A Fast and Fully Automatic Ear Recognition Approach Based on 3D Local Surface Features

S.M.S. Islam, R. Davies, A.S. Mian, and M. Bennamoun

The University of Western Australia, Crawley, WA 6009, Australia
{shams,rowan,ajmal,bennamou}@csse.uwa.edu.au

Abstract. Sensitivity of global features to pose, illumination and scale variations encouraged researchers to use local features for object representation and recognition. Availability of 3D scanners also made the use of 3D data (which is less affected by such variations compared to its 2D counterpart) very popular in computer vision applications. In this paper, an approach is proposed for human ear recognition based on robust 3D local features. The features are constructed on distinctive locations in the 3D ear data with an approximated surface around them based on the neighborhood information. Correspondences are then established between gallery and probe features and the two data sets are aligned based on these correspondences. A minimal rectangular subset of the whole 3D ear data only containing the corresponding features is then passed to the Iterative Closest Point (ICP) algorithm for final recognition. Experiments were performed on the UND biometric database and the proposed system achieved 90, 94 and 96 percent recognition rate for rank one, two and three respectively. The approach is fully automatic, comparatively very fast and makes on assumption about the localization of the nose or the ear pit, unlike previous works on ear recognition.

1 Introduction

Among the biometric traits used for computer vision, the face and the ear have gained most of the attention of the research community due to their non-intrusiveness and the ease of data collection. Face recognition with neutral expressions has reached its maturity with a high degree of accuracy. But changes of face geometry due to the changes of facial expression, use of cosmetics and eye glasses, aging, covering with beard or hair significantly affect the performance of face recognition systems. The ear is considered as an alternative to be used separately or in combination with the face as it is comparatively less affected by such changes. However, its smaller size and often the presence of nearby hair and ear-rings makes it very challenging to be used for non-interactive biometric applications.

As noted in the survey of Pun et al. [1] and Islam et al. [2], most of the proposed ear recognition approaches use either Principal Components Analysis (PCA) [3,4,5] or the ICP algorithm [3,4,5,6,7,8,9] or their combination [10] for matching purposes. Choras [11] and Yuizono et al, [12] proposed geometrical feature-based and genetic local search based approaches respectively. Both

J. Blanc-Talon et al. (Eds.): ACIVS 2008, LNCS 5259, pp. 1081–1092, 2008.
© Springer-Verlag Berlin Heidelberg 2008

of them reported error-free recognition but with comparatively smaller dataset containing high quality 2D ear images taken on the same day and without having any hair or ear-ring. Similarly, Hurley et al. [4] proposed the force field transformation for ear feature extraction and claimed 99.2% recognition on a smaller data set of only 63 subjects and without considering occlusions with ear-rings and hair.

The first ever ear recognition system tested with a larger database (415 subjects) is proposed by Yan and Bowyer [6]. Using an automatic ear detection based on the localization of the nose and the earpit, active contour based ear data extraction and finally, matching with a modified version of the ICP achieved an accuracy of 95.7% allowing occlusion and 97.8 % on examples without any occlusion (with an Equal-error rate (EER) of 1.2%). The system is not expected to work properly if the nose (for example, due to pose variation) or the ear pit (for example, due to its covering with hair) are not clearly visible which is a common case. In an experiment where a straight-on ear images was matched with twenty four 45 degree off images (a subset of Collection G of the UND database), it achieves only 70.8% recognition rate.

Most of the approaches above are based on global features. This require requires an accurate normalization of ear data with respect to pose, illumination and scale. These approaches are also inherently sensitive to occlusion. As demonstrated in this paper, local features are less affected by these factors. Recently, Chen and Bhanu [13] used a local surface shape descriptor to represent ear data. However, they only used the representation for a coarse alignment of the ear. The whole ear data was then used for matching with a modified version of ICP. They obtained 96.4% recognition on the Collection F of UND database (302 subjects) and 87.5% recognition for straight-on to 45 degree off images. They reported an ear detection accuracy of 87.1% only. Moreover, they assume that all the ear data are accurately extracted (manually, if needed) from the profile images prior to recognition.

