

Ear Biometrics in Human Identification

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Ping Yan, B.S., M.S.

Kevin W. Bowyer, Director

Graduate Program in Department of Computer Science and Engineering

Notre Dame, Indiana

June 2006

Ear Biometrics in Human Identification

Abstract

by

Ping Yan

Biometrics are physical or behavioral characteristics that can be used for human identification. Security plays an increasingly important role in our daily life, and biometric technologies are becoming the solution to highly secure recognition and verification of identity. In this dissertation, we propose the ear as a biometric and investigate its potential with both 2D and 3D data.

Our work is the largest experimental investigation of ear biometrics to date. The data set used for our experiments contains 415 persons, each with images acquired on at least two different dates. Approaches considered include a PCA (“eigen-ear”) approach with 2D intensity images and range images, a Hausdorff matching of edges from range images, and an ICP-based approach to matching the 3D data. Our experimental results demonstrate that the ICP-based approach outperforms the other approaches at a statistically significant level.

Furthermore, we develop the first fully automated biometric system using 3D ear shape. It is a complete automatic system starting with a raw 3D image, through automated segmentation of the ear, and 3D shape matching for recognition. The automatic system achieves a rank-one recognition rate of 97.6% on a 415-subject dataset. Our algorithm also shows good scalability of recognition rate with size of dataset. The results suggest a strong potential for 3D ear shape as a biometric.

In a biometrics scenario, gallery images are enrolled into the database ahead of the matching step, which provides the opportunity to build related information before the probe comes into the system. We present a novel approach, called “Pre-computed Voxel Closest Neighbors”, to reduce the computational time for shape matching in a biometrics context. The approach shifts the heavy computation burden to the enrollment stage, which can be done offline. Experiments in 3D ear and face biometrics demonstrate the effectiveness of the approach.

In addition, ear symmetry and partial ear shape experiments are investigated. The results indicate that most people’s left and right ears are approximately bilaterally symmetric. However, some people have ears with recognizably different shapes. Experimental results with partial ear shape suggest that minor hair covering does not affect the performance substantially, but large hair covering will certainly reduce the recognition rate. This suggests that even in circumstances where the complete ear shape cannot be captured, partial shape has potential for recognition. This lends support for using ear shape as a biometric.

Our experiments use the biometric database collected at the University of Notre Dame. This data set is available to other research groups.

CONTENTS

FIGURES	v
TABLES	viii
ACKNOWLEDGMENTS	x
CHAPTER 1: INTRODUCTION	1
1.1 Ear Biometrics	1
1.2 Overview of Our Work	3
1.3 Description of Our Work	4
CHAPTER 2: LITERATURE REVIEW	9
CHAPTER 3: DATA ACQUISITION AND PREPROCESSING	16
3.1 Data Acquisition	16
3.2 Data Preprocessing	19
3.2.1 2D Data Normalization	19
3.2.2 3D Data Normalization	19
3.2.3 Landmark Selection	25
3.3 Ear Extraction	27
CHAPTER 4: PCA-, EDGE- AND ICP-BASED APPROACHES ¹	29
4.1 PCA for 2D and 3D Ear Recognition	29
4.1.1 PCA ("Eigen Ear") Method	29
4.1.2 2D Ear Data	31
4.1.3 3D Ear Data	35

¹Sections 4.1 and 4.2 are based on the paper, "Ear Biometrics Using 2D and 3D Images", presented at the *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops: Advanced 3D Imaging for Safety and Security, 2005* [77]. Section 4.3 is based on the paper, "ICP-Based Approaches for 3D Ear Recognition", presented at the *Biometric Technology for Human Identification Conference, 2005* [81].

4.2	Hausdorff Range Edge Matching	35
4.2.1	Hausdorff Distance	35
4.2.2	Hausdorff Matching with Ear Image	37
4.3	ICP-Based Approach	41
4.3.1	ICP Algorithm	41
4.3.2	Initial Experiment	41
4.3.3	Noise Removal	42
4.3.4	Speed Limitation	43
4.3.5	Outlier Elimination	46
4.3.5.1	Outlier Elimination Using AV Distance Metric	48
4.3.6	Point to Point vs. Point to Triangle Approach	48
4.4	Statistical Significance Testing	52
4.4.1	Z-test with Bonferroni Correction	52
4.4.2	Friedman Test with Nemenyi Post-hoc Test	55
CHAPTER 5: A FAST ICP-BASED APPROACH ²		58
5.1	Introduction	58
5.2	Literature Review	60
5.3	Fast ICP Matching for 3D shapes	62
5.3.1	Volume Size	65
5.3.2	Implementation	66
5.4	Experiments	68
5.4.1	Voxel Size	69
5.5	Improvement	72
5.6	Summary And Discussion	74
CHAPTER 6: MULTI-BIOMETRICS ³		75
6.1	Fusion Levels and Score Normalization	76
6.2	Multi-modal Biometrics	76
6.3	Multi-algorithm Biometrics	83
6.4	Using Multiple Images to Represent a Person	83
6.5	Summary And Discussion	86
6.6	Improved ICP Algorithm	88
6.7	Multi-biometrics Using Ear and Face	88

²This section is based on the paper, “A fast algorithm for ICP-based 3D shape biometrics”, presented at the *Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID 2005)* [79].

³This section is based on the paper, “Multi-Biometrics 2D and 3D Ear Recognition”, presented at the *Audio- and Video-based Biometric Person Authentication Conference, 2005* [78].

CHAPTER 7: AUTOMATIC EAR RECOGNITION ⁴	92
7.1 Segmenting the Ear Region from a Profile Image	93
7.1.1 Ear Pit Detection	94
7.1.1.1 Preprocessing	94
7.1.1.2 Skin Region Detection	96
7.1.1.3 Surface Curvature Estimation	96
7.1.1.4 Surface Segmentation and Classification	98
7.1.2 Ear Segmentation Using an Active Contour Algorithm	101
7.2 Matching 3D Ear Shape for Recognition	109
7.2.1 ICP-based Matching	109
7.2.2 ICP - Perturbation of Starting Point	109
7.3 Experimental Results	113
CHAPTER 8: OCCLUSION AND SYMMETRY EAR EXPERIMENTS	116
8.1 Partial Ear Experiments	116
8.2 Ear Symmetry Experiment ⁵	120
8.3 Off-angle Experiment	120
CHAPTER 9: COMPARISON WITH OTHER BIOMETRICS	125
9.1 Comparison to 2D “Eigen-Ear” Results	125
9.2 Comparison to Commercial Face Recognition Results	125
9.3 Scalability of Recognition with Dataset Size	126
CHAPTER 10: CONCLUSIONS AND FUTURE WORK	128
10.1 Conclusions	128
10.2 Future Work	130
APPENDIX A: EAR BIOMETRICS DATA	131
APPENDIX B: EARRING DATA	132
APPENDIX C: MINOR HAIR COVERING	141
BIBLIOGRAPHY	146

⁴This section is based on the paper, “An Automatic 3D Ear Recognition System”, presented at *Third International Symposium on 3D Data Processing, Visualization and Transmission, 2006*.

⁵This section is based on the paper, “Empirical Evaluation of Advanced Ear Biometrics”, presented at the *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops: Empirical Evaluation Methods in Computer Vision, 2005 [80]*.

FIGURES

1.1	INS Form M-378 (6-92) Asking for the Right Ear to be Visible. . .	2
1.2	An Overview of Our Research Work.	3
2.1	Illustration of the Iannarelli System. In (a) 1. Helix Rim, 2. Lobule, 3. Antihelix, 4. Concha, 5. Tragus, 6. Antitragus, 7. Crus of Helix, 8. Triangular Fossa, 9. Incisure Intertragica (adapted from [12]). .	10
3.1	The Minolta VIVID 910.	17
3.2	Examples of Images Discarded for Quality Control Reasons. . . .	18
3.3	Three Points Used For Plane Fitting.	20
3.4	Two Step Rotation to Normalize Pose.	23
3.5	Pixel Size Calibration.	24
3.6	Hole Filling For 3D Range Data.	25
3.7	Example of Ear Landmarks.	26
3.8	Examples of Ear Mask and Cropped 2D and 3D Ear.	28
4.1	Experiments Using Different 2D Ear Size.	32
4.2	Eigen Ear Images Associated With 3 Largest Eigenvalues.	33
4.3	PCA Performance on 2D Ear Data Using Different Ear Size. . . .	34
4.4	Two Point Sets Illustrating the Hausdorff Distances.	36
4.5	2D and 3D Ear Images and Associated Edge Images. Canny Edge Detector Parameters are $\sigma = 1.0$, $T_{low} = 0.5$, $T_{high} = 0.5$	38
4.6	Hausdorff Ear Recognition Performance.	40
4.7	Examples of Noise Removal.	44
4.8	Examples for Outlier Elimination.	48
4.9	Four Incorrect Matches.	51
4.10	Performance of Different Approaches.	53

5.1	Voxelization of 3D Ear Data. In order to show it clearly, we present it from coarse to fine. In step 1 the volume is subdivided into 8 small voxels, and in step 2 each small voxel is subdivided into 8 even smaller voxels. And continue this subdivision until the size of each voxel is smaller than a threshold. (To implement this idea, we do not need to do it step by step, and instead we subdivide the volume once using a fixed voxel size).	63
5.2	Close Look at Voxels and Example Distance Between Voxel and Gallery Surface. (P1 is the center of the voxel 1 and the closest point on the gallery surface to P1 is P1'. P2 is the center of the voxel 2 and the closest point on the gallery surface to P2 is P2'). .	63
5.3	Steps to Calculate the Volume Size. In (b), a new coordinate system is generated from eigenvectors of the covariance matrix, where Y' is according to the direction of the largest variance in the dataset, X' to the second, and Z' to the third. Part(c) shows the new 3D shape after projecting every old point onto new coordinate system.	67
5.4	Gallery and Probe Images Show in the Same Volume. Each voxel in the volume corresponds to a point index on the gallery surface.	69
6.1	Multi-modal Biometrics Performance.	80
6.2	Relationship Between Correct Matches and Incorrect Matches. . .	81
6.3	Interval Distribution Between Correct and Incorrect Matches Using Both PCA and ICP approaches.	82
6.4	Multi-algorithm Biometrics Performance.	84
6.5	Multi-modal Biometrics using Ear and Face. The original matching score of ear alone does not return the correct match, but it is corrected by the face matching.	90
6.6	Multi-modal Biometrics using Ear and Face. The original matching score of face alone does not return the correct match, but it is corrected by the ear matching.	91
7.1	Data Flow of Automatic Ear Extraction.	94
7.2	Using Nose Tip as Center to Generate Circle Sector.	95
7.3	Ear Region with Skin Detection.	97
7.4	H and K Sign Test for Surface Classes [6].	98
7.5	Steps of Finding the Ear Pit: (a) 2D/3D raw data, (b) skin detection, (c) curvature estimation, (d) surface segmentation, (e) region classification, ear pit detection.	100
7.6	Active Contour Growing on Ear Image.	104

7.7	Active Contour Results using Only Color or Depth Information.	105
7.8	Active Contour Growing on A Real Image.	107
7.9	Effectiveness of Active Contour Algorithm when Dealing with Ear-ring and Blonde Hair.	108
7.10	Perturbation of Starting Point: The minimum error distance is 0.311 mm if using starting point from (1), the minimum error distance is 1.015 mm if using starting point from (b), and the the minimum error distance is 0.321 mm if using perturbation method, shown in (c).	110
7.11	Continued on Next Page.	111
7.12	Mismatched Probe and Gallery. ED is error distance value returned from our ICP algorithm with perturbation. The first column is the probe images, the second column is the real gallery images, and the third column is the mismatched gallery images.	112
7.13	The Performance of Ear Recognition.	114
8.1	Example of Partial Ear. The captions in part(b) through (u) indicate the percent of the ear removed.	117
8.2	Partial Ear Experiment Results.	119
8.3	After Removing Top 20% of the Ear, the Recognition Rank Jumps From Rank 23 to Rank 1.	119
8.4	Image Acquired for Ear Symmetry Experiments.	121
8.5	Examples of Asymmetric Ears.	122
8.6	Image acquired for Off Angle Experiments.	123
9.1	Scalability of Proposed Algorithm's Performance with Increased Dataset Size.	127

TABLES

2.1	Recent Ear Recognition Studies, * G=Gallery, P= Probe.	14
4.1	Recognition Rate Using Different Combinations of FM and BM. .	40
4.2	The ICP Performance by Using Different Masks for Gallery and Probe. * The time is for matching one image against 202 images in the gallery.	43
4.3	ICP Performance by Using Octree and K-d Tree. The performance shows the rank one recognition rate, and the run time is the CPU time to register one source surface to one target surface.	46
4.4	ICP Performance by Using Point-To-Point, Point-To-Triangle and Revised Version.	50
4.5	Proportion of Observed Correct Matches.	53
4.6	Statistical Test of the Difference between Performance, Using 0.05 level of significance. H_0 : There is no difference in performance between the two methods. (*: It changes from reject to accept after Bonferroni adjustment).	54
4.7	Friedman Test of the Difference between Performance, Using 0.05 level of significance. H_0 : There is no difference in performance between the two methods.	57
5.1	Ear Biometrics: Different Parameters Affected by Voxel Size. At 0.1 level of significance, there is no statistically significant difference between (1), (2) and (3) using both Z-test and Friedman test. . .	70
5.2	Face biometrics: Different Parameters Affected by Voxel Size. At 0.1 level of significance, there is no statistically significant difference between (1), (2) and (3) using both Z-test and Friedman test. . .	70
5.3	Ear Experiments: Difference Between Original ICP and Pre-computed ICP. ** At 0.1 level of significance, there is no statistically significant difference between (1), (2), (3) and (4) using both Z-test and Friedman test.	72

5.4	Ear Biometric: Run Time vs. Probe Size for Both Original ICP and Pre-computed ICP. (voxel = 0.5 mm)	73
6.1	Rank One Recognition for Fusion on Multiple Modalities (302 subjects).	77
6.2	Different Weights for Fusing the ICP and PCA Results.	78
6.3	Fraction of the Correct Match in the Different Interval Level.	79
6.4	Multi-algorithm Biometrics Using Weighted Sum Rule.	83
6.5	Fusion on Multiple Galleries and Probes (169 Subjects).	85
8.1	Experiment Results Using Partial Ear with a 415-subject dataset: All the results are shown as rank-one recognition rate. The rank-one recognition rate is 97.6% by using the whole ear as probe.	118
8.2	Off Angle Experiments with a 24-Subject Dataset.	124

ACKNOWLEDGMENTS

Biometrics research at the University of Notre Dame is supported by the National Science Foundation under grant CNS-0130839, by the Central Intelligence Agency, and by the Department of Justice under grant 2004-DD-BX-1224.

