# 4

# Recognizing 3D Ears Using Ear Helix/Anti-Helix

During the ear detection described in Chapter 3, the optimization procedure drives the initial global registration towards the ear helix and anti-helix parts, which results in the one-to-one correspondence of the ear helix and anti-helix between the reference ear shape model and the input image. We propose to match 3D ears using the ear helix/anti-helix representation [1]. First the correspondence of ear helix and anti-helix parts (available from the ear detection algorithm) between every gallery-probe ear pair is established and it is used to compute the initial rigid transformation. Then this transformation is applied to randomly selected control points of the hypothesized gallery ear in the database. A modified iterative closest point (ICP) algorithm is run to improve the transformation which brings the gallery ear and probe ear into the best alignment, for every gallery-probe pair. The root mean square (RMS) registration error is used as the matching error criterion. The subject in the gallery with the minimum RMS error is declared as the recognized person in the probe image.

## 4.1 Ear Helix/Anti-Helix Representation

The 3D coordinates of the ear helix and anti-helix parts are obtained from the detection algorithm and these coordinates form our ear helix/anti-helix representation. The detected ear helix/anti-helix parts are marked by the green points in Figure 4.1

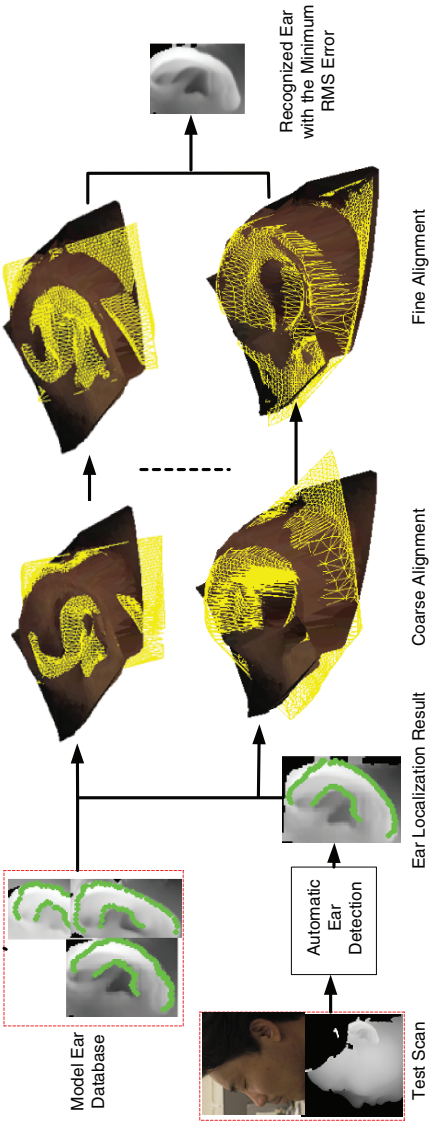



**Fig. 4.1.** The surface matching block diagram. The extracted ear helix and anti-helix parts are marked by the dots. The transformed model scan displayed as a mesh is overlaid on the textured test scan.

## 4.2 Surface Matching

As shown in the ear recognition part of Figure 4.1, the surface matching follows the coarse-to-fine strategy. Given a set of probe images, the ear and its helix and the anti-helix parts are extracted by running the detection algorithm described above. The correspondence of the helix and the anti-helix between the probe ear and the hypothesized gallery ear is used to compute the initial transformation that brings the hypothesized gallery ear into coarse alignment with the probe ear, and then a modified ICP algorithm is run to refine the transformation to bring gallery-probe pairs into the best alignment.

### 4.2.1 Coarse Alignment

Given two corresponding sets of $N_p$ 3D vertices $M$ and $S$ on the helix and the anti-helix parts, the initial rigid transformation, which brings the gallery and the probe ears into coarse alignment, can be estimated by minimizing the sum of the squares of theses errors (equation (4.1)) with respect to the rotation matrix $R$ and the translation vector $T$. The rotation matrix and translation vector are computed by using the quaternion representation [76].

$$\Sigma = \frac{1}{N_p} \sum_{l=1}^{N_p} |S_i - R * M_i - T|^2 \tag{4.1}$$

The unit quaternion is a vector $\mathbf{q_R} = [q_0, q_1, q_2, q_3]^T$, where $q_0 \geq 0$ and $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$. The solution of $R$ and $T$ using the quaternion representation is performed as follows:

1. Calculate the centroid of the gallery data set $M_m$ and the probe data set $S_s$ given by (4.2).

$$\mu_S = \frac{1}{N_p} \sum_{i=1}^{N_p} S_i \text{ and } \mu_M = \frac{1}{N_P} \sum_{i=1}^{N_p} M_i \tag{4.2}$$