In this paper, the 3D local surface features proposed for face recognition in [14] are adapted for the ear recognition. The authors of the work reported a very high recognition rate of 99% on neutral versus neutral and 93.5% on neutral versus all face data when tested on the FRGC v2 3D face data set. They also obtained a very good time efficiency of 23 matches per second on a 3.2 GHz Pentium IV machine with 1GB RAM. However, since ear features are different and more challenging than face features, we modified the feature creation and matching approach to make them suitable for ear recognition. Following [14] at first, a smaller number of distinctive 3D feature point locations are identified on each of the fully automatically detected 3D ear region. A 3D surface is then approximated around the selected keypoint based on the nearby data points and used as the feature for that point. A coordinate frame centred on the key point and aligned with the principal axes from PCA is used to make the features pose invariant. Correspondence is established between the gallery and the probe features and the matching decision is made based on the distance between the feature surfaces and the transformation between them. This yields a reasonable

recognition method based on only local features. However, the recognition performance is improved by aligning the probe and the gallery data set based on the initial transformation between the corresponding features and followed by the application of the Iterative Closest Point (ICP) algorithm on only a minimal rectangular subset of the whole 3D ear data containing the corresponding features only. This novel approach of extracting a reduced data set for final alignment significantly increases the efficiency in time also. Thus, the proposed system have three main advantages: 1) fully automatic 2) comparatively very fast and 3) makes on assumption about the localization of the nose or the ear pit, unlike previous works on ear recognition.

The paper is organized as follows. The proposed approach for 3D ear recognition is described in Sect. 2. The results obtained are reported and discussed in Sect. 3 followed by conclusions in Sect. 4.

2 Methodology

Our ear recognition system consist of seven main parts as shown in Fig. 1. Each of the components is described in this section.

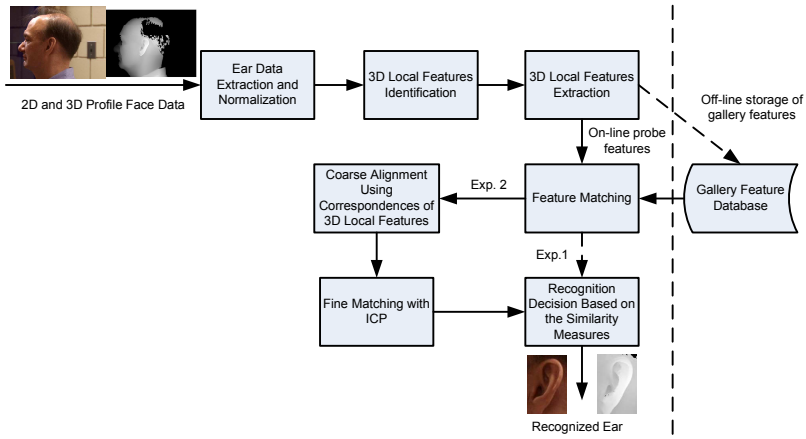

Fig. 1. Block diagram of the proposed ear recognition system

2.1 Ear Data Extraction and Normalization

The ear region is detected on 2D profile face images using the AdaBoost based detector described by Islam et al. [15]. This detector is chosen as it is fully automatic and also due to its speed and high accuracy of 99.89% on the UND profile face database with 942 images of 302 subjects [16]. The corresponding 3D data is then extracted from the co-registered 3D profile data as described in [16]. To ensure the whole ear is included and to allow the extraction of features on and slightly outside the ear region, we expanded the detected ear regions by an additional 25 pixels around each direction.

Consequently, the extracted 3D ear data varies in dimensions depending on the detection window. Hence, we normalized the 3D data by centering on the mean and then sampling on a uniform grid of 132 by 106. The surface fitting was performed using an interpolation algorithm at 0.5mm resolution. Since there were some missing data regions as shown in Fig. 2, we removed interpolated data for those regions after fitting to the grid.

2.2 Feature Location Identification

A 3D local feature can be depicted as a 3D surface constructed using data points within a sphere of radius r_1 centred at location p. As outline by Mian et al [14], the criteria to check while identifying feature locations is that it should be on a surface that is distinctive enough to differentiate between range images of different persons.

To avoid many matching features in a single smaller region (in other word, to increase distinctiveness), we only consider as possible feature points that lie on a 2mm grid. Then we find the distance of each data point from the boundary and take only those points with a distance greater than a predefined boundary limit. The boundary limit is chosen slightly longer than the radius of the 3D local feature surface (r_1) so that the feature calculation does not depend on regions outside the boundary and the allowed region corresponds closely with the ear. We call the points within this limit as seed points.