I would like to express my gratitude to my advisor Professor Kevin W. Bowyer. His direction and insight have immeasurably advanced my research. His relentless pursuit and strict attitude towards research set a great model for me. I would also like to thank Professor Patrick J. Flynn for years of mentorship.

I would like to thank Professor Gregory R. Madey and Professor Amitabh Chaudhary for their time and effort to be on my committee.

I would like to thank my husband Tim Schoenharl for his long time encouragement and patience. He showed his support through years of painful proof reading. Finally, I would like to acknowledge the lifelong support and encouragement that I have received from my parents and my sister. Words are unable to express the significance of their contribution.

CHAPTER 1

INTRODUCTION

1.1 Ear Biometrics

Biometrics is the study of measuring physical or behavioral characteristics of a person to verify or recognize his or her identity. Public safety and national security enhance the needs for biometric techniques, which are among the most secure and accurate authentication tools. External physical features such as fingerprints, face and iris have received wide attention. Fingerprint and iris are generally considered to allow more accurate biometric recognition than the face, but the face is more easily used in surveillance scenarios where fingerprint and iris capture are not feasible. However, the face by itself is not yet as accurate and flexible as desired for this scenario due to illumination changes, expression changes, makeup or eye glasses.

Ear images can be acquired in a similar manner to face images, and a number of researchers have suggested that the human ear is unique enough to each individual to allow practical use as a biometric. Several researchers have looked at using features from the ear's appearance in 2D intensity images [12, 14, 37, 38, 72]. A smaller number of researchers have looked at using 3D ear shape [8, 18]. In fact, the ear may already be used informally as a biometric. For example, the United States Immigration and Naturalization Service (INS) has a form giving

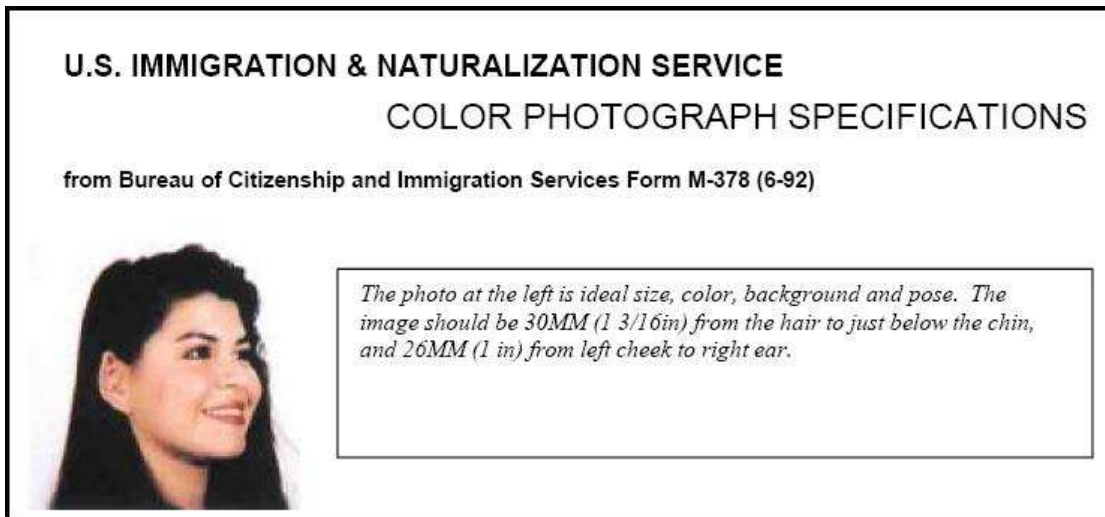


Figure 1.1. INS Form M-378 (6-92) Asking for the Right Ear to be Visible.

specifications for the photograph that indicate that the right ear should be visible [INS Form M-378 (6-92)], shown in Figure 1.1.

Researchers have suggested that the shape and appearance of the human ear is unique to each individual and relatively little change occurs during the lifetime of an adult [40]. The uniqueness of the ear appearance has not been proven, but two studies mentioned in [40] provide empirical supporting evidence. The medical report [40] shows that the variation over time is most noticeable during the period from four months to eight years old and over 70 years old. The ear growth between four months to eight years old is approximately linear, and after that it is constant until around 70 when it increases again [12]. The stretch rate due to gravity is not linear, but it mainly affects the lobe of the ear. Due to its stability and predictable changes, ear recognition is being investigated as potential biometric [8, 11, 12, 14, 37, 38, 40, 72]. Generally, ear images can be acquired in a manner similar to face images, and used in the same scenarios.

1.2 Overview of Our Work

Before we go into the details of our research, we present an overview of the different topics, as shown in Figure 1.2. First we considered three different approaches for ear recognition. The results of that work suggested two routes. One is combining two or more of the most effective approaches into a multi-modal approach. The other is to pursue the most promising single modality, which is the ICP-based approach on a 3D point set.

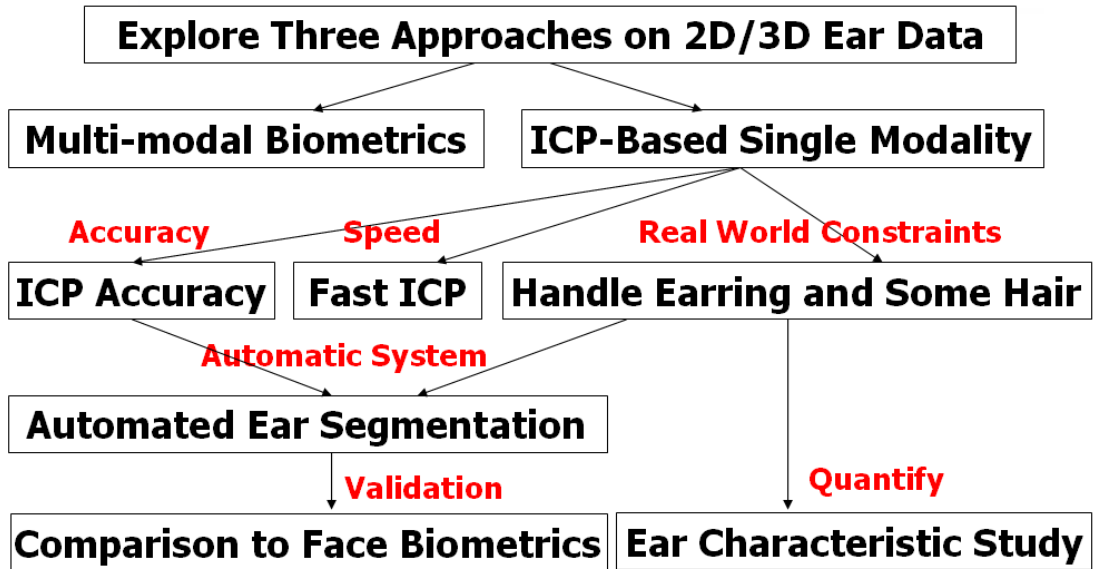


Figure 1.2. An Overview of Our Research Work.

We decided to focus on the single modality, an ICP-based approach on 3D ear shape. Three aspects were explored. First several enhancements were investigated to improve the accuracy of ICP algorithm. Next, in order to address the time limitation of the ICP algorithm, we proposed a fast ICP-based algorithm for 3D biometric application. Finally we examined the accuracy of the approach when subjected to real world constraints, such as earrings and hair covering the ear.

Our work demonstrated that 3D ear shape could be used for recognition. However, there were still several manual stages of processing the data. In a commercial system, these manual steps need to be automated. We chose to address the most challenging problem: segmenting the ear from the profile images.

As a last topic to explore for whether the ear is a good biometric source, we conducted two additional studies. First we compared the fully automatic ear recognition system with a commercial face biometric system. Second, we looked at ear characteristics as they affect matching.

1.3 Description of Our Work

A biometric system is essentially a pattern recognition system which uses a specific physiological or behavioral characteristic of a person to determine their identity or verify a claimed identity. Therefore, a biometric system can employ the methodologies from pattern recognition. For our initial experiments, we consider the use of both 2D and 3D images of the ear, using data from 302 subjects. Approaches considered include a PCA (“eigen-ear”) approach with 2D intensity images, achieving 63.8% rank-one recognition; a PCA approach with range images, achieving 55.3%; and a Hausdorff matching of edges from range images, achieving 67.5%. Starting with the general ICP algorithm proposed by [5], we

obtained a 84.3% rank-one recognition rate on a 3D ear biometric. The promising experimental results of the ICP-based approach suggest the strong potential for 3D ear shape as a biometric, and encourage us to investigate the ICP algorithm further both for recognition accuracy and for computational time.

In a biometrics scenario, gallery images are enrolled into the database prior to the matching step, which gives us the opportunity to build related information before the probe comes into the system. In this dissertation, we present a novel approach, called “Pre-computed Voxel Closest Neighbors”, to reduce the computational time for ICP shape matching in a biometrics context. The approach shifts the heavy computation burden to the enrollment stage, which can be done offline. Experiments in 3D ear biometric with 369 subjects and 3D face biometric with 219 subjects demonstrate the effectiveness of the approach.

Three algorithms have been explored on 2D and 3D ear images, and based on that, three kinds of multi-biometrics are considered: multi-modal, multi-algorithm and multi-instance. Various multi-biometric combinations all result in improvement over a single biometric. Multi-modal 2D PCA together with 3D ICP gives the highest performance. To combine 2D PCA-based and 3D ICP-based ear recognition, a new fusion rule using the interval distribution between rank one and rank two outperforms other simple combinations. The rank one recognition rate achieves 91.7% with 302 subjects in the gallery. In general, all the approaches perform much better when multiple images are used to represent one subject. In our dataset, 169 subjects had 2D and 3D images of the ear acquired on at least four different dates, which allowed us to perform multi-instance experiments. The highest rank one recognition rate was 97% with the ICP approach used to match a two-image-per-person probe against a two-image-per-person gallery. In addition,

we found that different fusion rules perform differently on different combinations. The **min** rule works well when combining the multiple presentations of one subject, while the **sum** rule works well when combining multiple modalities.

Extending our previous work, we developed the first fully automated system for ear biometric using 3D shape. There are two major parts to the system: automatic ear region segmentation and 3D ear shape matching. Starting with the multi-modal 3D+2D image acquired in a profile view, the system automatically finds the ear pit by using skin detection, curvature estimation and surface segmentation and classification. After the ear pit is detected, an active contour algorithm using both color and depth information is applied, and the contour expands to find the outline of the visible ear region.

We have found that the active contour (AC) algorithm is well suited for determining the ear region. The ear pit makes an ideal starting point. The AC algorithm grows until it finds the ear edge, and is robust in its ability to exclude earrings and occluding hair. When the active contour finishes, the outlined shape is cropped from the 3D image, and the corresponding 3D data is then used as the ear shape for matching. The matching algorithm achieves a rank-one recognition rate of 97.6% on a 415-subject dataset. In order to demonstrate the competence of our automatic ear segmentation, two other experiments are conducted on the same dataset. A comparable rank-one recognition rate using the previous manually-assisted ear cropping that uses a fixed template rather than the active contour is 96.4%. And a comparable rank-one recognition rate with everything being the same except the the location of the ear pit is manually marked is 97.1%.

This dissertation is organized as follows. A review of related work is given in chapter 2. Chapter 2 covers general ear biometrics. A review of ICP related work

is given in the chapter 5. In chapter 3, we describe the experimental method and materials used in our work. The preprocessing steps described here are relevant mainly to the methods in chapter 4, 5 and 6.

Chapter 4 details three approaches we have examined for ear biometrics: PCA-based, edge-based and ICP-based approaches. In chapter 4, we consider each of these approaches individually, and compare their performance. We find that an ICP approach on 3D data performs the best of the approaches considered. Therefore the remaining chapters focus on developing and evaluating this approach.

Chapter 5 addresses the problem of creating a faster version of the general ICP approach by exploiting the biometrics application context. The “Pre-computed Voxel Closest Neighbor” strategy improves the speed of the ICP algorithm without losing recognition accuracy.

Multi-modal biometric approaches are discussed in Chapter 6. We explore multi modalities, multi algorithms and multi instance on 2D and 3D ear data, and all improve over the performance of a single biometric. This is consistent with other multi-modal biometrics results. We also consider multi modality using 3D ear and 3D face shapes. With a 174-subject dataset, we achieve a 100% rank-one recognition rate.

Chapter 7 presents the details of the automatic ear recognition system. Using curvature information and an active contour algorithm, we are able to locate the ear from the profile image. The algorithm demonstrates the robustness of the handling hair and earring.

In chapter 8, we discuss the ear symmetry and partial ear experiments. In order to make a strong claim that the ear can be used as a biometric, we need to quantify its characteristics. We chose to look at these areas: How much of the ear

do you need to capture in order to use it for recognition? Are the left and right ears of one person symmetric?

Chapter 9 gives the comparison between our ear biometrics with other face biometrics. The results validate the promising potential of the ear as a biometric. Conclusions and future work are presented in chapter 10.