2. Calculate the covariance matrix $\Sigma_{MS}$ of the sets $M_m$ and $S_s$ is given by

$$\Sigma_{MS} = \frac{1}{N_p} \sum_{i=1}^{N_p} [(M_i - \mu_M)(S_i - \mu_S)^T] \tag{4.3}$$

3. Construct the matrix $A_{ij} = (\Sigma_{MS} - \Sigma_{MS}^T)_{ij}$ which is used to form the column vector $\Delta = [A_{23}\ A_{31}\ A_{12}]^T$. This vector is then used to form the symmetric $4 \times 4$ matrix $Q$.

$$Q = \begin{bmatrix} tr(\Sigma_{MS}) & \Delta^T \\ \Delta & \Sigma_{MS} + \Sigma_{MS}^T - tr(\Sigma_{MS}) I_3 \end{bmatrix} \tag{4.4}$$

where $I_3$ is the $3 \times 3$ identity matrix and $tr$ is the matrix trace operator.

4. The maximum eigenvalue of $Q$ is corresponding to the $\mathbf{q_R}$.

5. The $3 \times 3$ rotation matrix generated from a unit rotation quaternion is calculated by (4.5).

$$R = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1 q_2 - q_0 q_3) & 2(q_1 q_3 + q_0 q_2) \\ 2(q_1 q_2 + q_0 q_3) & q_0^2 + q_2^2 - q_1^2 - q_3^2 & 2(q_2 q_3 - q_0 q_1) \\ 2(q_1 q_3 - q_0 q_2) & 2(q_2 q_3 + q_0 q_1) & q_0^2 + q_3^2 - q_1^2 - q_2^2 \end{bmatrix} \tag{4.5}$$

The translation vector $T$ is calculated as $T = \mu_S - R\mu_M$.

#### 4.2.2 Fine Alignment

Given the estimate of initial rigid transformation, the purpose of iterative closest point (ICP) algorithm [63] is to determine if the match is good and to find a refined alignment between them. If the probe ear is really an instance of the gallery ear, the ICP algorithm will result in a good registration and a large number of corresponding points between gallery and probe ear surfaces will be found. Since ICP algorithm requires that the probe be a subset of the gallery, a method to remove outliers based on the distance distribution is used [70]. The basic steps of the modified ICP algorithm are summarized below:

- Input: A 3D gallery ear range image, a 3D probe ear range image and the initial transformation obtained from the coarse alignment.
- Output: The refined transformation between the two ears.
- Procedure:
  (a) Select control points ($\sim$180) in the gallery ear range image randomly and apply the initial transformation to the gallery ear image.

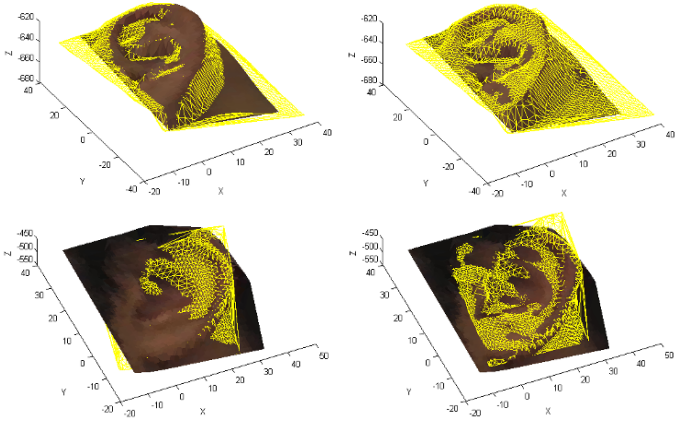
(b) Find the closest points of the control points in the probe ear image and compute the statistics [70] of the distances between the corresponding pairs in the gallery and probe images.
(c) Discard some of the corresponding pairs by analyzing the statistics of the distances (a threshold is obtained based on the mean and standard deviation of distances) [70].
(d) Compute the rigid transformation between the gallery and the probe ears based on the correspondences.
(e) Apply the transformation to the gallery ear range image and repeat step b) until convergence.

Starting with the initial transformation obtained from the coarse alignment, the modified ICP algorithm is run to refine the transformation by minimizing the distance between the control points of the gallery ear and their closest points of the probe ear. For each gallery ear in the database, the control points are randomly selected and the modified ICP is applied to those points. For a selected gallery ear, we repeat the same procedure 15 times and choose the rigid transformation with the minimum root mean square (RMS) error. The subject in the gallery set with the minimum RMS error is declared as the recognized person. In the modified ICP algorithm, the speed bottleneck is the nearest neighbor search. Therefore, the K-d tree structure is used in the implementation. Figure 4.2(a) shows the coarse alignment after applying the initial rigid transformation; Figure 4.2(b) shows the refined alignment after applying the modified ICP algorithm. In Figure 4.2, the gallery ear represented by the mesh is overlaid on the textured 3D probe ear. We observe a better alignment after applying the modified ICP algorithm.