To check whether the data points around a seed point contain enough descriptive information, we adopt the approach of Mian et al. [14] discussed in short as

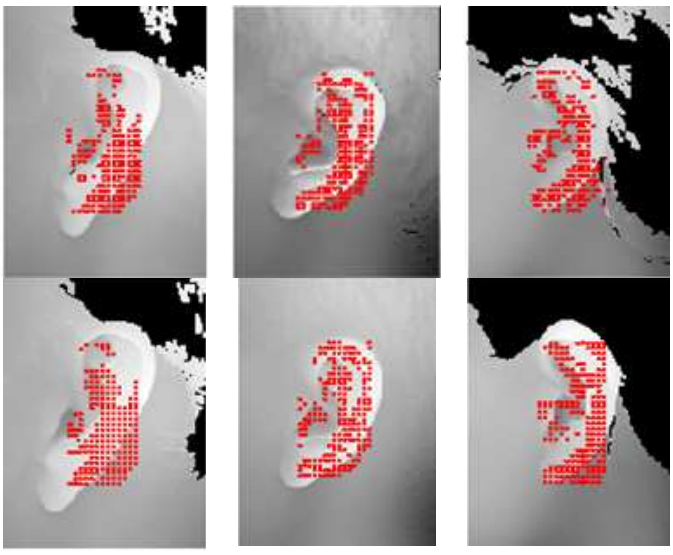

Fig. 2. Locations of local features (shown with dots) on the range images of different views (in rows) of different individuals (in columns). (This figure is best seen in color).

follows. We randomly choose a seed point and take a sphere of data points around that point which are within a distance of r_1. We apply the PCA on those data points and align them with their principal axes using the computed rotation matrix. The difference between the ranges of the first two principal axes of the local region is computed as δ. It is then compared to a threshold (δ_t). We only accept a seed point to be a distinctive feature location if the δ is higher than δ_t. The higher δ_t the less number of features we get. But lowering δ_t can result in the selection of less significant feature points. This is because, the value of δ indicates extent of unsymmetrical variation in depth in that point cloud. For example, δ_t of zero for a point cloud means it could be completely planar or spherical.

We continue selecting feature locations from the available seed points until we get a significant number of points (F_n). For a seed resolution of 2mm, r_1 of 10, δ_t of 2 and F_n of 200, for most of the gallery and the probe ear we found 200 feature locations. We found however, as low as 65 features particularly for cases where missing data occurs. The value of these parameters were empirically chosen. However, it is reported by Mian et al. [14] that the performance of the feature point detection algorithm does not vary significantly with small variations of these parameters.

Fig. 2 shows the suitability of our local features on the ear data. It illustrates as it appears that local feature locations are different for ear images of different individuals. It also shows that these features have a high degree of repeatability for the ear data of the same individual. Here by repeatability we mean the