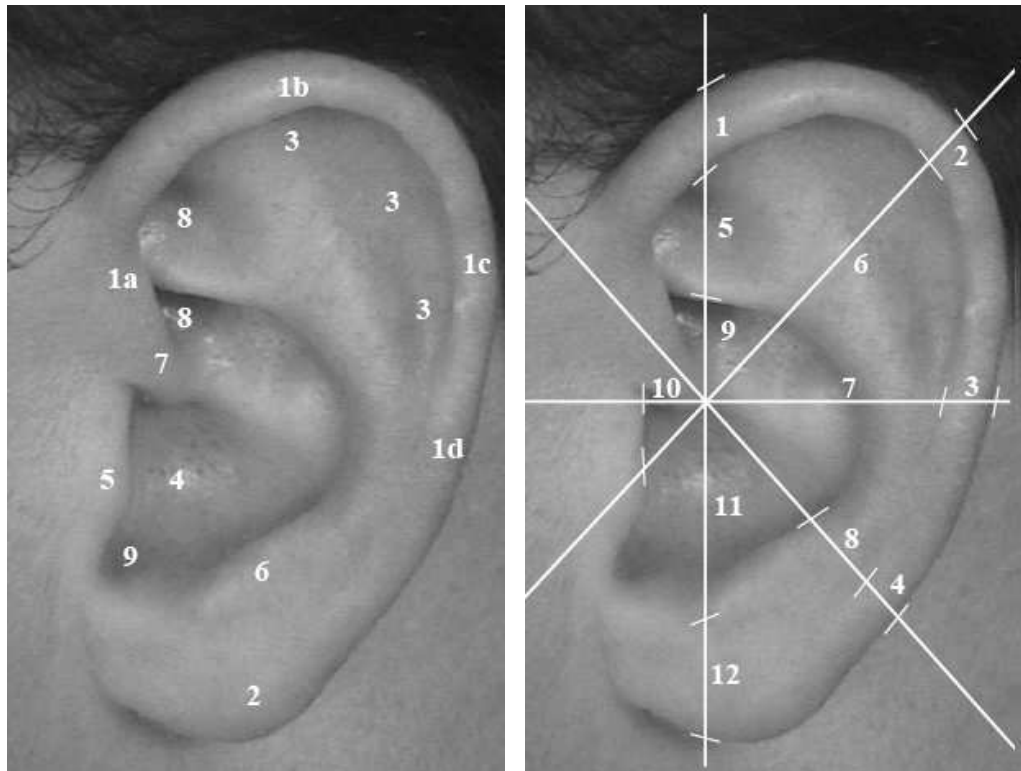
CHAPTER 2

LITERATURE REVIEW

As we mentioned before, many research studies have proposed the ear as a biometric. Researchers have suggested that the shape and appearance of the human ear is unique to each individual and relatively unchanging during the lifetime of an adult. There are several studies that address the question of uniqueness and classification of ears. No one can really prove the uniqueness of the ear, but two studies mentioned in [11, 12] provide empirical supporting evidence. In 1906, Imhofer found that only 4 characteristics were needed to distinguish a set of 500 ears [35]. The most prominent work is perhaps that done by Iannarelli (1989)[40]. In his work, over 10,000 ears were examined and no indistinguishable ears were found. Iannarelli developed an anthropometric technique for ear identification. A total of 12 measurements are used in the “Iannarelli System”, illustrated in Figure 2.1(b).

The system requires special manual alignment and normalization of the right ear 2D images. Here is the process as described by Iannarelli:

Once the ear is focused and the image is contained within the easel boundaries, adjust the easel carefully until the oblique guide line is parallel to the outer extreme tip of the tragus flesh line... The oblique line should now be barely touching the tip of the tragus. [The left two



(a) Anatomy

(b) 12 Measurements used in 'Iannarelli System'

Figure 2.1. Illustration of the Iannarelli System. In (a) 1. Helix Rim, 2. Lobule, 3. Antihelix, 4. Concha, 5. Tragus, 6. Antitragus, 7. Crus of Helix, 8. Triangular Fossa, 9. Incisure Intertragica (adapted from [12]).

short white lines in Figure 2.1(b)] Move the easel slightly, keeping the oblique line touching the tip of the tragus, until the upper section of the oblique guide line intersects the point of the ear image where the start of the inner helix rim overlaps the upper concha flesh line area just below the slight depression or hollow called the triangular fossa ...

When the ear image is accurately aligned using the oblique guide line, the ear image has been properly positioned. The technician must now focus the ear image to its proper size. The short vertical guide line [The long vertical white line in Figure 2.1(b)] on the easel is used to enlarge or reduce the ear image to its proper size for comparison and classification purposes. ([40] pp.83-84)

These examples indicate that the ear is possibly uniquely distinguishable on a limited number of features or characteristics. The medical report [40] shows that the variation over time is most noticeable during the period from four months to eight years old and over 70 years old. Due to the ear's stability and predictable changes, ear features are potentially a promising biometric for use in a human identification [8, 11, 12, 14, 37, 40, 72].

Moreno et al. [54] experiment with three neural net approaches to recognition from 2D intensity images of the ear. Their testing uses a gallery of 28 persons plus another 20 persons not in the gallery. They find a recognition rate of 93% for the best of the three approaches. They consider three methods of combining results of the different approaches - Borda, Bayesian, and weighted Bayesian combination- but do not find improved performance over the best individual method.

PCA (Principal Component Analysis) on 2D intensity images for ear biometrics has been explored by Victor [72] and Chang [14]. The two studies obtained

different results when compared with the performance of facial biometrics. Both ear and face show similar performance in Chang's study, while ear performance is worse than face in Victor's study. Chang suggested that the difference might be due to differing ear image quality in the two studies.

Yuizono [82] implemented a recognition system for 2D intensity images of the ear using genetic search. In their experiments they had 660 images from 110 persons, with 6 images per person. And the 6 images were selected from a video stream. The first three of these are used as gallery images and the last three are probe images. They reported that the recognition rate for the registered persons was approximately 100%, and the rejection rate for unknown persons was 100%.

Bhanu and Chen presented a 3D ear recognition method using a local surface shape descriptor in [8]. Twenty range images from 10 individuals are used in the experiments and each subject has two images to represent. A 100% recognition rate is reported for their ear recognition system. In [18], Chen and Bhanu use a two-step ICP algorithm on a dataset of 30 subjects with 3D ear images. They reported that this method yielded 2 incorrect matches out of 30 persons. In these two works, ears are manually extracted from profile images. They also presented an ear detection method in [19]. In the offline step, they build an ear model template from each of 20 subjects using the average histogram of the shape index [48]. In the online step, first they use a step edge detection and thresholding to find the sharp edge around the ear boundary, and then apply dilation on the edge image and connected-component labeling to search for ear region candidates. Each potential ear region is a rectangular box, and it grows in four directions to find the minimum distance to the model template. The region with minimum distance to the model template is the ear region. They get 91.5% correct detection with

2.5% false alarm rate. No ear recognition is reported based on this ear detection method.

Hurley *et al.* [37] developed a novel feature extraction technique by using force field transformation. Each image is represented by a compact characteristic vector, which is invariant to initialization, scale, rotation and noise. The experiment displays the robustness of the technique to extract the 2D ear. Their extended research applies the force field technique to ear biometrics [38]. In the experiments, they use 252 images from 63 subjects, with 4 images per person collected during 4 sessions over a 5-month period, and no subject is included if the ear is covered by hair. A classification rate of 99.2% is claimed on this 63-person dataset. The dataset comes from the XM2VTS face profiles database [53].

Choraš [21] analyzes the possibility of using the human ear as a biometric. The paper explores the dimension of 2D ear space, and introduces an ear recognition method based on geometric feature extraction. They claim that error-free recognition is obtained on “easy” images from their database. The “easy” images are the images of very high quality, with no earring and hair covering, and without illumination changes. No detailed experiment setup is reported.

Pun and Moon [60] have surveyed the comparatively small literature on ear biometrics. They summarized elements of five approaches for which experimental results have been published [8, 11, 14, 37, 82]. In Table 2.1 we compare different aspects of these and other published works.

In this dissertation, we will first look at various methods of 2D and 3D ear recognition. Our work found that an ICP-based approach statistically significantly outperformed the other approaches considered for 3D ear recognition, and also statistically significantly outperformed the 2D “eigen-ear” result [14]. There-

TABLE 2.1

Recent Ear Recognition Studies, * G=Gallery, P= Probe.

Reference	Data Used	Dataset Size	Time Lapse	Number of G/P*	Method Applied	Earrings /Occlu.	Reported Perfor.
Chen & Bhanu [18]	3D	30×2	Same Day	1/1	ICP	NO	93%
Hurley & Nixon [38]	2D	63×4	5 Month	1/1	PCA	NO	99.2%
Moreno et al. [54]	2D	28×6	Different Days	1/1	Neural Net	Not mention	93 %
Yuizono et al. [82]	2D	110×6	Same Day	3/3	Genetic search	Not mention	100%
Victor & Chang [14]	2D	88×2	15 Month	1/1	PCA	No	73%
Choraš [21]	2D	N/A	Same Day	N/A	Feature-based	No	100%

fore an ICP-based approach is extensively investigated and used as the matching algorithm in our later work. Thus although other approaches to 2D ear recognition may perform differently, it appears that ear recognition based on 3D shape is more powerful than based on 2D appearance.

Of the publications reviewed here, only two works [8, 18] deal with biometrics based on 3D ear shape. The largest dataset, in number of persons, is 110 [82]. And the presence or absence of earrings is not mentioned, except for [14, 80], in which earrings are excluded.

Comparing with the publications reviewed above, the work presented in this dissertation is unique in several points. We report results for the largest ear biometrics study to date, in terms of number of persons (415) and in terms of number of images (830). Our work is able to deal with the presence of earrings.

Ours is the only work to fully automatically detect the ear from a profile view and segment the ear from the surroundings. Because we use a large experimental dataset, we are also able to explore how the different algorithms scale with number of persons in the gallery. Ours is the only work to show the effect of multi-image representations, and is the most extensive work on multi-modal ear biometrics.

CHAPTER 3

DATA ACQUISITION AND PREPROCESSING

3.1 Data Acquisition

All the images used in this work were acquired at the University of Notre Dame. In each acquisition session, the subject sat approximately 1.5 meters away from the sensor, with the left side of the face facing the camera. Data was acquired with a Minolta VIVID 910 range scanner, shown in Figure 3.1. The technology of the VIVID 910 scanner is based on using a projected laser light strip. The light reflected from the scanned object is imaged by a CCD camera, and 3D data is then generated by employing triangulation using the known geometric relation between the projector and camera. In high resolution mode, 640 x 480 individual points can be measured per scan. In addition to 3D data, one 640 x 480 color image data is acquired nearly simultaneously with the same CCD by applying a rotating filter to separate the acquired light [2].

From 497 people that participated in two or more image acquisition sessions, there were 415 who had good 2D and 3D ear images in two or more sessions. No special instructions were given to the participants to make the face profile images particularly suitable for this ear biometrics study, and 817 out of 2,709 images were dropped for various quality control reasons: 381 instances with hair obscuring most of the ear, 74 cases with artifacts due to motion during the scan,



Figure 3.1. The Minolta VIVID 910.

and 362 cases with poor image quality in either the 3D and / or the 2D. See Figure 3.2 for examples of these problems. Using the Minolta scanner in the high resolution mode that we used may make the motion artifact problem more frequent, as it takes 8 seconds to complete a scan. There is a “fast mode” that has a shorter acquisition time but only 320 X 240 sample points.

If the subject sits still during acquisition, the 3D and 2D scans are registered as obtained. In some cases with slight subject motion inbetween the 3D and 2D scans, the registration could be manually corrected by marking the same point in each scan and translating the 2D in X and Y. We used the “Incisure Intertragica” (shown in Figure 2.1) for each ear in our experiment. of In cases of large subject motion, or motion during the several seconds of the 3D scanning time, the data had to be discarded.

The earliest good image for each of the 415 persons was used to create the gallery. The gallery is the set of images that a “probe” image is matched against for identification. The latest good image of each person was used as the probe for that person. This results in an average of 8.7 weeks time lapse between the gallery and probe images used in our experiments.



Figure 3.2. Examples of Images Discarded for Quality Control Reasons.

3.2 Data Preprocessing

The purpose of the preprocessing is to minimize the variation in the acquired image, while keeping the characteristic features of the subject. Different preprocessing methods were applied to 2D intensity data and 3D range data.

3.2.1 2D Data Normalization

We performed the 2D data normalization in two steps. First is the geometric normalization. Ears were aligned using two manually identified landmark points. The distance between the two points was used for scale, which means that all the extracted ears have the same distance between the Triangular Fossa and the Incisure Intertragica [40] (see Figure 3.7 for an example). Similarly, the orientation of the line between the two points is used for rotation. After normalization, the line between these two points is vertical in the xy plane. The second step is histogram equalization, which is used to compensate for lighting variation between images. These preprocessing steps are analogous to those in standard use in face recognition from 2D intensity images [7] and those used in previous PCA-based ear recognition using 2D intensity images [14].

3.2.2 3D Data Normalization

The normalization discussed next applies to preparing the range image from the 3D data for the 3D PCA and 3D edge-based approaches. No preprocessing is applied for the 3D ICP. 3D image normalization is more complicated than 2D normalization, due to z-direction rotation, holes and missing data [8, 14]. Three steps of the 3D normalization are 3D pose normalization, pixel size normalization for the range images and hole filling.

Normalization of 3D ear pose is required to create the range image for the 3D PCA and Hausdorff edge matching. In this study, the pose of the ear is determined by the orientation of the face plane connected with the ear. Three points are marked near the ear on the z-value image, as shown in Figure 3.3. To obtain a more stable result, all the range points inside the triangle are used for plane fitting (1). The normal of the face plane P is $\vec{n} = \langle a, b, c \rangle$, where

$$ax + by + cz = d, \text{ and } a^2 + b^2 + c^2 = 1. \quad (3.1)$$

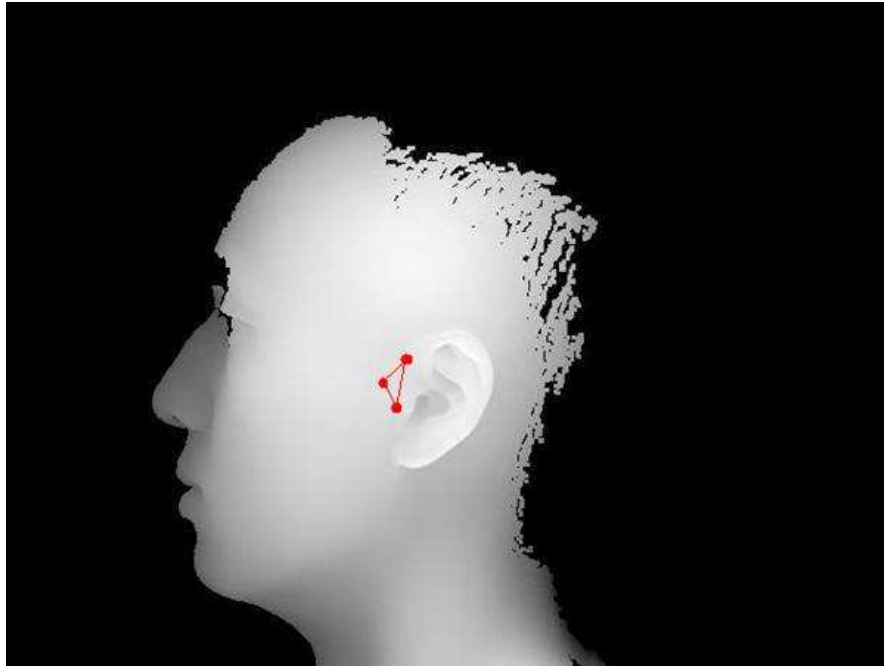


Figure 3.3. Three Points Used For Plane Fitting.

