]Corresponding author: cslinzhang@tongji.edu.cn

method. Efficacy and efficiency of our approach are corroborated by the experimental results.

The remainder of this paper is organized as follows. Section 2 describes our proposed algorithm for ear detection and recognition. Section 3 reports the experimental results and section 4 concludes the paper.

2. SPARSE REPRESENTATION BASED EAR RECOGNITION

In this section, we will present our proposed ear recognition approach in detail, which consists of three sub-sections, ear detection, ear feature extraction and ear identification.

2.1. Ear detection

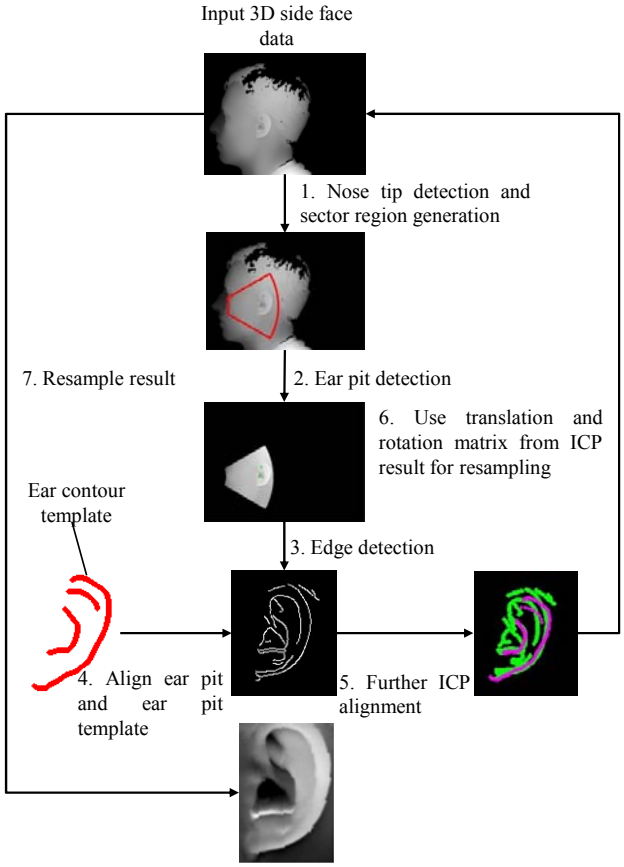


Fig. 1: Illustration for the proposed ear detection scheme.

In our system for each ear, the input 3D ear data is just a 640×480 side face range scan with vertices positions texcoords combined with a binary image utilized to indicate whether a pixel contains a vertex.

2.1.1. Ear pit detection

The binary image is a mask, each pixel of which indicates whether the corresponding pixel in the range scan contains a vertex. Similar as [9], to locate the ear pit, we first need to locate the nose tip which will greatly narrow the ear pit position range in the following process. Identifying the location of the nose tip is accomplished in the image space of the binary mask. Here, we briefly review this process:

1. Record the X values along each row at which we first encounter a white pixel in the binary image and find the mean value X_{mean} of X values.
2. Within a 100 pixel range left of X_{mean}, like step 1, record the starting Y values along each row and find the mean value Y_{mean}. Within a 60 pixel range above and below of the Y_{mean}, the valid point with minimum X value is the nose tip ($X_{NoseTip}$, $Y_{NoseTip}$).
3. Using the point ($X_{NoseTip}$, $Y_{NoseTip}$) as the center of a circle, we generate a sector region spanning +/- 30 degree from the horizontal. In this region, we reject pixels of which the corresponding vertices have a distance larger than 16 cm or smaller than 4 cm to the vertex of nose tip.

Based on the location of the nose tip, we can generate a corresponding sector which the ear pit should fall in. We then use the Z value in the range scan as the intensity for each pixel to generate a range image, and the further locating of ear pit is performed in the image space of the range image. To this end, we introduce a simple yet effective method. We assume that the ear pit should be the point of which the Z value is the lowest within a circle range like a pit. Following this rule, we pre-generate a random sample pattern which consists of 150 points within a 30-pixel radius circle range. And for each point fall in the sector we sample its neighboring points via the pre-generated sample pattern. The points with lowest depth value in its local neighboring are the candidates of ear pit. From these candidates distribution, we found that the true ear pit must lie on a cluster which contains several candidates. So we remove the isolated candidates with no other candidates around, which could be caused by noise or hair interference. Then we regard the candidate with lowest depth value from the remaining as the ear pit.

2.1.2. Ear contour alignment

For handling pose change, we adopt the ICP algorithm to align each ear contour to an ear contour template. The idea is similar to the one suggested in [8]; however, there are some differences. At first, we have detected the position of ear pits so we could reduce the computational cost of ICP contours matching by aligning the ear pits first. Secondly, in our case, ICP matching is performed in the 2D image space, which is much faster than the one working in the 3D space.