## 4.3 Experimental Results

In order to evaluate and compare the matching performance on the selected datasets, all the ears are correctly extracted. In these limited cases where the ears are not successfully detected in an automated manner, they are correctly extracted by human interaction.

In the UCR dataset there are 155 subjects with 902 images. The data are split into a gallery set and a probe set. Each set has 155 subjects and every subject in the probe set has an instance in the gallery set. In



(a) Coarse alignment (b) Fine alignment

**Fig. 4.2.** Two examples of coarse and fine alignment. The gallery ear represented by the mesh is overlaid on the textured 3D probe ear.

order to evaluate the proposed surface matching schemes, we perform experiments under three scenarios:

- one frontal ear of a subject is in the gallery set and another frontal ear of the same subject is in the probe set;
- two frontal ears of a subject are in the gallery set and the rest of the ear images of the same subject are in the probe set;
- two ears of a subject are randomly chosen and put in the gallery set and the rest of ear images for the same subject are in the probe set.

All the experiments are repeated five times. The three scenarios are denoted by $ES_1$, $ES_2$ and $ES_3$ respectively. $ES_1$ is used for testing the performance of the system to recognize ears with the same pose; $ES_2$ is used for testing the performance of the system to recognize ears with pose variations; $ES_3$ is used for testing the robustness of the system.

In the UND dataset Collection F, there are 302 subjects and each subject has two images. The gallery set has 302 images and the probe set has the corresponding 302 images. The experimental results on the UND dataset Collection F are obtained using the same parameters of the ear recognition algorithm as those used on the UCR dataset.

Note that the resolution of the sensors for the UCR and UND datasets are different. We anticipate improvement in performance by fine tuning the parameters on the UND dataset. However, these experiments are not performed since we wanted to keep the algorithm parameters fixed across datasets.

#### 4.3.1 Identification Performance

Given a probe ear image, the root mean square (RMS) error is calculated for every enrolled subject in the gallery set and the subject in the gallery with the minimum RMS error is declared as the recognized person in the probe image. The matching error matrix $\{ME(i,j), i = 1,2,\ldots,N_t, j = 1,2,\ldots,N_m\}$ on the UCR dataset $ES_1$ using the helix/anti-helix representation, where $N_t$ and $N_m$ are the number of probe ears and gallery ears respectively, is displayed as an intensity image shown in Figure 4.3. The smaller the matching error is the more likely the two ears match. From Figure 4.3, we can see that most of the diagonal pixels are darker than the other pixels on the same row, which means correct recognition. The average time to match a pair of ears, which includes the coarse and fine alignment, is about 1.1 seconds

**Table 4.1.** Cumulative matching performance on the UCR dataset and the UND dataset Collection F. The number of images in the gallery and probe sets are listed in the parenthesis.

| Dataset | Rank-1 | Rank-2 | Rank-3 | Rank-4 | Rank-5 |
|---|---|---|---|---|---|
| $ES_1(155, 155)$ | 96.77% | 98.06% | 98.71% | 98.71% | 98.71% |
| $ES_2(310, 592)$ | 94.43% | 96.96% | 97.80% | 98.31% | 98.31% |
| $ES_3(310, 592)$ | 92.91% | 95.95% | 96.96% | 97.30% | 97.47% |
| | 94.26% | 95.78% | 96.28% | 96.45% | 96.96% |
| | 92.74% | 94.76% | 95.44% | 95.78% | 95.95% |
| | 92.23% | 95.44% | 95.61% | 96.45% | 96.79% |
| | 92.23% | 94.93% | 95.44% | 95.78% | 95.78% |
| $UND(302, 302)$ | 96.03% | 96.69% | 97.35% | 97.68% | 98.01% |

with C++ implementation on a Linux machine with an *AMD Opteron* 1.8GHz CPU.

The identification performance is evaluated by the cumulative match characteristics (CMC), which describes "is the right answer in the top rank-r matches?". Table 4.1 shows the rank-r recognition rates for the UCR dataset and the UND dataset Collection F. In Table 4.1, the numbers of images in the gallery and the probe sets are listed in the parenthesis following the name of the dataset. We achieve 96.77% rank-1 recognition rate (150 out of 155) on the UCR dataset $ES_1$ and 96.03% rank-1 recognition rate (290 out of 302) on the UND dataset Collection F. As expected, the system performs better on $ES_1$ with the same pose and the performance degrades slightly on $ES_2$ with pose variations. By an inspection of the CMC values on $ES_3$ listed in Table 4.1, it also can be seen that the system is robust to recognize ears under different conditions. We observe that without retuning the parameters of the proposed algorithm we still achieved good recognition performance on the UND dataset which has several weeks of time lapse between the gallery and the probe.