After computing the normal of the plane, we align it to a reference plane with uniform normal $\vec{n}_2 = \langle 0.199, 0, -0.98 \rangle$. The rotation is around the axis $(\vec{n}_1 \otimes \vec{n}_2)$, and the angle of the rotation is $\arccos(\vec{n}_1 \odot \vec{n}_2)$. At the same time, the two-point landmark is used to correct the orientation of the ear in the x-y plane. This rotation is around the Z axis $\langle 0, 0, 1 \rangle$.

Rotation is implemented by the quaternion, represented as:

$$q = w + xi + yj + zk; \quad (3.2)$$

where $i^2 = j^2 = k^2 = -1$, and w is a real number.

A rotation around an arbitrary axis $\langle ax, ay, az \rangle$ in 3D space by θ can be converted to a quaternion as:

$$\begin{aligned} w &= \cos(\theta/2) \\ x &= ax \times \sin(\theta/2) \\ y &= ay \times \sin(\theta/2) \\ z &= az \times \sin(\theta/2) \end{aligned} \quad (3.3)$$

Compared with Euler angle representation, quaternion representation provides an easier way to concatenate several rotations. Given two unit quaternions, Q1 = (w1, x1, y1, z1) and Q2 = (w2, x2, y2, z2), a combined rotation of Q1 and Q2 is achieved by:

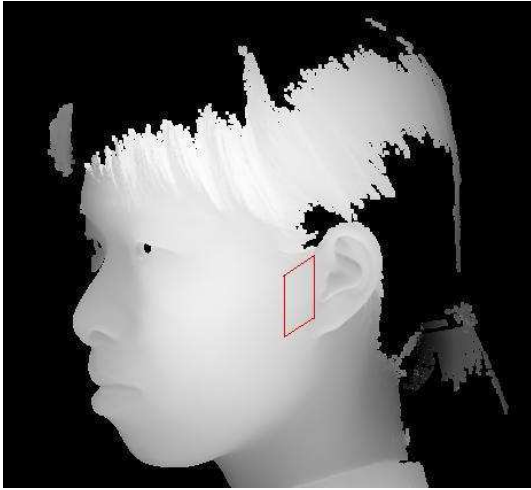
$$\begin{aligned} w &= w1w2 - x1x2 - y1y2 - z1z2 \\ x &= w1x2 + x1w2 + y1z2 - z1y2 \\ y &= w1y2 + y1w2 + z1x2 - x1z2 \\ z &= w1z2 + z1w2 + x1y2 - y1x2 \end{aligned} \quad (3.4)$$

At the end of the orientation normalization, the ear is facing at $\langle 0.199, 0, -0.98 \rangle$, and the two-point landmark is perpendicular to the x-z plane. Figure 3.4 shows the two rotations separately.

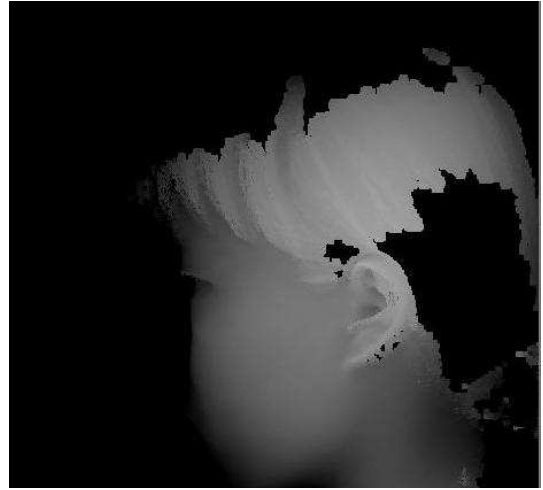
The second step of the 3D data normalization is the 2D x-y calibration. The images used for the PCA-based and edge-based methods are cropped depth images taken from the 3D range data. In the raw 3D data obtained from the Minolta, the x and y values are not evenly spaced across the image. The spacing varies with the distance between the subject and camera. Thus one preprocessing step is that the x and y distance between two pixels is normalized to be same in all images, so that the size of a particular person's ear in pixels is constant across different range images of that person.

Figures 3.5(a) - 3.5(b) show the images of an ear taken at different times. It is obvious that the second ear appears larger than the first in the original images. But after normalization, the two ears shown in Figure 3.5(c) - 3.5(d) have the same size.

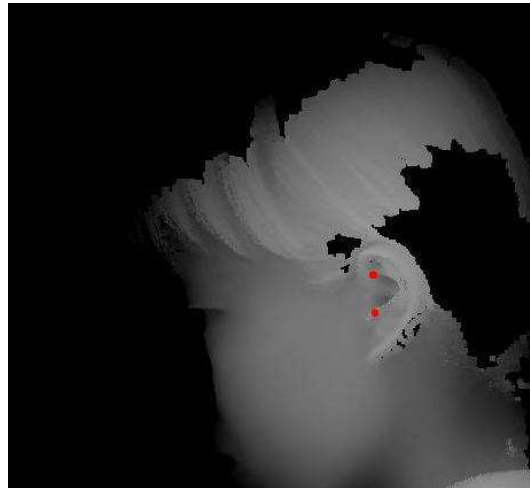
The third component of the 3D normalization is a hole filling process. There are two ways in which "missing" data can occur. The first is caused by the 3D sensor due to oily skin or lighting condition, and the second is because of the rotation of the data. Before rotation, the intervals between the x- neighborhoods and y- neighborhoods are almost evenly distributed. But after rotation, the interval between the x-, y- neighborhoods is distorted, especially when the point of view is changed significantly. From initial experiments, we decided to use a median filter to fill the holes caused by the sensor, and correct the rotation missing data by a mean filter. The window size for both operations is 5×5 . Figure 3.6 shows an image before and after hole filling. The problem of "missing" data after a 3D



(a) Original Image



(b) Rotation around the Normal of the Plane



(c) Rotation around Z axis

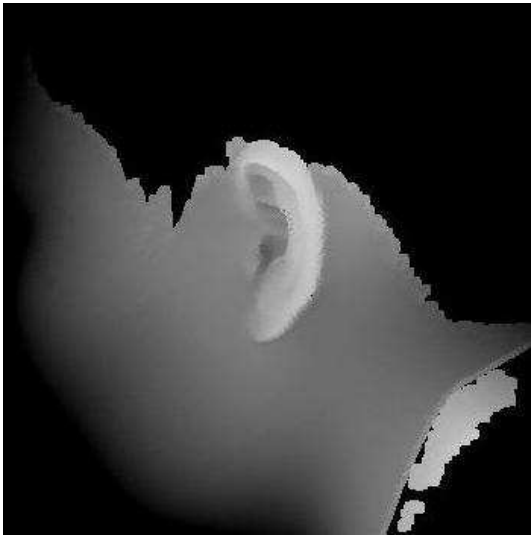
Figure 3.4. Two Step Rotation to Normalize Pose.



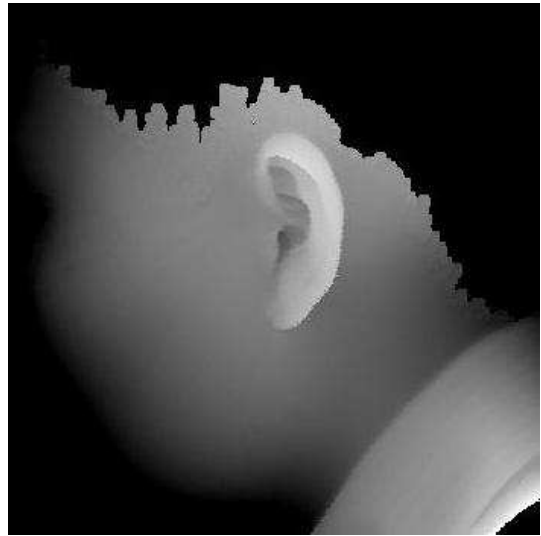
(a) Original Image 1



(b) Original Image 2



(c) After x, y Calibration of (a)



(d) After x, y Calibration of (b)

Figure 3.5. Pixel Size Calibration.

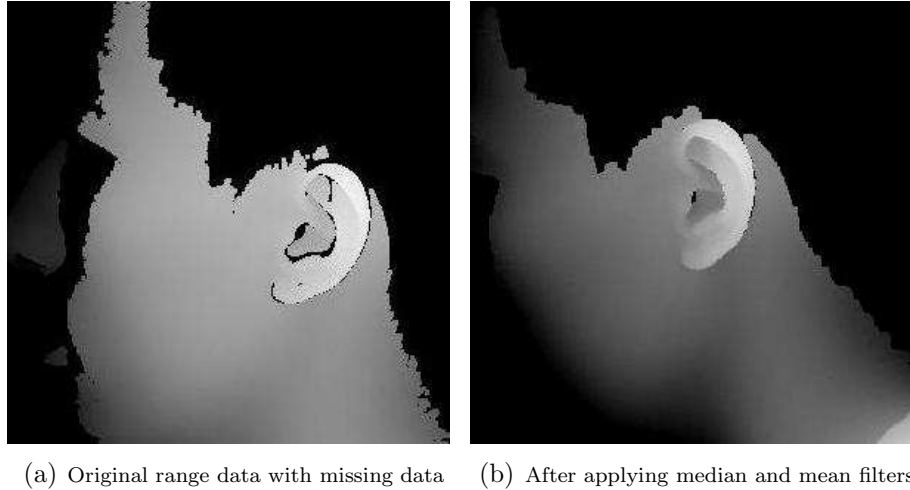
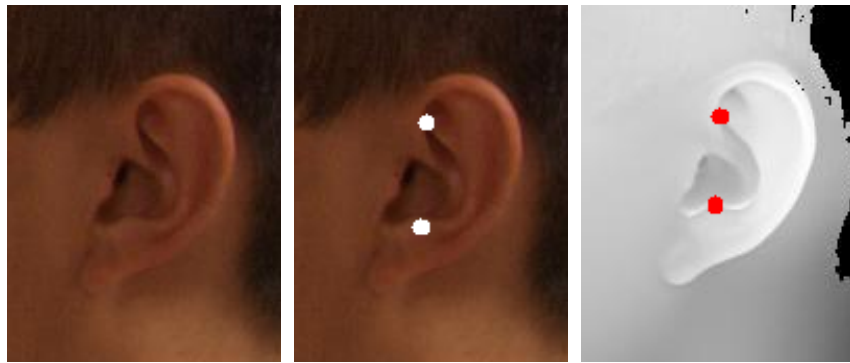


Figure 3.6. Hole Filling For 3D Range Data.

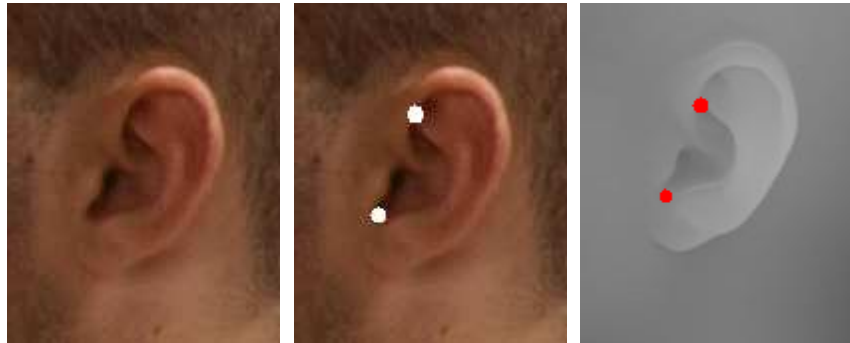
image is rotated to be viewed from a standard orientation is an inherent complication. In general, matching techniques must either “fill in” the missing data in some way, or allow for subset matching between datasets.

3.2.3 Landmark Selection

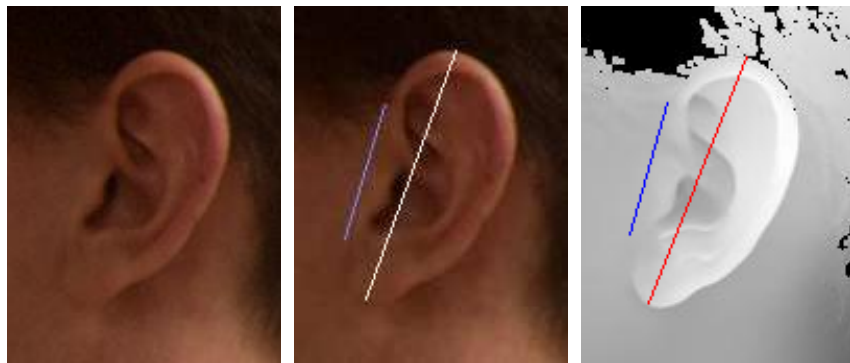
We have investigated three different landmark selection methods. The first is the two-point landmark described in a study of “eigen-ears” with 2D intensity images [14]. The upper point is the Triangular Fossa, and the lower point is the Antitragus [40], see Figure 3.7(a). However, we found that these two points are not easily detected in all images. For instance, many ears in our study have a small or indistinct Antitragus. In order to solve this problem, two other landmark methods were investigated. The second is similar to the first two-point landmark, but we used the Incisure Intertragica instead of the Antitragus as the second point, as shown in Figure 3.7(b). The orientation of the line connecting these two points is used to determine the orientation of the ear, and distance between them



(a) Landmark1: Using Triangular Fossa and Antitragus



(b) Landmark2: Using Triangular Fossa and Incisure Intertragica



(c) Landmark3: Using Two Lines

Figure 3.7. Example of Ear Landmarks.

is used to measure the size of the ear. The third method uses a two-line landmark, shown in Figure 3.7(c). One line is along the border between the ear and the face, and the other is from the top of the ear to the bottom. Unlike the two-point landmark, the two-line landmark promises to find the most part of the ear. In our experiments, the second method is adopted for further ear extraction in PCA-based and edge-based algorithm, since it is good at blocking out background and avoiding ambiguity. The two-line landmark is used in the ICP-based algorithm, since it is better suited to the ICP algorithm properties. ICP uses the real 3D range data in the matching procedure and the two matching surfaces should overlap. The two-line landmark gives the opportunity to extract the whole ear for matching, but at the same time, it always includes some background, which increases the background variation, and affects the PCA-based and edge-based performance.