We built an ear contour template by manually selecting corresponding points from the subset of UND-J ear database and calculating the average position for each point. The built ear contour template is shown in Fig. 1. For each query ear, we extract the sector region (see 2.1.1) from the depth image. Then we apply a Canny edge detector on the extracted part of the depth image and we get the edge map which can be viewed as the contour part of a query ear. After that, we first align the edge map to the ear contour template by aligning the detected ear pit to the ear pit template. Secondly, an ICP

algorithm was applied in the image space to fulfill the further alignment. Finally by sampling the original input 3D range image using the translation and rotation matrix obtained in ICP alignment, we get the 3D ear region, which actually locates in a common standard coordinate system defined by the ear contour template.

By using the template-based contour alignment, the final extracted 3D ear region is robust to ear pose change and there is little difference for the ears extracted from the same person. Fig. 1 summarizes the proposed ear detection method.

2.2. Ear feature extraction

In order to adopt the sparse representation framework, we need to map the extracted 3D ear region into a feature vector. The widely used features on 3D shape are point curvature, graphic distance, triangle area and so on [17]. But the difficulty is to build a one-to-one correspondence between these 3D-shapes. In our case, however, since we have aligned and scaled all 3D-ear data to a standard template in the ear detection step, the one-to-one correspondences have already been built automatically and the feature extraction could be directly applied to range images.

Instead of using traditional features like curvatures, we adopted a local PCA-based feature introduced by Islam *et al.* in [20]. The extracted 3D-ear could be represented as a point cloud: $E = [x_i, y_i, z_i]^T$, where $i = 1,...,n$. For each point in E, let $L_i = [x_j, y_j, z_j]^T$, where $j = 1,\ldots,n_i$ be the points set in a region cropped by a sphere of radius r centered at this point $p_i = [x_i, y_i, z_i]^T$. Given L_i, we can calculate the mean vector m_i and covariance matrix C_i.

$$m_i = \frac{1}{n_i}\sum_j^{n_i} L_{ij} \tag{1}$$

$$C_i = \frac{1}{n_i}\sum_j^{n_i}\left(L_{ij}L_{ij}^T - m_i m_i^T\right) \tag{2}$$

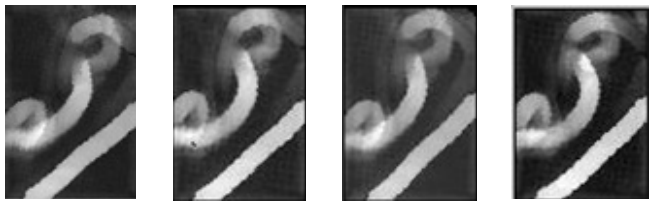
L_{ij} represents the j-th point of L_i. Then a PCA is performed on the covariance matrix C_i and the result yields a matrix V_i of eigenvectors and a diagonal matrix D_i of eigenvalues of C_i

$$C_i V_i = D_i V_i \tag{3}$$

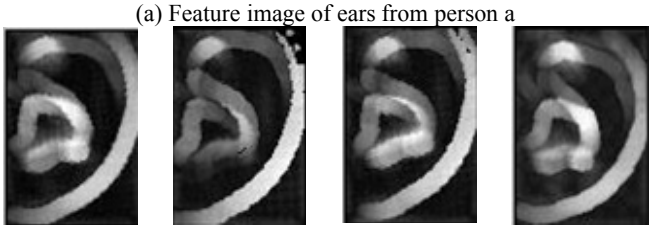
Let $L_{ix}^{'}$ and $L_{iy}^{'}$ be the projections of points L_i on its first and second principal components, respectively (these two principal components correspond to the first two largest eigenvalues). The feature value at point p_i could be defined as

$$\delta_i = \max\left(L_{ix}^{'}\right) - \min\left(L_{ix}^{'}\right) - \max\left(L_{iy}^{'}\right) + \min\left(L_{iy}^{'}\right) \tag{4}$$

According to [20], δ_i can actually represent the difference of first two principal axes at a local region around p_i and it will be *zero* if the point cloud L_i is planar or spherical. By calculating $\{\delta_i\}$ for all the points $\{p_i\}$, we can obtain a feature map.



(a) Feature image of ears from person a



(b) Feature image of ears from person b

Fig. 2: Feature maps calculated for some samples taken from UND-J ear database, with radius $r = 5$ mm.

Fig. 2 shows two examples of our calculated features. From it we can find that the feature mainly describes the ear anatomical information and there is a large similarity for feature images from the same ear. To reduce the feature dimension, we adopt the random projection introduced by Wright *et al.* in [16].