Figure 4.4 shows three examples of the correctly recognized gallery-probe ear pairs with a large pose variation. Figure 4.4(a) shows the side face color images of the gallery and the probe alternately; Figure 4.4(b) shows the range images of the ears that are automatically extracted; Figure 4.4(c) shows the gallery ear represented by the mesh overlaid on the textured 3D probe ear images. We observe that the cases with a large pose variation are correctly handled.

We show three special cases of correctly recognizing gallery-probe ear pairs in Figures 4.5. In these two figures, the probe ear is rendered

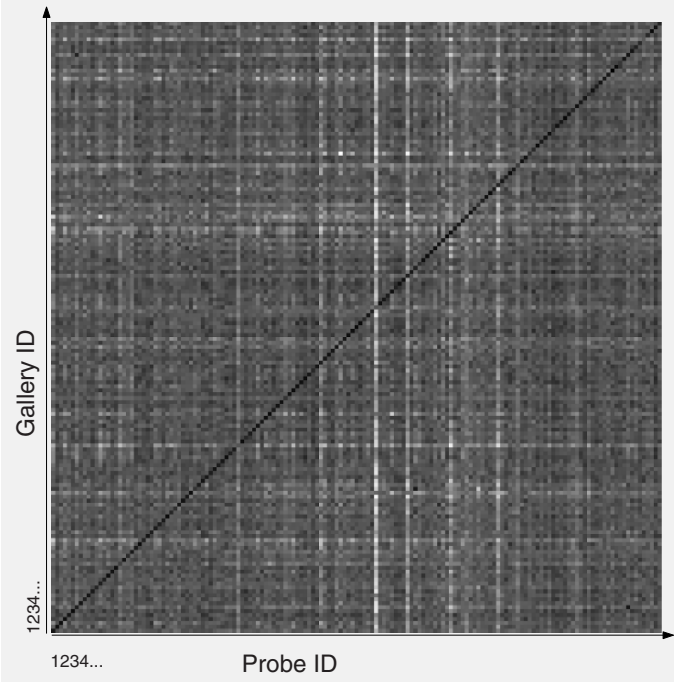



**Fig. 4.3.** Matching error matrix on the UCR dataset $ES_1$ using the ear helix/anti-helix representation displayed as an intensity image (smaller values correspond to darker pixels). The gallery ID is labeled horizontally and the probe ID is labeled vertically.