3.3 Ear Extraction

Ear extraction is based on the landmark locations in the original images. The original ear images are cropped to (87x124) for 2D and (68x87) for 3D ears. The reason for different ear size for the 2D and 3D data is due to different pre-processing methods. The two-point landmark is used to calibrate both the orientation and scale of the 2D ear data, but only the orientation of the 3D ear data. For each extracted 2D ear, the two landmarks are mapped into the same locations in the image, and the ear size is scaled with the template size. But 3D data is in an absolute 3D space value, and the ear size different from person to person. When we used the same 2D ear template size for the 3D data, a larger ear background appeared in the cropped ear. Therefore, we shrank the template size for 3D ears to reduce the amount of background. The normalized images are

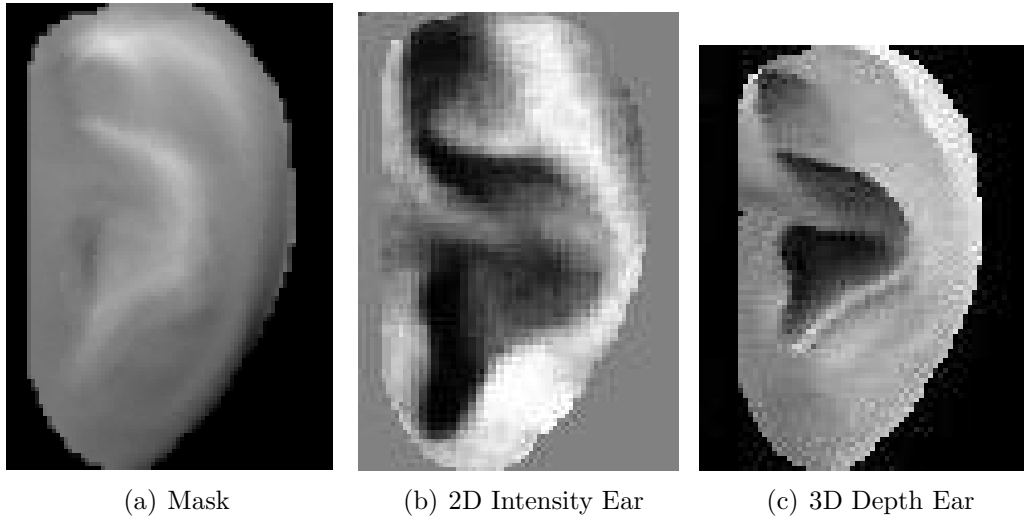


Figure 3.8. Examples of Ear Mask and Cropped 2D and 3D Ear.

masked to “gray out” the background and only the ear is kept. Figure 3.8 shows the mask and examples of the cropped 2D and 3D ear images.

CHAPTER 4

PCA-, EDGE- AND ICP-BASED APPROACHES ¹

Three approaches are compared in this chapter: PCA-based, edge-based and ICP-based. The PCA-based approach is experimented with using both 2D and 3D ear data, while the edge-based approach is applied to the 3D range image and the ICP-based approach to the 3D point cloud. The ICP-based approach has been extensively investigated with respect to both recognition accuracy and speed.

While we were first investigating these approaches, we had a dataset with just 202 subjects. After more image acquisitions, the dataset was expanded to the 302-subject version used in this chapter. All the experiments run on the 202-subject dataset are rerun on the 302-subject dataset to validate our conclusions.

4.1 PCA for 2D and 3D Ear Recognition

4.1.1 PCA ("Eigen Ear") Method

The PCA (Principal Component Analysis) based approach has been widely used in face recognition [13, 15, 56, 69, 71]. It was also used previously by Chang [14] in evaluation of 2D ear and face biometrics.

¹Sections 4.1 and 4.2 are based on the paper, "Ear Biometrics Using 2D and 3D Images", presented at the *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops: Advanced 3D Imaging for Safety and Security, 2005* [77]. Section 4.3 is based on the paper, "ICP-Based Approaches for 3D Ear Recognition", presented at the *Biometric Technology for Human Identification Conference, 2005* [81].

The eigenface approach for recognition was first proposed by Turk and Pentland [71]. They used PCA to create an eigenspace for all subjects in the database. The PCA subspace is defined by a covariance matrix formed by the training images. Each image contains $n = \text{rows} \times \text{columns}$ pixels, and these n pixels can be viewed as a column vector v . A set of m training images $T = \{I_1, I_2, \dots, I_m\}$ thus can be represented by a set of column vectors $V = \{v_1, v_2, \dots, v_m\}$. The covariance matrix Ω of V is defined as follows:

$$\bar{V} = \frac{1}{m} \sum_{i=1}^m v_i \quad (4.1)$$

$$A = V - \bar{V} \quad (4.2)$$

$$\Omega = AA^T \quad (4.3)$$

The eigenspace of the training images is spanned by the eigenvectors of Ω , $E = \{e_1, e_2, \dots, e_n\}$. Usually the eigenvectors are sorted by their corresponding eigenvalues from large to small. In order to reduce the data dimension, only the top x eigenvectors E_x which represent the major variance in the dataset are used for the “face space” or in this case, the “ear space”.

The PCA recognition algorithm is the nearest neighbor classifier operating on the eigen-subspace E_x . Given two test images I_1, I_2 , the similarity of I_1 and I_2 is measured by the distance between the projections of two images onto the eigen-subspace, I'_1 and I'_2 .

$$I'_i = E_x(I_i - \bar{V}) \quad (4.4)$$

There are several possible measurements of the distance between I'_1 and I'_2 , and a detailed description can be found in the manual of [7]. In our experiments,

a standard PCA based algorithm is used. Figure 3.8 shows an example of the images we used for PCA.

For each of the 302 subjects, the earliest good quality 2D and 3D images are used for the 2D and 3D ear space galleries, respectively. The latest good quality images are used as probes. For PCA-based algorithms, eigenvalues and eigenvectors are computed from the images in the training set. In our experiment, the training set is the set of gallery images. The “ear space” is picked out from the eigenvectors corresponding to all the eigenvalues. Starting with all of the eigenvectors, we drop eigenvectors one at a time from the beginning of the list until the performance gets worse. Then we also drop eigenvectors from the end of the list using the same procedure. The best rank-one recognition rate for 2D ear data is 63.6% when dropping the first 2 and the last 23 eigenvectors. The best performance for the 3D ear data is 55.3% when dropping the first two eigenvectors. The Yambor Angle [26] distance metric is used. The Euclidean and MahCosine distance were also tested but gave lower performance.

4.1.2 2D Ear Data

Two versions of the same gallery/probe datasets with different scaling of the ear sizes are examined on 2D data. One is set as the actual size of the ear, and the other is set at 1.25 times the size of ear (see Figure 4.1).

Figure 4.3 shows the performance of PCA using regular 2D ear size (Figure 4.1(a)). The performance is lower than that reported by Chang in his study of 2D “eigen-ears” [14]. Looking closely at the images created from the eigenvectors associated with the 3 largest eigenvalues (Figure 4.2(a)), it was apparent that each of them had some space behind the contour of ear. It is hypothesized that this

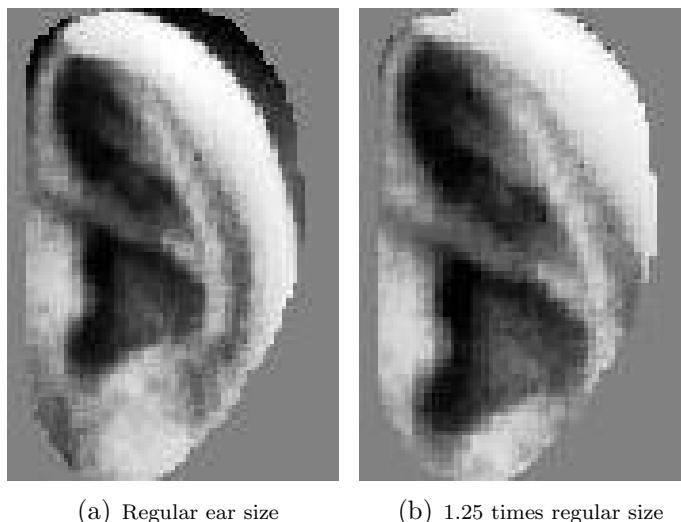
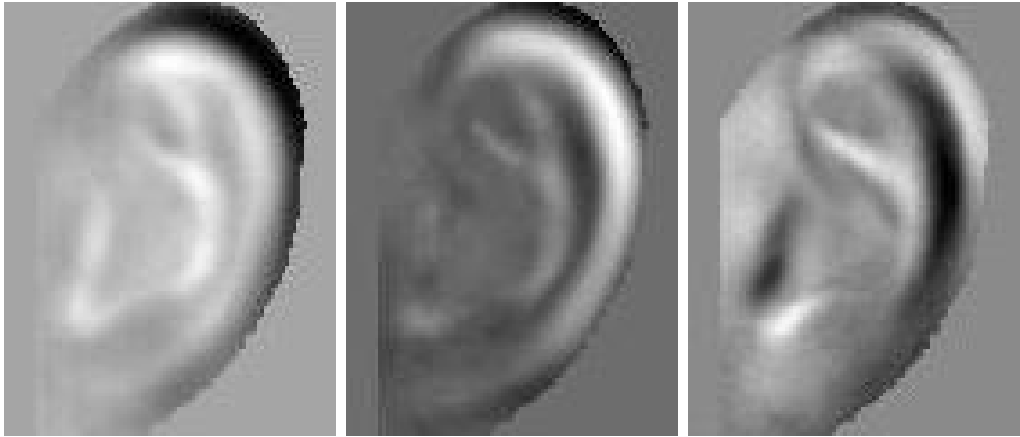


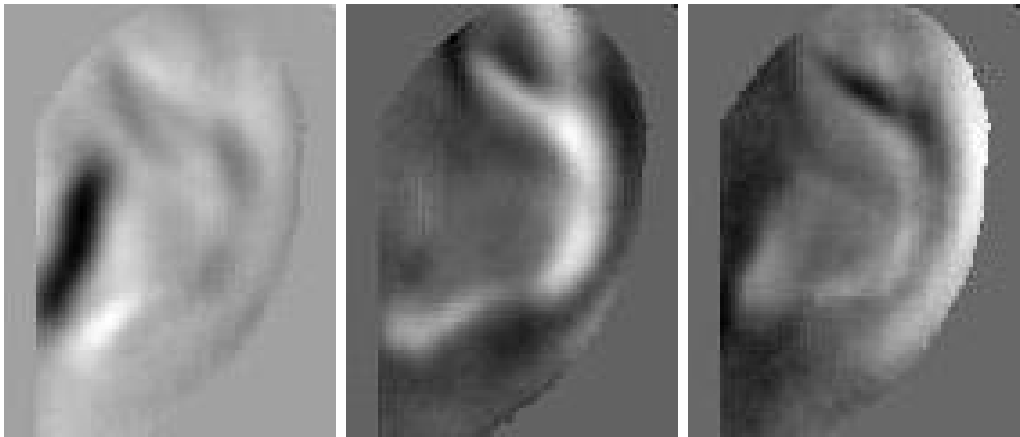
Figure 4.1. Experiments Using Different 2D Ear Size.

space caused the decrease in performance. After enlarging the ear to 1.25 times the original size (Figure 4.1(b)), there was no space behind the contour of ear in Figure 4.2(b). The rank-one recognition rate increased from 66.9% to 71.4% when using 202 subjects. Using the enlarged ear, the performance was 63.6% when using 302 subjects, as shown in Figure 4.10. Here, in our eigen-ear results, we go from 76.1% for a 88-person dataset to 63.6% with 302 subjects.

Chang obtained 73% rank-one recognition with 88 persons in the gallery and a single time-lapse probe image per person [14]. Our rank one recognition rate for PCA-based ear recognition using 2D intensity images with first 88 persons is 76.1%, which is similar to the result obtained by Chang, even though we used a completely different image data set acquired by a different sensor and different landmark points. Lower performance in the presence of a larger gallery is a well known general effect in face recognition [57].



(a) Eigen images of regular ear size



(b) Eigen images of enlarged ear size

Figure 4.2. Eigen Ear Images Associated With 3 Largest Eigenvalues.

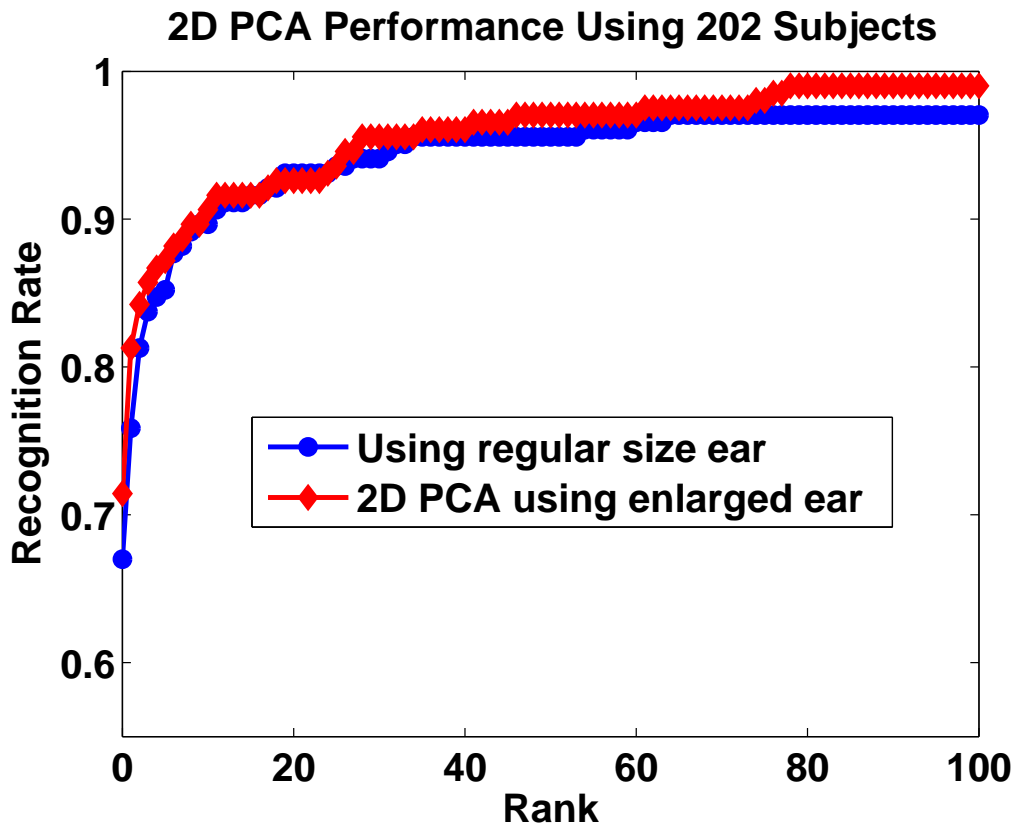


Figure 4.3. PCA Performance on 2D Ear Data Using Different Ear Size.