2.3. Ear identification

Given an ear gallery, we calculate a feature vector v for each ear in the gallery and form them to a dictionary matrix $A=[v_{11},..., v_{1k}, v_{21},..., v_{2k}, v_{n1},..., v_{nk}] \in \mathrm{IR}^{m\times n}$, where m here represents the feature dimension, n represents the number of subjects and k represents the number of samples for each subject in gallery.

Given a query sample y, the recognition problem can be viewed as solving the over-completed linear equation:

$$\hat{x}_1 = \arg\min \|x\|_1 \quad \text{subject to} \quad y = Ax \tag{5}$$

With the obtained $\hat{x}_1$, we calculate the residual of each subject i by

$$r_i = \left\| y - \sum_{j=1}^{k} \hat{x}_{1ij} v_{ij} \right\|_2 \tag{6}$$

We choose the index with the minimum residual r_i as the identity of the query sample y.

To better handle the noise and corruption problem, we use an extensional sparse representation by substituting $B = [A, I]$ for the original A where I is identity matrix $\in \mathrm{IR}^{m\times m}$. So the final sparse representation algorithm we adopt is

$$\hat{w}_1 = \arg\min \|w\|_1 \quad \text{subject to} \quad y = Bw \tag{7}$$

And we extract $\hat{x}_1$ by decomposing of $\hat{w}_1 : \hat{w}_1 = [\hat{x}_1, e]$ and the residual calculation should also be substituted with

$$r_i = \left\| y - e - \sum_{j=1}^{k} \hat{x}_{1ij} v_{ij} \right\|_2 \tag{8}$$

We can easily recognize any given query by solving the above l_1-minimization problem in one matching step. The algorithm of l_1-minimization solver we used is DALM [21] which is the fastest algorithm for l_1-minimization problem.

3. EXPERIMENTAL RESULTS

In this section, we will present the evaluation results of the proposed method. The database we used in our experiment is UND collection J dataset [22]. The UND-J dataset is currently the largest 3D ear dataset which consists of 2436 side face 3D scan from 415 different persons. Several samples are shown in Fig. 3.

Fig. 3: Samples of 3D side face range data in UND-J Database.

3.1. Evaluation of the ear detection performance

To validate our ear detection algorithm, we manually marked ear pits for the whole UND-J dataset. In our experiment, we run the experiment on all the 2436 ears of UND-J. For each 3D ear, we compared the automatic detected ear pit location ($X_{\text{auto\_earpit}}$, $Y_{\text{auto\_earpit}}$) with the ground-truth ear pit location ($X_{\text{gt\_earpit}}$, $Y_{\text{gt\_earpit}}$) and if the Euclidean distance of two positions were larger than 16 pixels, we regarded the ear detection for this ear as failure.

Under these experimental settings, our algorithm could achieve a 90.87% detection rate which is much higher than 85% reported in [9] on the same dataset.

3.2. Evaluation of the ear identification performance

Although there are 415 subjects in UND-J database, most subjects have only 2 samples. Since the recognition based on sparse representation needs sufficient samples for each subject [16], we cannot run our experiment on the whole database.

In our experiment, we selected three subsets from UND-J database. The first subset contained 185 subjects of UND-J, each of which had more than 5 samples and the second subset contained 125 subjects, each of which had more than 7 samples and the third subset contained 85 subjects, each of which had more than 10 samples. In subset1, we randomly selected 5 samples from each subject to form the gallery and the rest of ears were formed to the test set. So the gallery size for subset1 was 925, and the test set size was 885. The principle of sample selection for subset2 and subset3 were the same. So the gallery size and the test set size of subset2 were 1209 and 588. For subset3, they were 850 and 291.

We applied our ear recognition algorithm on those three different subsets respectively and we also evaluated the performance of ICP with the same condition. Since there were multiple samples for each subject, we used ICP to match a query ear with all the samples for each subject and selected the minimum matching error to represent the matching error for this subject. Finally we selected the index with the minimum matching error as the identity for this query ear. Table 1 lists the rank-1 recognition rates achieved by using our algorithm and ICP, where M stands for the number of samples for each subject. Table 2 lists the time cost consumed by one identification operation, where N stands for the gallery size.

Table 1: Rank-1 recognition rate

	$M = 5$	$M = 7$	$M = 10$
ICP	83.83%	89.64%	94.09%
Our Algorithm	**87.79%**	**91.53%**	**95.23%**

Table 2: Time cost for 1 identification operation (seconds)

	$N = 850$	$N = 925$	$N = 1016$
ICP	127.45	144.31	152.45
Our Algorithm	**0.041**	**0.047**	**0.05**

3.3. Discussions

From experimental results shown in Table 1, it can be seen that our proposed method performs better than ICP in terms of rank-1 recognition rate.