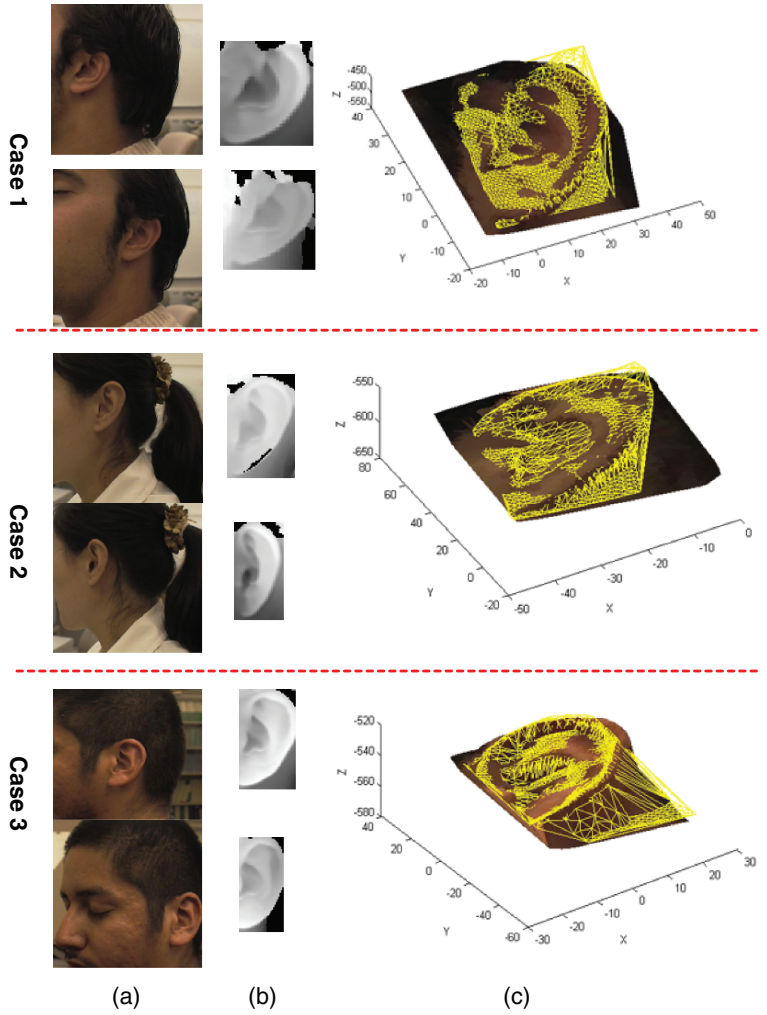



**Fig. 4.4.** UCR dataset: Three cases of the correctly recognized gallery-probe ear pairs using the ear helix/anti-helix representation with a large pose variation. (a) Side face color images. (b) Range images of the detected ears. In columns (a) and (b), the gallery image is shown first and the probe image is shown second. (c) The probe ear with the corresponding gallery ear after alignment. The gallery ear represented by the mesh is overlaid on the textured 3D probe ear. The units of x, y and z are millimeters (mm). In case 1, the rotation angle is $33.5^\circ$ and the axis is $[0.0099,\ 0.9969,\ 0.0778]^T$. In case 2, the rotation angle is $-33.5^\circ$ and the axis is $[-0.1162, 0.9932, 0.0044]^T$. In case 3, the rotation angle is $32.9^\circ$ and the axis is $[0.0002,\ 0.9998,\ 0.0197]^T$.

as a textured 3D surface; the gallery ear is displayed as a mesh. In order to examine our results visually, we display the gallery ear and probe ear in the same image (Figure 4.5(b)) and also the transformed gallery and probe ear in the same image (Figure 4.5(c)). From Figure 4.5, we observe that the ear recognition system can handle partial occlusion. Twelve more examples of correctly recognized gallery-probe ear pairs are shown in Figure 4.6. The images in the columns (a) and (c) display test ears and their corresponding gallery ears before alignment; the images in the columns (b) and (d) show probe ears and correctly recognized gallery ears after alignment. From Figure 4.6, we see that each gallery ear is well aligned with the corresponding probe ear.

During the recognition, some errors are made and four cases are illustrated in Figure 4.7. Figure 4.7(a) and (b) show the color images of two visually similar probe and gallery ears that belong to different subjects; Figure 4.7(c) shows the true gallery ear overlaid on the textured 3D probe ear after registration; Figure 4.7(d) shows the falsely recognized gallery ear overlaid on the textured 3D probe ear after alignment. In Figure 4.7(d), the root mean square error for the falsely recognized ear is smaller than the error for the correct ear in Figure 4.7(c). Since we pick up the gallery ear with the minimum RMS error as the recognized ear, we made the errors. In this figure, we obtain good alignment between the gallery and model ears from different persons since these ears are quite similar in 3D.

#### 4.3.2 Verification Performance

The verification performance of the proposed system is evaluated in terms of the two popular methods, the receiver operating characteristic (ROC) curve and the equal error rate (EER). The ROC curve is the plot of genuine acceptance rate (GAR) versus the corresponding false acceptance rate (FAR). GAR is defined as the percentage of the occurrences that an authorized user is correctly accepted by the system, while FAR is defined as the percentage of the occurrences that a non-authorized user is falsely accepted by the system. The EER, which indicates the rate at which false rejection rate ($FRR = 1 - GAR$) and the false acceptance rate are equal, is a threshold independent performance measure.

During the verification, the RMS distance is computed from matching the gallery ears to the probe ears and it is then compared to a