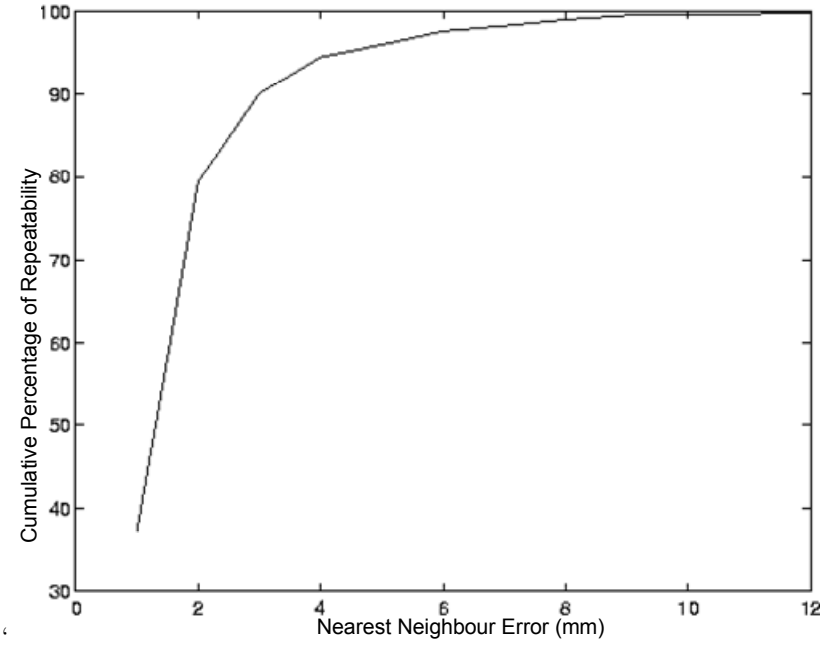

Fig. 3. Repeatability of local feature locations

proportion of probe feature points that have a corresponding gallery feature point within a particular distance. Similar to [14], the probe and gallery data of same individual are aligned using the ICP as in before computation of the repeatability. The cumulative percentage of repeatability as a function of nearest neighbor error between gallery and probe features of ten different individual is shown in Fig. 3. The repeatability reaches around 80% at an error of 2mm which is the sampling distance between the seed points.

2.3 3D Local Feature Extraction

After a seed point qualifies as a keypoint, we extract a surface feature from its neighborhood. As described in Sect. 2.2, while testing for suitability of the seed point we take the sphere of data points r_1 away from that seed point and aligned to their principal axes. We use these rotated data points to construct the 3D local surface feature. Similar to [14], the principal direction of the local surface is used as the 3D coordinates to calculate the features. Since the coordinate basis is defined locally based on the shape of the surface, the computed features are potentially stable and pose invariant.

We fit a uniformly sampled (with resolution of 1mm) 3D surface of 30×30 lattice to these data points. In order to avoid the boundary effects, we crop the inner region of 20×20 lattice from the bigger surface. This smaller surface is then concatenated to form a feature vector to be used for matching. Consequently, the dimension of our feature vector is 400. An example of 3D local surface feature is shown in Fig. 4.

For surface fitting, we use a publicly available surface fitting code [17]. The motivation behind the selection of this algorithm is that it builds a surface over the complete lattice, extrapolating (rather than interpolating) smoothly into the corners. Therefore, it is less sensitive to noise and outliers in the data.

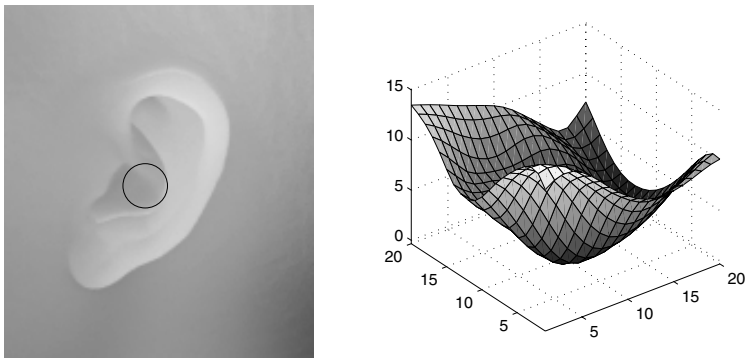

Fig. 4. Example of a 3D local surface (right image). The region from which it is extracted is shown by a circle on the left image.

2.4 Feature Matching

The similarity between two features is calculated as the Root Mean Square (RMS) distance between corresponding points on the 20×20 grid generated when the feature is created (aligned following the axes in the PCA). The RMS distance is computed from each probe feature location to all the gallery feature locations. Matching gallery features which are located more than a threshold (th) away are discarded to avoid matching in quite different areas of the cropped image. The RMS distance of the probe feature and the remaining gallery feature is then computed. The gallery feature that corresponds to the minimum distance is considered as the corresponding gallery feature for that particular probe feature. The mean of the distances for all the matched probe and gallery features is used as a similarity measure.

Unlike [14], we also used the implied rotation between probe and gallery for each pair of matching features which is calculated from the rotations used to generate the two features. The angle between each of these rotations and all the others is calculated and the rotation which has the most similar rotations (within five degrees) is chosen as the most representative rotation. The ratio of the size of the largest cluster of rotation angles to the total number of matching features is used as an additional similarity measure.