The greatest advantage of our algorithm over ICP is that it has a low time cost. Table 2 compares the computational time cost for one ear query process. To recognize the identity for one query ear, the ICP based algorithm has to compare the query ear to all the gallery ears and each comparison is an ICP alignment for two ear shapes. Without a previous rejection process, the whole recognition process will cost a lot of time. Different from ICP, the recognition process based on sparse representation just solves an l_1-minimization problem based on pre-calculated features, so it is much faster than ICP based approaches. With gallery size rising, the computational time of ICP approach will hugely rise. However, the computational time of sparse representation based on DALM algorithm changes little with the enlargement of the gallery size, which can also be reflected from the results listed in Table 2.

4. CONCLUSION

In this paper, we proposed a novel 3D ear recognition approach. In order to make use of the sparse representation framework for identification, we proposed a new template-based ear detection method. By using this method, extracted ear regions are in a common standard coordinate system defined by the ear contour template, which highly facilitates the following feature extraction and recognition steps. Experimental results indicate that the proposed method could achieve high ear detection rate, high identification accuracy and low computational cost.

ACKNOWLEDGEMENT

This work is supported by the Fundamental Research Funds for the Central Universities under grant no. 2100219033, the Natural Science Foundation of China under grant no. 61201394, and the Innovation Program of Shanghai Municipal Education Commission under grant no. 12ZZ029.

5. REFERENCES

[1] A. Iannarelli, *Ear Identification*, Paramont Publishing Company, 1989.

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

[3] M. Burge and W. Burger, "Ear biometrics in computer vision," *ICPR'00*, pp. 822-826, 2000.

[4] D. Hurley, M. Nixon, and J. Carter, "Force field feature extraction for ear biometrics," *CVIU*, vol. 98, pp. 491-512, 2005.

[5] K. Chang, K.W. Bowyer, S. Sarkar, and B. Victor, "Comparison and combination of ear and face images in appearance-based biometrics," *IEEE Trans. PAMI*, vol. 25, pp. 1160-1165, 2003.

[6] H. Chen and B. Bhanu, "Contour matching for 3D ear recognition," *Proc. Seventh IEEE Workshop Application of Computer Vision*, pp. 123-128, 2005.

[7] P. Besl and N. McKay, "A method for registration of 3-D shapes," *IEEE Trans. PAMI*, vol. 14, pp. 239-256, 1992.

[8] H. Chen and B. Bhanu, "Shape model-based 3D ear detection from side face range images," *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshop Advanced 3D Imaging for Safety and Security*, 2005.

[9] P. Yan and K.W. Bowyer, "Biometric recognition using three-dimensional ear shape," *IEEE Trans. PAMI*, vol. 29, pp. 1297-1308, 2007.

[10] L.D. Cohen, "On active contour models and balloons," *CVGIP*, vol. 53, pp. 211-218, 1991.

[11] H. Chen and B. Bhanu, "Human ear recognition in 3D," *IEEE Trans. PAMI*, vol. 29, pp. 718-737, 2007.

[12] S.M. Islam, R. Davies, M. Bennamoun, and A.S. Mian, "Efficient detection and recognition of 3D ears," *Int. J. Comput Vis.*, vol. 95, pp. 52-73, 2011.

[13] P. Viola and M. J. Jones, "Rapid object detection using a boosted cascade of simple features," *CVPR'01*, pp. 511-518, 2001.

[14] A. Mian, M. Bennamoun, and R. Owens, "Keypoint detection and local feature matching for textured 3D face recognition," *Int. J. Computer Vis.*, vol. 79, pp. 1-12, 2008.

[15] D. Donoho. "Compressed sensing," *IEEE Trans. Information Theory*, vol. 52, pp. 1289-1306, 2006.

[16] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. "Robust face recognition via sparse representation," *IEEE Trans. PAMI*, vol. 31, pp. 210-227, 2009.

[17] X. Li, T. Jia, and H. Zhang. "Expression-insensitive 3d face recognition using sparse representation," *CVPR'09*, pp. 2575-2582, 2009.

[18] P.J. Besl and R.C. Jain, "Invariant surface characteristics for 3D object recognition in range images," *CVGIP*, vol. 33, pp. 33-80,1986.

[19] P. Flynn and A. Jain, "Surface classification: Hypothesis testing and parameter estimation," *CVPR'88*, pp. 261-267, 1988.

[20] S. Islam, R.davies, S. Mian, and M.Bennamoun, "A fast and fully automatic ear recognition approach based on 3D local surface features," *ACICS'08*, pp. 1081-1092, 2008 .

[21] J. Yang and Y. Zhang, "Alternating direction algorithms for l1-problems in compressive sensing," *Technical Report, Rice University*, 2009.

[22] CVRL Datasets, http://www3.nd.edu/~cvrl.