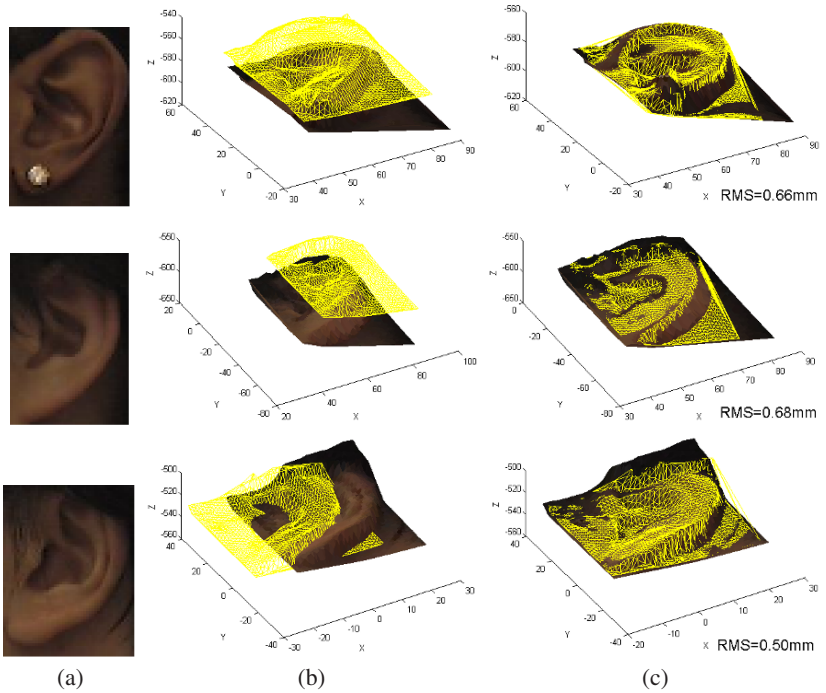



**Fig. 4.5.** Three special cases of correctly recognized gallery-probe ears using the ear helix/anti-helix representation, in which ears are partially occluded by an earring or by the hair. (a) Color images of the ears. (b) Examples of probe ears with the corresponding gallery ears before alignment. (c) Examples of test ears with the correctly recognized gallery ears after alignment. The gallery ear represented by the mesh is overlaid on the textured 3D test ear. The units of x, y and z are millimeters (mm).

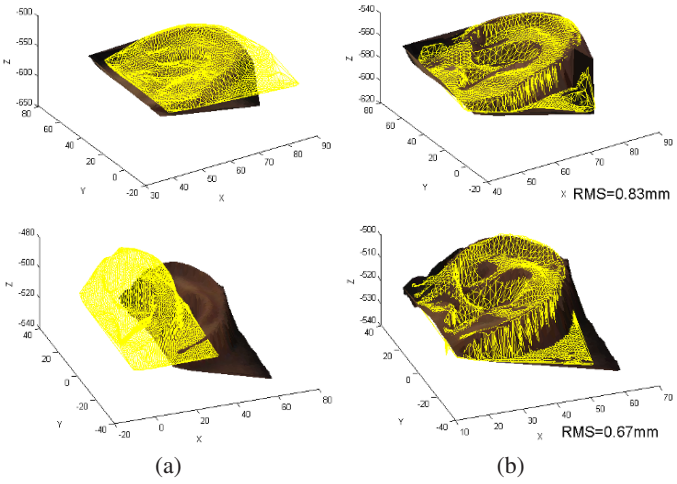
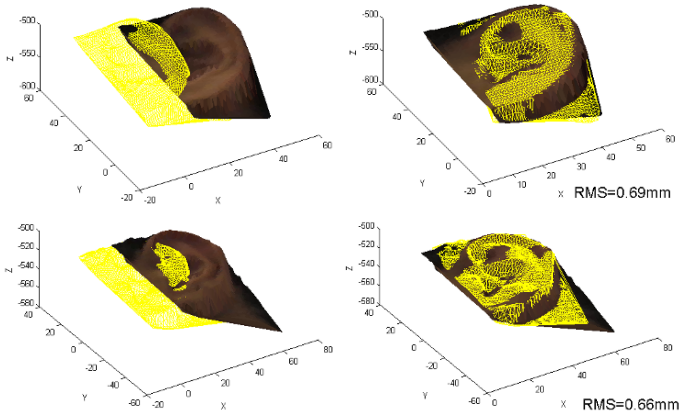



**Fig. 4.6.** UCR dataset: Twelve cases of the correctly recognized gallery-probe pairs using the ear helix/anti-helix representation. (a) Examples of probe ears with the corresponding gallery ears before alignment. (b) Examples of probe ears with the correctly recognized gallery ears after alignment. The gallery ear represented by the mesh is overlaid on the textured 3D probe ear. The units of x, y and z are millimeters (mm).




(a) (b)

**Fig. 4.6.** Figure 4.6 Continued.

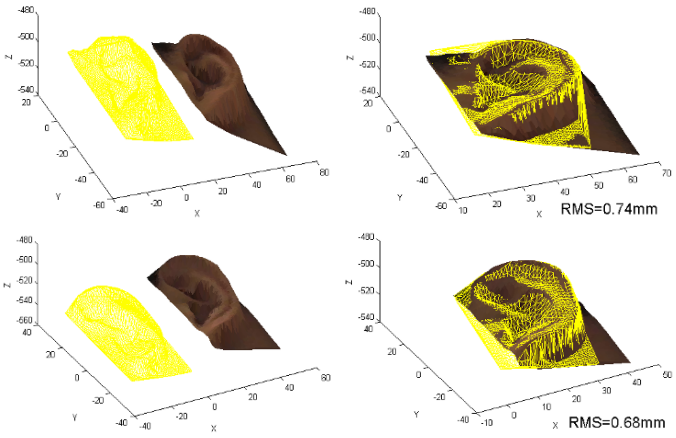
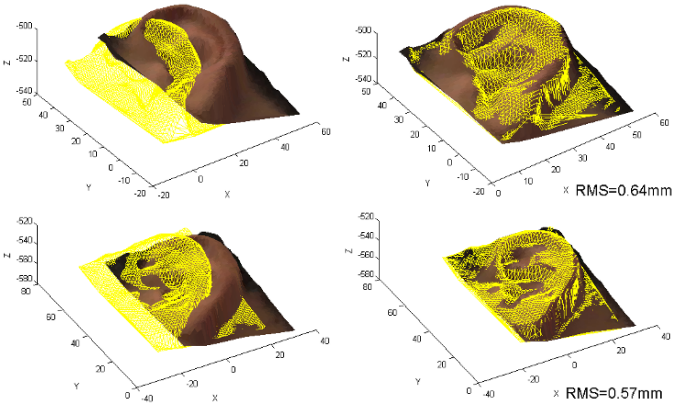