2.5 Coarse Registration of Gallery and Probe Data

The input profile images of the gallery and the probe may have pose (rotation and translation) variations. To minimize the effect of such variations, unlike [14], we use the correspondence and rotation information obtained from the 3D local feature matching for the initial or coarse registration of the probe to the gallery image data. We applied the following approaches for this purpose. In the first approach Singular Value Decomposition (SVD) is used to find the rotation and translation matrix from the gallery and probe data points corresponding to the matched 3D local features (see Sect. 2.4). In the second approach, the rotation and the translation with the maximum number of occurrences (within five degrees and 2mm of limit respectively) are used to coarsely align the probe data to the gallery.

2.6 Fine Matching with ICP

The Iterative Closest Point (ICP) algorithm [18] is considered to be one of the most accurate algorithm for registration of two clouds of data points provided the data sets are roughly aligned. Since ICP is computationally expensive, we extracted a reduced rectangular region fed to a modified version of ICP in [19]. The minimum and maximum co-ordinate values of the matched local 3D features were used to extract the reduced rectangular region from the originally detected gallery and probe ear data. This smaller but feature-rich region also minimizes the probability of being affected by the presence of hair and ear-rings.

2.7 Final Similarity Measures

The final decision regarding the matching is made based on the results of ICP as well as the local feature matching. Therefore, the final similarity measures are: (i) Distance between local 3D features (ii) ICP error and (iii) The ratio of the size of the largest cluster of rotation angles to the total number of matching features (RR).

As in [14], each of the similarity measures was scaled to 0 and 1 for its minimum and maximum values respectively. A weight factor is then computed as the ratio of the difference of the minimum value from the mean to that of the second minimum value from the mean of a similarity measure. The final result is the weighted summation of all the similarity measures. However, the third similarity measure (RR) is subtracted from one before multiplication with the corresponding weight factor as it is opposite to other measures (the higher this value the better are the results).

3 Results and Discussions

The recognition performance of our proposed approach is evaluated in this section. The results with and without ICP are reported separately. Examples of correct and misclassifications are analyzed. The time requirement for matching is also reported.

3.1 Data Set Used

The Collection F from the University of Notre Dame Profile face database is used to perform recognition experiments of the proposed approach. We have taken 200 profile images of the *first* 100 different subjects. Among these images, 100 of the images that were collected in the year 2003 are used in the gallery database and the first 100 images of the same subjects collected in the year 2004 are used in the probe database.

3.2 Recognition Rate with 3D Local Features Only

In our first experiment, we performed recognition considering the matching errors with local 3D features only. We have obtained 84%, 88% and 90% identification rate for rank-1, rank-2 and rank-3 respectively for this experiment. The results are shown in the plot of Fig. 5.

3.3 Fine Matching with the ICP

Some of the matching failures using the local 3D features only are recovered by fine alignment with the ICP after the initial alignment using the local 3D features. Using the combined similarity measure described in Sect. 2.7, identification rate improved to reach 90%, 94% and 96% respectively for rank-1, rank-2 and rank-3. The results are shown in Fig. 5.