4.1.3 3D Ear Data

Due to the range data preprocessing, the absolute distance between two pixels in the X or the Y direction is same over all the images. Therefore, no scale process is applied in the 3D ear extraction. Also, pre-alignment uses the landmark to align the ear with same orientation. Two different experiments were conducted on the 3D ear data. One is using the original ear range data. The other is applying mean and median filters on the original data to fill the holes of the cropped ear (see Figure 3.6). The rank-one recognition rate is improved from 58.4% to 64.8% with hole-filling when using 202 subjects. This is still not very good in an absolute sense. One possible reason is that the ear structure is quite complex, and so using mean and median filter alone might not be good enough to fill holes in the 3D range data. Applying hole filling on the 302 subjects, the performance stays at 55.3% rank one recognition rate.

4.2 Hausdorff Range Edge Matching

4.2.1 Hausdorff Distance

Hausdorff distance (HD) [39] is an appropriate metric for 2-D object recognition and view-based 3-D recognition. Given as input the binary edge images of a model image and a scene image, the algorithm computes the Hausdorff distance between all possible relative positions of two images [76].

Given two point sets $A = \{a_1, a_2, \dots, a_m\}$ and $B = \{b_1, b_2, \dots, b_n\}$, the Hausdorff distance between set A and set B is defined as:

$$H(A, B) = \max(h(A, B), h(B, A))$$

where

$$h(A, B) = \max_{a \in A} \min_{b \in B} \| a - b \|$$

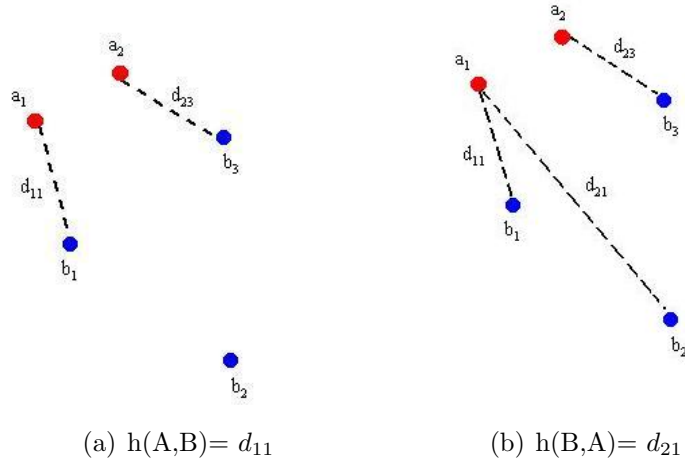


Figure 4.4. Two Point Sets Illustrating the Hausdorff Distances.

Figure 4.4 shows an example of Hausdorff Distance of point sets A and B , where $A = \{a_1, a_2\}$ and $B = \{b_1, b_2, b_3\}$. Also the algorithm considers simple transformations such as rotation or scaling. Without loss of generality, it allows only set B to translate and lets set A remain unchanged. The Hausdorff distance under translation is measured as:

$$M_T(A, B) = \min_t H(A, \text{Translate_of}(B))$$

where H is the Hausdorff distance as defined in equation (1), and *Translate_of* is a function used to implement scaling or rotation of set B . After the set B has been translated, especially by the scaling, we might lose some property which is true of the original image. In our experiments, we align all the ear images into same size and orientation, therefore rotation is not considered in the Hausdorff matching.

Achermann and Bunke [3] use an extension of the Hausdorff distance matching for the 3D face registration. Instead of using original 2D Hausdorff distance, they introduce a 3-D version of the partial Hausdorff distance. All the computation is based on the 3D space. In our experiment, the matching is between two edge images, therefore, only 2D Hausdorff distance is computed during the procedure.

4.2.2 Hausdorff Matching with Ear Image

In order to use Hausdorff distance to measure the similarity of two ears, we consider the binary edge image of both 2D and 3D depth ear image. We noticed that the 3D depth data looks much “cleaner” than the 2D intensity data. Figures 4.5(a) and 4.5(b) show the 2D and 3D images taken on two different days of the same person’s ear. The Canny edge detector with the same parameters is applied to those 2D and 3D ear data, and the edge images are shown in Figure 4.5(c) and 4.5(d). Here, single isolated edge pixels are eliminated from the edge images. It is obvious that edges from the range image are much more connected and smoother than in the 2D edge images. This is the motivation to develop an edge-based Hausdorff distance method using the range image.

The model images are the edge images from the gallery, and the test images are the edge images from the probe. The Canny edge detector is applied on all of the model and test images. The parameters are set to $\sigma = 1.00$, $T_{low} = 0.50$, $T_{high} =$

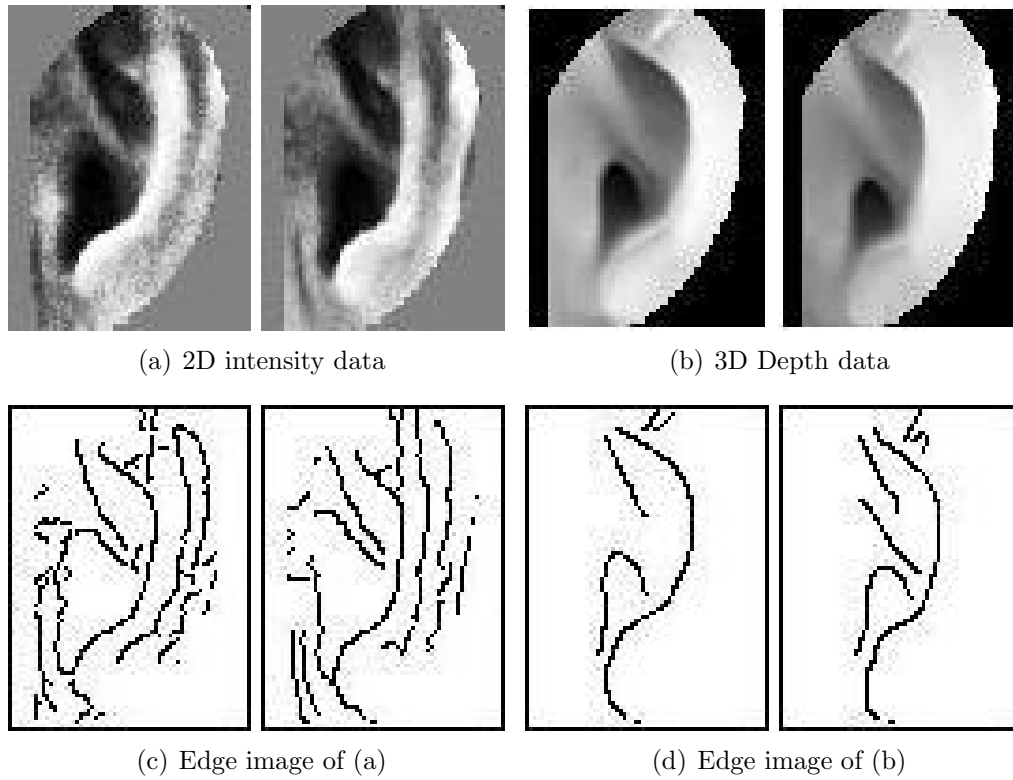


Figure 4.5. 2D and 3D Ear Images and Associated Edge Images. Canny Edge Detector Parameters are $\sigma = 1.0$, $T_{low} = 0.5$, $T_{high} = 0.5$.

0.50. Also only the inner edges of an ear is used. The outside contour is from the mask, and it is the same over all the images, therefore we drop it, as shown in Figure 4.5(d). By applying the Hausdorff algorithm, the maximum distance D between two ear edge images is obtained. Given a model image and a test image, let the model image remain fixed, move the test image $\pm D$ pixels in x direction and $\pm D$ pixels in y direction. At each position, the similarity between the test image and model image is computed. The maximum similarity in this $D \times D$ range is used as the measurement for matching.

The similarity is evaluated by the matching rate. Given an edge pixel in the model image, if there are edge pixels within the threshold of T_{match} pixels in the test image, then the pixel is a match. Here $T_{match} = 3$. The matching rate = $\frac{\text{number of matching pixels}}{\text{total edge pixels in model image}}$. The forward matching rate is the matching rate from test image to model image, and the backward matching rate is the matching rate from model image to test image. The forward and backward matching rate are combined for the similarity measure. In our experiments, different weightings on the forward and backward matching rate were tested, shown in Table 4.1.

There is no significant difference between various combinations of weighting between 0.3 and 0.7. Applying a forward matching weight of 0.6 and a backward matching weight of 0.4 on 302 subjects, the rank one recognition rate achieved is 67.5%, which is significantly better than the 3D PCA performance, shown in Figure 4.6.

TABLE 4.1

Recognition Rate Using Different Combinations of FM and BM.

Forward Matching	Backward Matching	Recognition Rate
0.7	0.3	142 out of 202
0.6	0.4	148 out of 202
0.5	0.5	146 out of 202
0.4	0.6	144 out of 202
0.3	0.7	142 out of 202

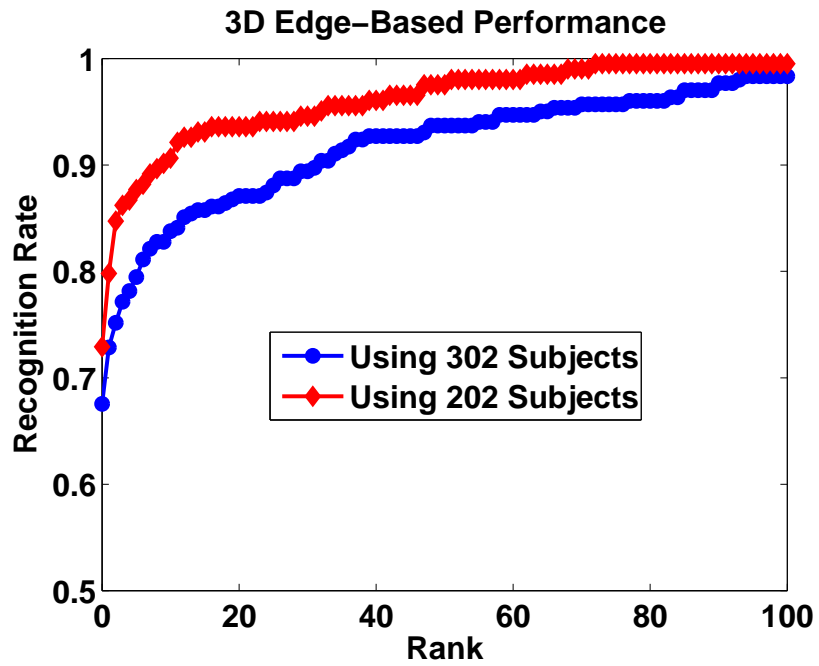


Figure 4.6. Hausdorff Ear Recognition Performance.

4.3 ICP-Based Approach

4.3.1 ICP Algorithm

Given a set of source points P and a set of model points X , the goal of ICP is to find the rigid transformation T that best aligns P with X . Beginning with a starting estimate T_0 , the algorithm iteratively calculates a sequence of transformations T_i until the registration converges. At each iteration, the algorithm computes correspondences by finding closest points, and then minimizes the mean square difference between the correspondences. A good initial estimate of the transformation is required, and all scene points are assumed to have correspondences in the model. The centroid of the extracted ear is used as a starting point in our experiments in this chapter.

The general ICP algorithm requires no extracted features, or curvature computation [5]. The only preprocessing of the range data is to remove the outliers. In a 3D face image, the eyes and mouth are common places to cause holes and spikes. 3D ear images do exhibit some spikes and holes due to oily skin or sensor error, but much less than in the 3D face images. Our initial experiment does not have outlier removal. We also consider a version of ICP that does some outlier removal as part of the algorithm.

4.3.2 Initial Experiment

Our initial experiments used the ICP implementation in the VTK library [65]. This implementation uses an oct-tree data structure. In the first baseline experiment, we use the same template size on both the gallery and probe ear images, and it gives us a 74.8% rank one recognition rate. When a smaller ear template is used on both gallery and probe ears, the performance increases to 79.7%. The

smaller ear template helps to exclude some background points, which improves the ICP performance. Given a starting registration, the ICP process is guaranteed to converge to a local minimum if the set of source points is a subset of the set of the destination points [5]. To give the probe a better chance to be a subset of the gallery ear, we used a larger template for the gallery ear, and a smaller one for the probe ear. This improves the rank one recognition rate to 85.1%, as shown in Table 4.2.

Another issue for ICP is computation speed. It takes 25-35 minutes to compute one probe ear against 202 gallery ears if all the ear points are used. The average number of data points for an ear is 5272 in our database, based on imaging the ear with the Minolta Vivid 910 with the “Tele” lens. Sub-sampling of the original range data will reduce the computational time significantly. Instead of using all the points, every other row and column in the 3D image is selected. After sub-sampling, the average number of the points on the ear is reduced to 1186. If both gallery and probe ears are sub-sampled, it only takes 3-5 minutes to compute one probe ear against all the gallery ears, but the rank one recognition rate decreased to 83.7%. If only probe ears are subsampled, it takes 5-8 minutes for the same computation, and the rank one recognition rate stays at 84.7%.