**Fig. 4.6.** Figure 4.6 Continued.




(a) (b)

**Fig. 4.6.** Figure 4.6 Continued.

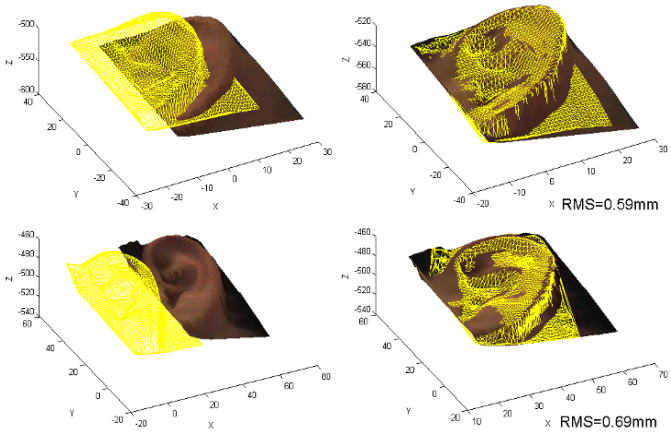
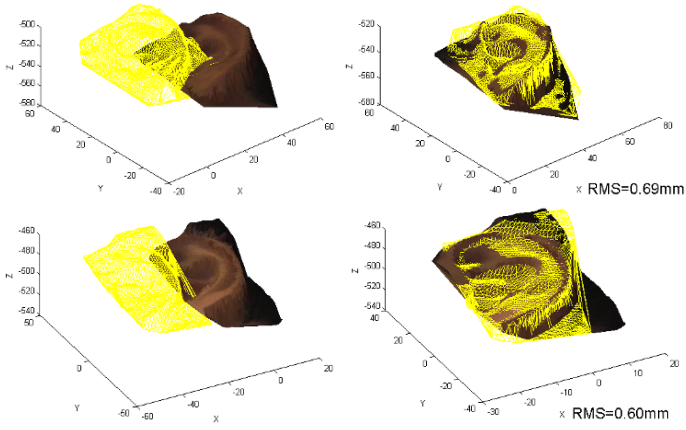



**Fig. 4.6.** Figure 4.6 Continued.




(a) (b)