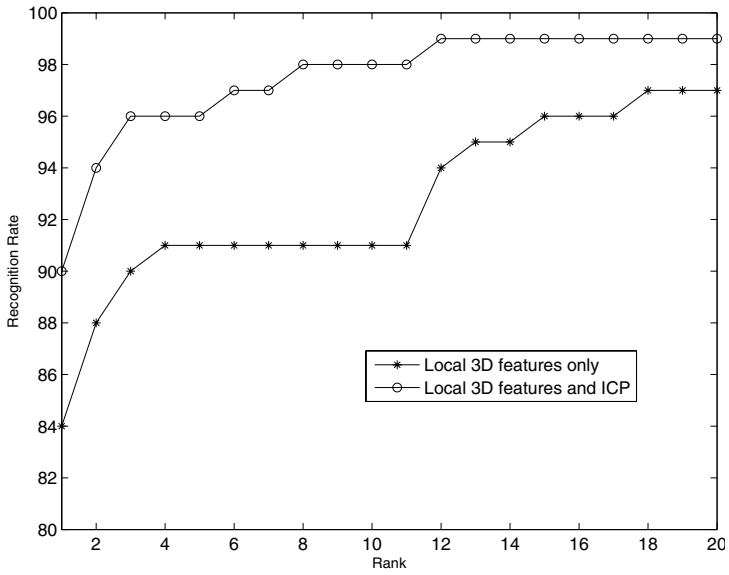

Fig. 5. Identification results

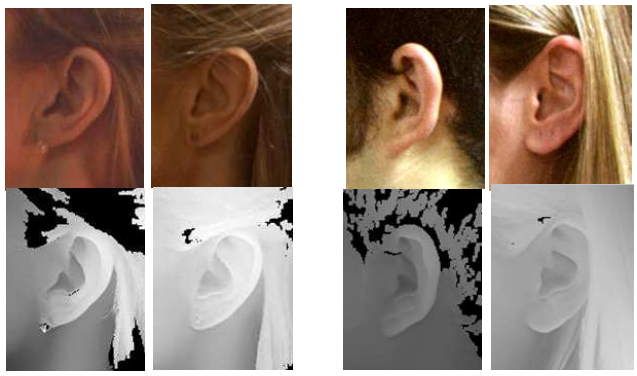

Fig. 6. Examples of correct recognition in the presence of occlusions. (a) With ear-rings. and (b) With hair. (2D and the corresponding range images are placed in the top and bottom row respectively).

3.4 Occlusion and Pose Invariance

Our approach with local 3D features is found to be robust to the presence of partial occlusions due to hair and ear-rings. Some examples are illustrated in Fig. 6.

Fig. 7. Example of correct recognition of gallery-probe pairs with pose variations

The proposed local 3D features are pose-invariant due to the way they have been created. However, if the pose variation (specially the out-of-plain) causes self-occlusions, the number of 3D local features and their repeatability decreases in the gallery-probe pair. Therefore, we noticed some misclassifications using local 3D features only in presence of some pose variations. However, with the finer registration with ICP, most of those failures were recognized correctly. Fig. 7 shows two such examples. Profile images are used for the example on the right to illustrate the pose variations.

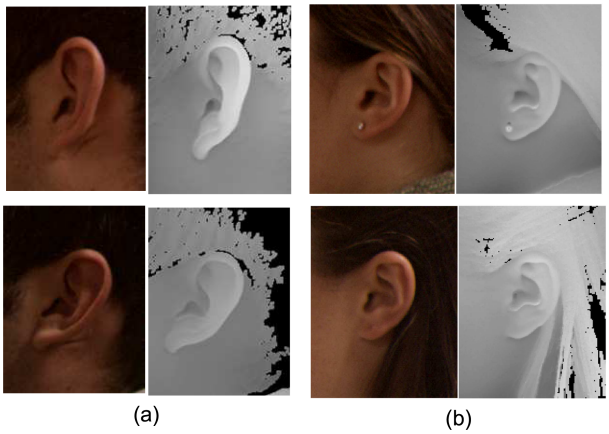

Fig. 8. Example of misclassification. (a) With large pose variations. (b) With ear-ring, hair and pose variations.

3.5 Analysis of the Failures

The proposed system fails mostly in cases of missing data (due to sensor error or ear-rings), large pose variations causing self-occlusion and severe occlusions with hair and ear-rings.

The repeatability of the local 3D features in misclassified images was found to be very low. The worst misclassified gallery-probe pair has only around 33% repeatability at 2mm (the feature point resolution).

Two examples of misclassification are illustrated in Fig. 8: one has large in-plain and out-of-plain rotation, and the other has large out-of-plain rotation, ear-ring and hair covering a portion. The hair around the gallery ear of the second example hides the depth variation at the edges (see the range image in Fig. 8b).

3.6 Recognition Speed

An un-optimized implementation of the recognition system is performed on MATLAB on a Pentium 4, 3.6 GHz and 3.25 GB RAM. It takes around 0.3762 sec for matching 3D local features in a gallery-probe pair. This timing is a bit longer than what is reported in [14] as we are not using any compression of the feature vector. We also perform the computation for finding the rotation similarity measure (see Sect. 2.4) during matching the features. Time required with the ICP is 8.6 sec for each match on the same platform.

4 Conclusion

In this paper, a robust 3D local features based approach is proposed for 3D ear recognition. The approach is fully automatic and comparatively very fast. It is shown to be robust to pose and scale variations and occlusion due to hair and ear-rings. It is also not based on any assumption about the localization of the nose or the ear pit. The large variation in rank-2 and rank-3 matches also indicates that finer tuning of the parameters in our system is likely to improve the performance. The time efficiency of the system can also be improved by reducing the dimensionality of the local feature vector by projecting the features to the PCA subspace.