4.3.3 Noise Removal

Given a profile image, it is very difficult to isolate the ear without any background and noise around it. This problem will affect the ICP performance. One observation is that the noise mostly occurs on the top part of the ear. The bottom part of the ear is relatively clean, except when an earring appears. The blue line in the truthwriting (on the right in Figure 3.7(c)), which goes through the ear

TABLE 4.2

The ICP Performance by Using Different Masks for Gallery and Probe.

* The time is for matching one image against 202 images in the gallery.

Initial Experiments with 202 Person Dataset	
Variation of ICP Algorithm	Rank One Recognition Rate
Mask for max ear size for gallery and probe	74.8%
Mask for min ear size for gallery and probe	79.7%
Additional reduced mask for probe	85.1% (25-35 minutes*)
Subsample points for probe only	84.7% (5-8 minutes*)

from top to the bottom, defines the bottom boundary of the ear clearly. Taking advantage of the fact that the ear edge is a continuous curve, we start from the bottom point, and use a seed-growing method to trace the ear edge and eliminate the noise. The big jump in the z depth value indicates a new ear edge point. Also, this edge point should connect to the old edge point, which is defined in the previous step. Figure 4.7(d)- 4.7(h) shows that this method effectively gives us much “cleaner” images, although sometimes when the hair is too close to the ear, this method fails. Figure 4.7(i)- 4.7(j) shows some missed cases. Rank-one recognition on the 302 person dataset increased from 84.1% to 87.7% with this additional step.

4.3.4 Speed Limitation

It is well known that the basic ICP algorithm is effective but time consuming for 3D object registration [4, 22, 31, 32, 44, 50, 63]. In order to make it more practical, it is necessary to speed up the algorithm. Two steps which are intended

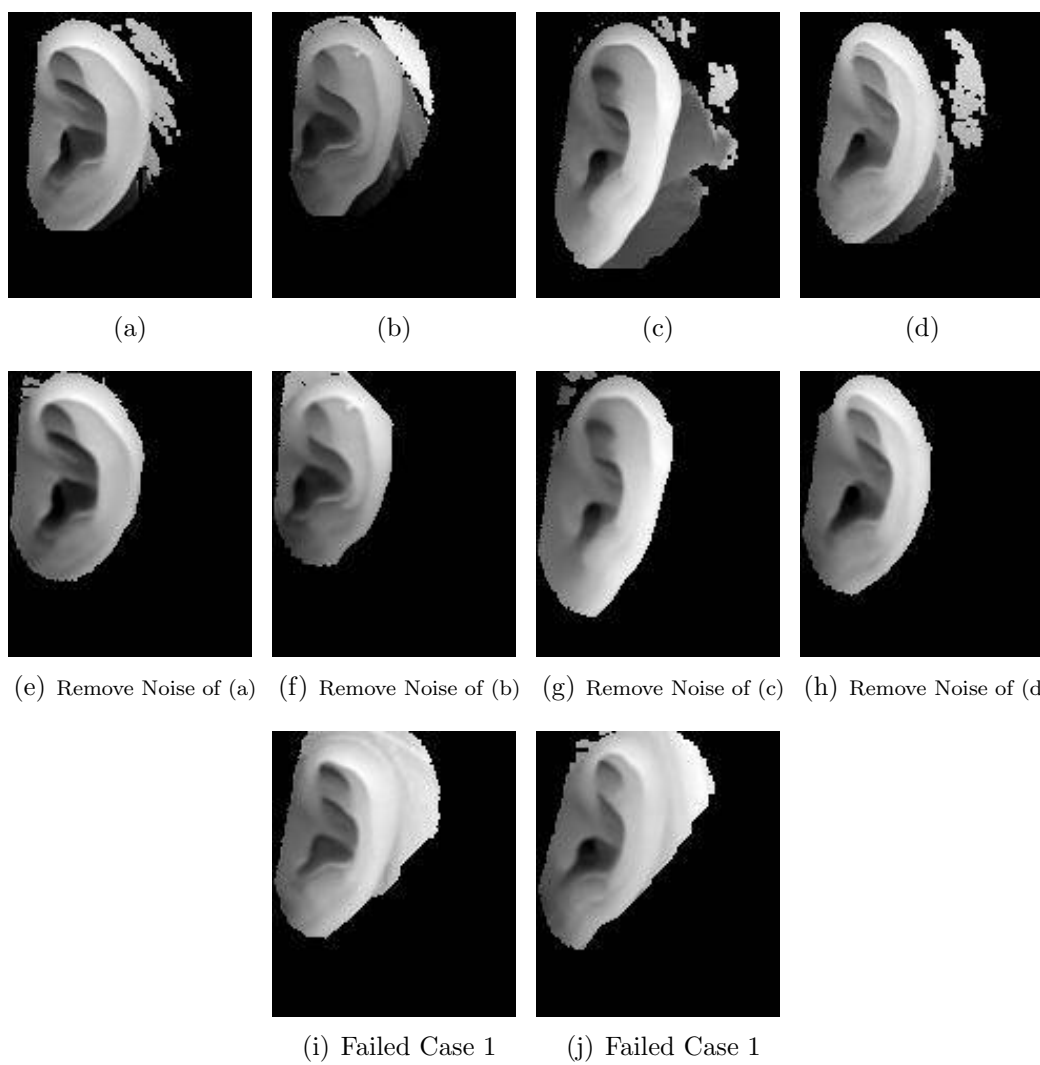


Figure 4.7. Examples of Noise Removal.

to make the algorithm faster are considered in this section. One is to control the number of iterations, and the other is to use appropriate data structures to shrink the running time for each iteration.

The number of iterations is initially set as 50, but we found the error distance decreases much faster in the first few iterations than in the later iterations. So instead of using a fixed number of iterations, we measure the drop in the average distance between paired points between two consecutive iterations. Using a threshold of 0.0001 for the average point difference in mm, the average of the number of iterations decreases from 50 to 25.74, and the performance stays the same.

In their experiments, Besl and McKay reported that 95% of the running time is attributed to determining the corresponding points. Our initial ICP implementation used an octree as the data structure. Following the suggestion of [5, 63, 67], a k-d tree data structure was implemented. When using ICP to register two surfaces T and S, the total time includes two major parts. One is to build a tree, and the other is to find the closest point in the surface T for each point in surface S. The search for the closest point can be speed up by using a tree structure to store the point set. The traversal time for both octree and kd-tree are $\log n$, where n is the number of cells in the space after subdivision. Since our 3D data can have spikes and noise, it is very easy for the octree to be unbalanced, but the k-d tree will not suffer from this.

Table 4.3 shows that the ICP-based recognition performance using octree and kd-tree ICP are essentially the same. The reason why octree and kd-tree have slight different results is due to the different data structures. The octree subdivides the 3D space into small cells, and builds the tree upon these. When searching the

TABLE 4.3

ICP Performance by Using Octree and K-d Tree. The performance shows the rank one recognition rate, and the run time is the CPU time to register one source surface to one target surface.

Subsample probe	Octree		K-D Tree	
	Performance	Run Time (s)	Performance	Run Time (s)
1	87.1%	22.3	86.7%	3.41
2	87.4%	3.82	87.1%	0.98
4	86.4%	1.34	87.7%	0.61
8	75.2%	0.68	73.2%	0.48
16	14.6%	0.52	16.2%	0.40

closest point, the octree attempts to find the closest cell to the given point and returns the point that is a projection from the given point to the plane created by all the points in that closest cell. The comparison between the ICP performance using Octree and K-d tree is presented in Table 4.3.

4.3.5 Outlier Elimination

When using the ICP algorithm to align two surfaces, the quality of alignment depends highly on selecting good pairs of corresponding points from two surfaces. When outliers or missing points occur, their corresponding points will distract the alignment and generate a wrong transformation matrix. For the ear biometric, hair is the most common cause of outliers, and sometimes minor hair covering part of the ear is inevitable. Therefore outlier elimination becomes a requirement for our system.

An “outlier” match can occur when there is noise in one of the two point sets or when there is a poor match. To improve performance, outlier elimination is added to the original ICP implementation. Outlier elimination is accomplished in two stages. During the calculation of the transformation matrix, we focus on eliminating the spike noise [42, 84]. The approach is based on the assumption that for a given noise point p on the probe surface, the distance from p to the associated closest point g_p on the gallery surface will be much larger than the average distance. For each point p on the probe surface, we find the closest point g_p on the gallery surface. Let $D = d(p, g_p)$ represent the distance between the two points. Only those pairs of points whose D is less than a threshold are used to calculate the transformation matrix. Here the threshold is set as mean distance $+ \frac{R^2}{2}$, where the mean distance is the average distance of all point pairs, and R is the resolution of the gallery surface.

The second stage occurs outside the transformation matrix calculation loop. After the first step, one transformation matrix is generated to minimize the error metric. We apply this transformation matrix on the source surface S and obtain a new surface S' . Each point on the surface S' will have a distance to the closest point on the target surface. We sort all the distance values, and use only the lower 90% to calculate the final mean distance. Other thresholds have been tested, but 90% gives the best performance, which is consistent with the experience of other researchers [59]. The two-stage outlier elimination method improved rank-one recognition from 87.7% to 91.4%.

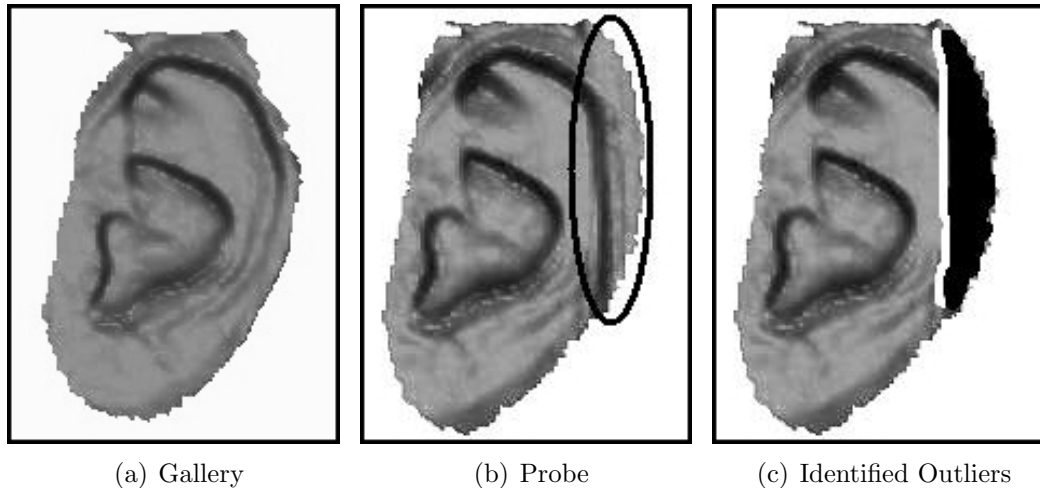


Figure 4.8. Examples for Outlier Elimination.

4.3.5.1 Outlier Elimination Using AV Distance Metric

Since both outlier rejection and using the AV distance metric result in higher performance, we combine these two approaches, by using the absolute distance as the threshold. The rank-one recognition improved from 87.7% to 97.4%. This might be because absolute value is more robust when outliers appear in the set of correspondences.

Figure 4.8 shows an example of one pair of ears which is not correctly recognized before using outlier elimination, but is correctly recognized after the addition of this procedure. There is some hair around the right top edge of the ear which affects the mean distance before using outlier elimination.

4.3.6 Point to Point vs. Point to Triangle Approach

Two approaches are considered for the ICP algorithm to match the points, point-to-point [5] and point-to-triangle [20]. In the point-to-point approach we try to find the closest point on the target surface. In the point-to-triangle approach we

use the result from the point-to-point algorithm. After obtaining the closest point on the target surface, all the triangles around this point are extracted. Then the real closest point is the point on any of these triangles with the minimum distance to the source point. From our experiments, it seems that each of the approaches has pros and cons, and there is no clear winner when considering all situations.

As shown in Table 4.4, the point-to-point approach is fast, and accurate when all the points on the source surface can find a good closest point on the target surface. But if the gallery is subsampled, the point-to-point approach loses accuracy. Since the probe and gallery ear images are taken on different days, they vary in orientation. When both gallery and probe images are subsampled, it is difficult to match points on the probe surface to their corresponding points on the gallery surface. This serves to increase the overall mean distance value. But this approach is much faster than point-to-triangle.

On the other hand, the greatest advantage of the point-to-triangle approach is that it is accurate through all the different subsample combinations. Even when the gallery is subsampled to every fourth row and column, the performance is still acceptable. This behavior comes at a substantial computational expense. And in our experiment, where we just take one scan of the ear, the presence of occlusion is unavoidable. Therefore point-to-triangle will certainly give us more accurate error distance.

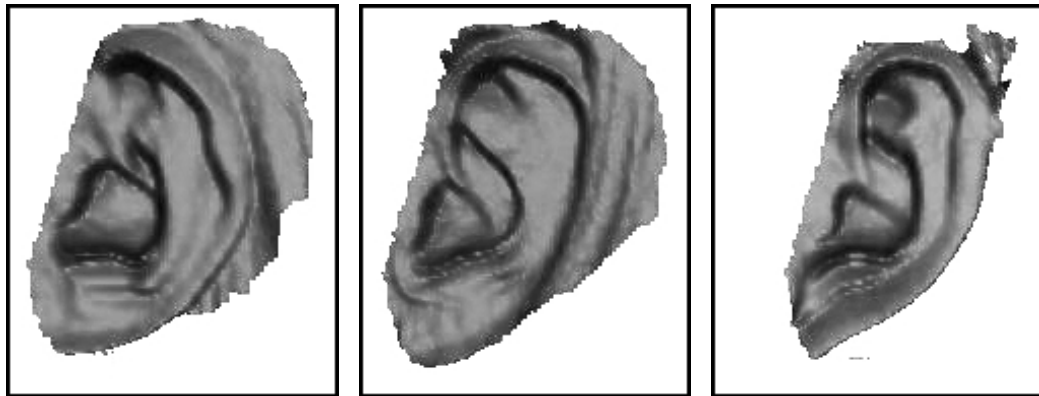
Our final algorithm attempts to reduce the tradeoff between performance and speed. The point-to-point approach is used during the iterations to compute the transformation matrix. One more point-to-triangle iteration is done after obtaining the transformation matrix to compute the error distance. We have learned that using a gallery surface that is denser than the probe surface will sometimes

yield better performance. The final results reflecting the revised algorithm are shown in Table 4.4. After using the revised ICP implementation, when gallery and probe are both subsampled, the performance is close to the point-to-triangle approach, but requires significantly less computation time.