**Fig. 4.6.** Figure 4.6 Continued.

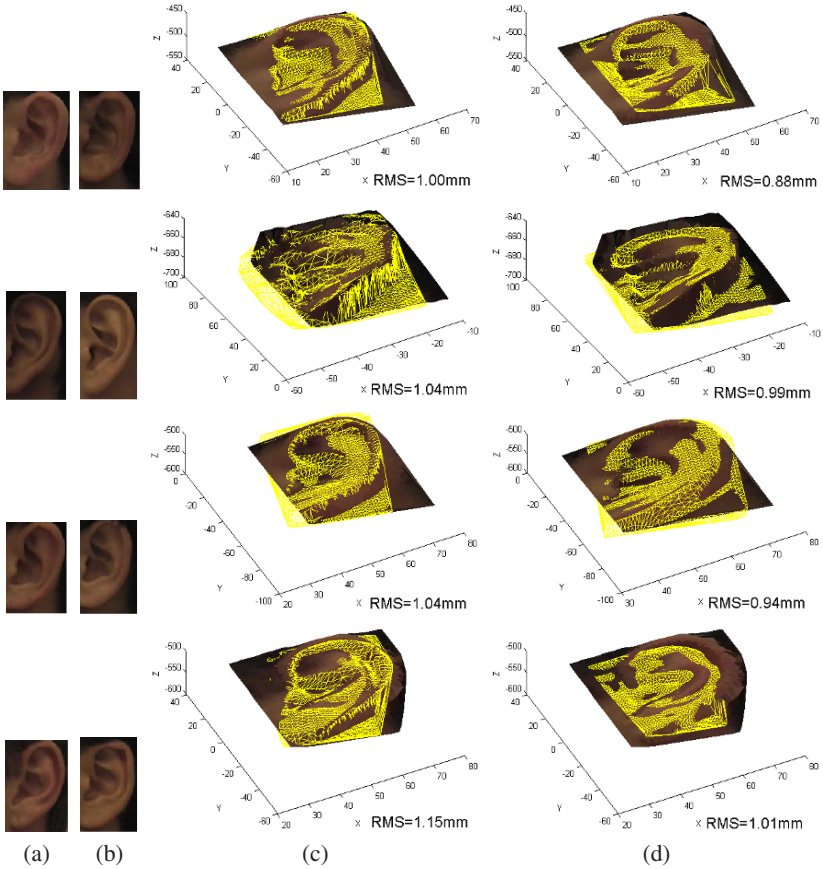
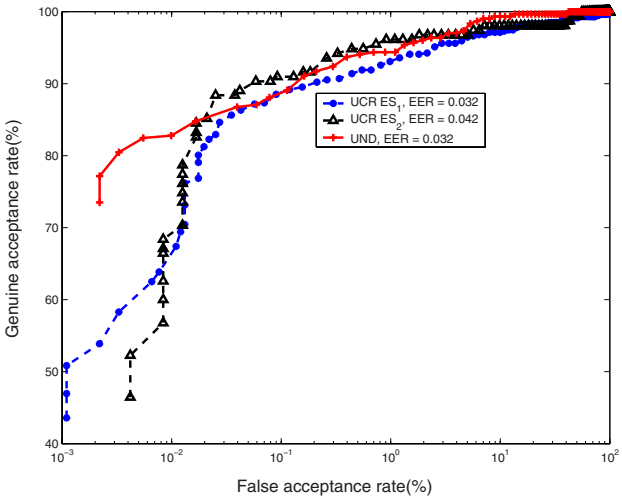
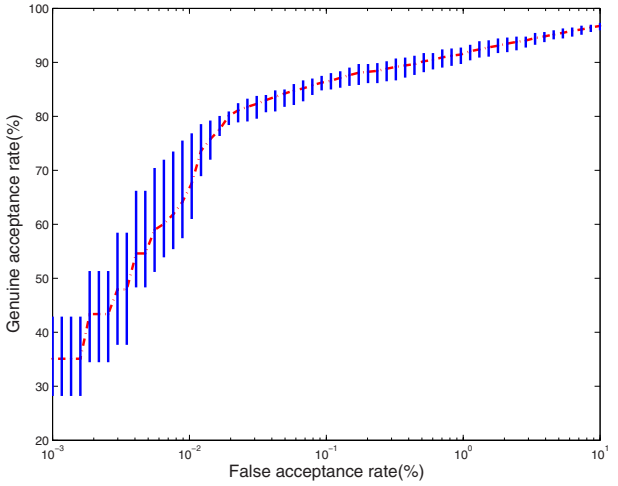



**Fig. 4.7.** UCR dataset: Four cases of incorrectly recognized gallery-probe pairs using the ear helix/anti-helix. Each row shows one case. The gallery ears represented by the mesh are overlaid on the textured 3D probe ears. The units of x, y and z are millimeters (mm). (a) Color images of the probe ears. (b) Color images of falsely recognized gallery ears. (c) True gallery ears after alignment are overlaid on the textured 3D probe ears. (d) The falsely recognized gallery ears after alignment are overlaid on the textured 3D probe ears. Note that for the incorrect matches the gallery ears in column (d) achieve a smaller value of RMS error than the gallery ears in column (c).




(a)




(b)

**Fig. 4.8.** UCR dataset and UND dataset Collection F: Verification performance as a ROC curve using the ear helix/anti-helix representation. (a) ROC curves on the UCR dataset $ES_1$, $ES_2$ and the UND dataset. (b) ROC curve on the UCR dataset $ES_3$. The dotted line is the average of GAR over a particular FAR and the vertical line indicates the range of GAR for a particular FAR over 5 runs. The EERs for the 5 runs are 0.042, 0.043, 0.049, 0.045, and 0.055, respectively.

threshold to determine if the probe is an authorized user or an imposter. By varying the threshold FAR and GAR values are computed and plotted in Figure 4.8. Figure 4.8(a) shows the ROC curves on the UCR dataset $ES_1$, $ES_2$ and the UND dataset; Figure 4.8(b) shows the ROC curve on the UCR dataset $ES_3$. As expected, the system performs best on $ES_1$ with a 0.032 EER. It can be seen that EERs changed slightly on the three scenarios, which suggests that the system is capable of verifying ears with pose variations and partial occlusions.

## 4.4 Conclusions

We have proposed the ear helix/anti-helix representation for surface matching in 3D. The 3D coordinates of the ear helix/anti-helix parts are obtained from the detection algorithm and they are used to estimate the initial transformation between a gallery-probe pair. Then a modified ICP algorithm iteratively refines the transformation which brings the hypothesized gallery and a probe into the best alignment. We have performed extensive experiments the UCR dataset (155 subject with 902 images) and the UND dataset Collection F (302 subject with 302 time-lapse pairs). It is encouraging to observe that the proposed ear recognition system is capable of recognizing ears with pose variation and partial occlusions.