Acknowledgements

We acknowledge the use of the UND Biometrics databases for ear detection and recognition. We also like to thank D'Errico for the surface fitting code. This research is sponsored by ARC grants DP0664228 and DP0881813.

References

1. Pun, K.H., Moon, Y.S.: Recent advances in ear biometrics. In: Proc. of the Sixth IEEE Int'l Conf. on Automatic Face and Gesture Recognition, pp. 164–169 (2004)
2. Islam, S., Bennamoun, M., Owens, R., Davies, R.: Biometric Approaches of 2D-3D Ear and Face: A Survey. In: Proc. of Int'l Conf. on Systems, Computing Sciences and Software Engineering, SCSS 2007 (2007)

3. Zhang, H.J., Mu, Z.C., Qu, W., Liu, L.M., Zhang, C.Y.: A novel approach for ear recognition based on ica and rbf network. In: Proc. of Int'l Conf. on Machine Learning and Cybernetics, 2005, pp. 4511–4515 (2005)
4. Hurley, D.J., Nixon, M.S., Carter, J.N.: Force field feature extraction for ear biometrics. Computer Vision and Image Understanding 98, 491–512 (2005)
5. Chang, K., Bowyer, K., Sarkar, S., Victor, B.: Comparison and combination of ear and face images in appearance-based biometrics. IEEE Transactions on Pattern Analysis and Machine Intelligence 9, 1160–1165 (2003)
6. Yan, P., Bowyer, K.W.: Biometric recognition using 3D ear shape. IEEE Trans. on PAMI 29, 1297–1308 (2007)
7. Yan, P., Bowyer, K.W.: An automatic 3D ear recognition system. In: Proc. of the Third Int'l Symposium on 3D Data Processing, Visualization and Transmission (2006)
8. Yan, P., Bowyer, K.W.: Icp-based approaches for 3d ear recognition. In: Jain, A.K., Ratha, N.K. (eds.) Biometric Technology for Human Identification II. Proc. of SPIE, vol. 5779, pp. 282–291 (2005)
9. Chen, H., Bhanu, B.: Contour matching for 3d ear recognition. In: Proc. of the 7th IEEE Workshops on Application of Computer Vision, pp. 123–128 (2005)
10. Yan, P., Bowyer, K.W.: Empirical evaluation of advanced ear biometrics. In: Proc. of Conf. on Empirical Evaluation Methods in Computer Vision (2005)
11. Choras, M.: Ear biometrics based on geometrical feature extraction. Electronic Letters on Computer Vision and Image Analysis 5, 84–95 (2005)
12. Yuizono, T., Wang, Y., Satoh, K., Nakayama, S.: Study on individual recognition for ear images by using genetic local search. In: Proc. of Congress on Evolutionary Computation, pp. 237–242 (2002)
13. Chen, H., Bhanu, B.: Human ear recognition in 3d. IEEE Trans. on PAMI 29, 718–737 (2007)
14. Mian, A., Bennamoun, M., Owens, R.: Keypoint Detection and Local Feature Matching for Textured 3D Face Recognition. International Journal of Computer Vision (IJCV) 79, 1–12 (2008)
15. Islam, S., Bennamoun, M., Davies, R.: Fast and Fully Automatic Ear Detection Using Cascaded AdaBoost. In: Proc. of IEEE Workshop on Application of Computer Vision WACV 2008, pp. 1–6 (2008)
16. Islam, S., Bennamoun, M., Mian, A., Davies, R.: A Fully Automatic Approach for Human Recognition from Profile Images Using 2D and 3D Ear Data. In: Proc. of the Fourth International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT 2008) (2008)
17. D'Errico, J.: Surface fitting using gridfit. MATLAB Central File Exchange Select (2006)
18. Besl, P.J., McKay, N.D.: A Method for Registration of 3-D Shapes. IEEE Trans. on PAMI 14, 239–256 (1992)
19. Mian, A., Bennamoun, M., Owens, R.: An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face Recognition. IEEE Trans. on PAMI 29, 1927–1943 (2007)