Table 4.4 leads to two conclusions. The first is that when the gallery and probe surfaces have similar resolution, the revised ICP is fast and reasonably accurate. The second is that when the gallery surface is much denser than the probe surface, point-to-point is both fast and accurate. The highest rank-one recognition is obtained with the point-to-triangle matching and no subsampling of the data. However, this is computationally expensive. A more practical result is obtained with our revised algorithm and subsample by 4 on the probe. Figure 4.9 shows the four mis-matches on the 302-subject dataset.

TABLE 4.4
ICP Performance by Using Point-To-Point, Point-To-Triangle and
Revised Version.

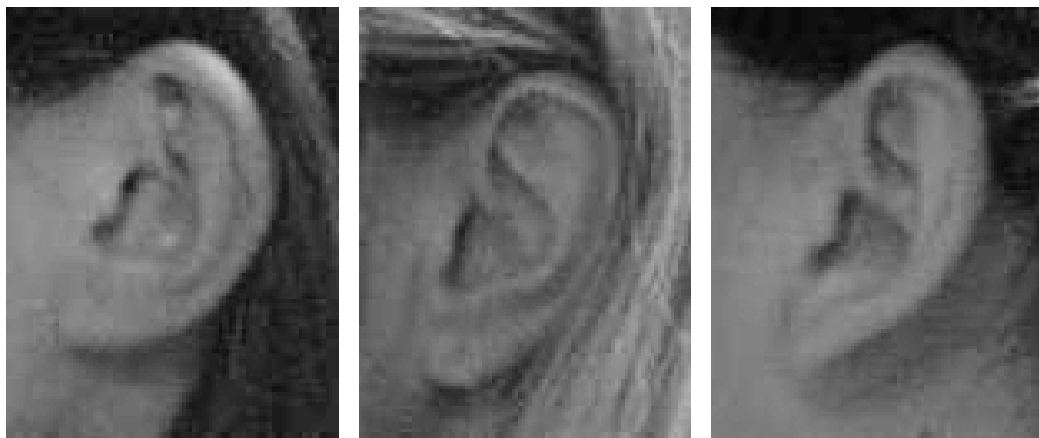
G = Gallery	Point-To-Triangle		Point-To-Point		Revised version	
	Perf.	Run Time	Perf.	Run Time	Perf.	Run Time
G1P1	98.7%	4-5 hrs	95.7%	20-25 mins	95.4%	20-25 mins
G1P4	97.7%	20-25 mins	97.4%	1-2 mins	97.4%	1-2 mins
G2P4	97.7%	6-7 mins	94.7%	0.5-1 min	96.4%	0.8-1 mins
G4P4	94.7%	1 min	74.5%	0.2-0.3 min	89.4%	0.3-0.4 min



(a) Hair Covered

(b) Hair Covered

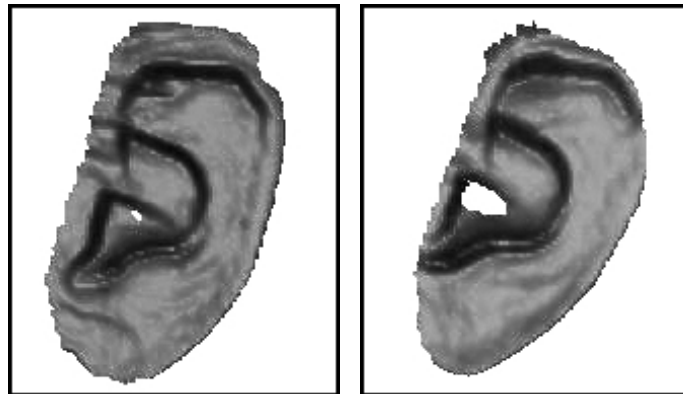
(c) Hair Covered



(d) 2D Image of (a)

(e) 2D Image of (b)

(f) 2D Image of (c)



(g) Gallery

(h) Probe

Figure 4.9. Four Incorrect Matches.

4.4 Statistical Significance Testing

4.4.1 Z-test with Bonferroni Correction

Four single-biometric experiments were explored extensively in the previous sections, represented as CMC curves in Figure 4.10. The ICP-based approach has the highest performance, followed by the 3D edge-based approach, then followed by PCA approach on 2D intensity images, and PCA on the 3D range images. Another 3D ear recognition method due to Bhanu and Chen [8] was initially considered but dropped in favor of the other methods described. In order to analyze the performance differences between methods, statistical significance tests were conducted.

A test for significance of difference in observed rank one recognition rate can be addressed as a binomial distribution problem. The correct matching rate is the probability of success p and incorrect matching rate is the probability of failure q , where $p + q = 1$. When the sample size becomes larger, the binomial distribution begins to converge to a normal distribution. That is, for a large enough sample size N , a binomial variable X is approximately to $N(Np, Npq)$. Fairly good results are usually obtained when $Npq \geq 3$. Here, \hat{p} is the proportion of observed correct matches. \hat{p} value for each method is shown in Table 4.5. In our circumstance, sample size $N = 302$, with all $N\hat{p}\hat{q} \geq 3$.

Given two methods, with sample size as N_1 and N_2 , and proportion of observed correct matches as \hat{p}_1 and \hat{p}_2 , the test statistic for $H_0 : p_1 = p_2$ is

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\left(\frac{N_1 + N_2}{N_1 N_2}\right) \left(\frac{X_1 + X_2}{N_1 + N_2}\right) \left(1 - \frac{X_1 + X_2}{N_1 + N_2}\right)}}$$

where $X_1 = \hat{p}_1 \times N_1$ and $X_2 = \hat{p}_2 \times N_2$.

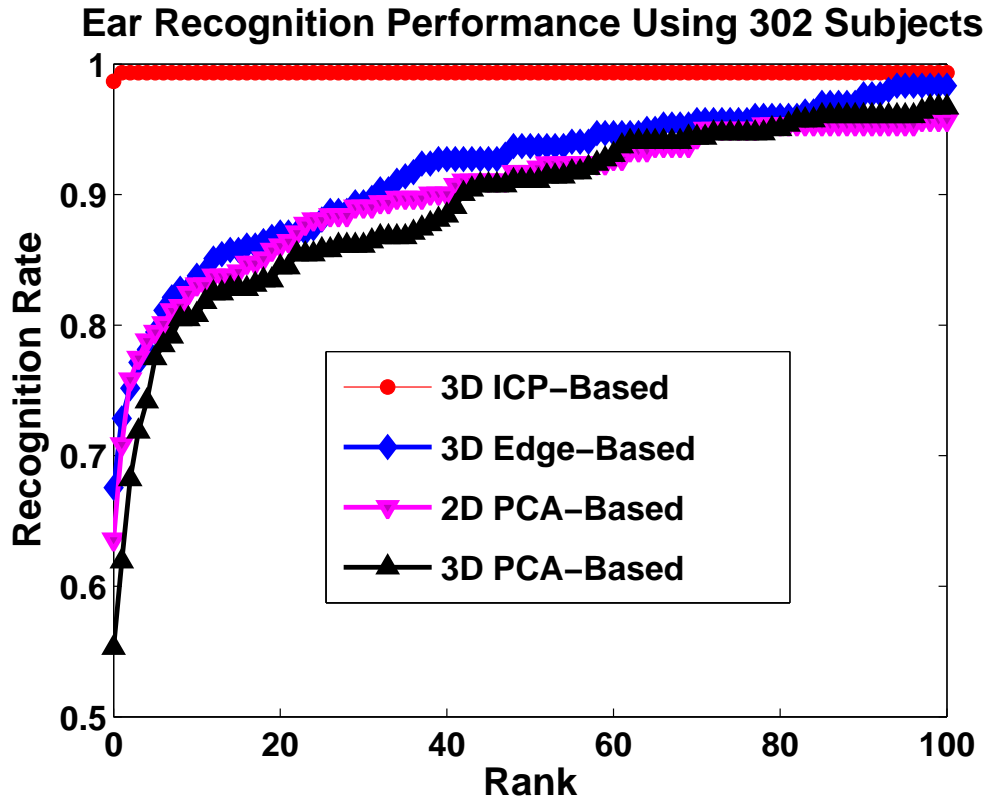


Figure 4.10. Performance of Different Approaches.

TABLE 4.5

Proportion of Observed Correct Matches.

	ICP	2D PCA	3D PCA	Edge-based
\hat{p}	0.987	0.636	0.553	0.675
\hat{q}	0.013	0.364	0.447	0.325
$N\hat{p}\hat{q}$	3.87	69.91	74.65	66.25

Table 4.6 is constructed using 0.05 level of significance. It is well known that significance levels from the pairwise comparisons might be misleading. Simply, this means that when too many comparisons are carried out, there exists the probability that the result suggests a statistically significant difference even if no difference exists. And this kind of error is called type I error.

The Bonferroni correction has been used post hoc to determine the significance of multiple tests [33]. It basically multiplies each of the significance levels from the z test by the number of tests performed. If this value is greater than 0.05, a significance level of 0.05 is used. This obviously makes it harder to claim a significant result, and reduces the chance of making a Type I error to an acceptable level.

TABLE 4.6

Statistical Test of the Difference between Performance, Using 0.05 level of significance. H_0 : There is no difference in performance between the two methods. (*: It changes from reject to accept after Bonferroni adjustment).

	Edge	2D PCA	3D PCA
ICP	11.03(Reject)	12.67(Reject)	10.23(Reject)
Edge		1.01(Accept)	3.08(Reject)
2D PCA			2.08(Accept*)

The “reject” in Table 4.6 refers to rejecting the null hypothesis of no significant difference. The performance of the ICP-based algorithm is statistically significantly better than the other three methods. The edge-based performance is statistically significantly better than the 3D PCA-based method.

4.4.2 Friedman Test with Nemenyi Post-hoc Test

The z-test with following Bonferroni correction is one of the binomial tests. However, the binomial testing assumes that the samples are drawn from normal distribution. As noted by Salzberg in [64], when this assumption is not met, the binomial testing lacks the power of the better non-parametric tests and the Bonferroni correction is “overly radical”. Demšar suggested using the Friedman test [29, 30] with the corresponding pot-hoc tests [55], a pair of non-parametric tests, for statistical comparisons [25]. There is no proof whether our dataset is normally distributed. We were interested in determining whether there exists a difference between the z-test result and the Friedman test result. As we described, the z-test uses the rank-one recognition performance of each algorithm on a 302-subject dataset. For the Friedman test, it takes each subject (probe ear) as an observer, and looks at the rank of every algorithm on each subject.

The Friedman statistic is given by:

$$\chi_f^2 = \frac{12N}{k(k+1)} \left[\sum_j R_j^2 - \frac{k(k+1)^2}{4} \right] \quad (4.5)$$

As in [25], N is the number of observers, here $N = 302$; k is the number of algorithms to be compared. Assume r_i^j be the rank of the j -th algorithm on the i -th observer, $R_j = \frac{1}{n} \sum_i r_i^j$, which indicates the average rank of algorithm j . The null-hypothesis is that there is no difference in performance among algorithms.

If the null hypothesis is rejected, the Nemenyi post-hoc test [55] is applied. The performance of two algorithms i, j is significantly different only if their average ranks R_i, R_j differ by at least the critical difference (CR):

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}} \quad (4.6)$$

$$z = (R_i - R_j) / \sqrt{\frac{k(k+1)}{6N}} \quad (4.7)$$

$q_{0.05}$ is 2.569 and $q_{0.10}$ is 2.291 when $k = 4$.

Using the Friedman Test, $\chi_f^2 = 147.5966$, and we reject the null-hypothesis. The average rank for each algorithm and their pair-wise z value are shown in Table 4.7. Compared to the z -test results, the only difference is between 3D PCA and edge-based. The z -test with the Bonferroni correction detects the significant difference between performance of these two algorithms, while the Friedman test with the Nemenyi post-hoc test does not. For any statistical test in the following chapters, we will apply both methods.

TABLE 4.7

Friedman Test of the Difference between Performance, Using 0.05 level of significance. H_0 : There is no difference in performance between the two methods.

	Edge	2D PCA	3D PCA
Average Rank	= 2.5977	= 2.6689	= 2.7864
ICP = 1.9470	6.1936 (Reject)	6.8713 (Reject)	7.9897(Reject)
Edge = 2.5977		0.6777 (Accept)	1.7961 (Accept)
2D PCA = 2.6689			1.1184 (Accept)

CHAPTER 5

A FAST ICP-BASED APPROACH ¹

In a biometrics scenario, gallery images are enrolled into the database ahead of the matching step, which gives us the opportunity to build related information before the probe comes into the system. In this chapter, we present a novel approach, called “Pre-computed Voxel Closest Neighbor,” to reduce the computational time for shape matching in a biometrics context. The approach shifts the heavy computation burden to the enrollment stage, which can be done offline. Experiments in 3D ear biometrics with 369 subjects and 3D face biometrics with 219 subjects demonstrate the effectiveness of our approach.

5.1 Introduction

Since its introduction by Chen and Medioni [20] and Besl and McKay [5], the Iterative Closest Point (ICP) algorithm has been widely used for 3D shape matching [20, 31, 43, 83], and demonstrates appealing accuracy. It has been used in a wide range of application areas, including the integration of range images [34, 70] and alignment of CT and MR images [27]. Here, we are specifically interested in 3D shape matching for biometrics [17, 18, 20, 51, 52, 81]. The

¹This section is based on the paper, “A fast algorithm for ICP-based 3D shape biometrics”, presented at the *Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID 2005)* [79].

ICP algorithm is known to be computationally expensive. With two clouds of points or meshes, source S (probe) and target T (gallery), the complexity of a typical single ICP iteration is $O(N_S \log(N_T))$ in the expected case, where N_S is the number of points in the source and N_T is the number of points in the target. The improvement to $O(N_S \log(N_T))$ from $O(N_S \times N_T)$ results from using a K-D tree data structure [81]. Also the ICP algorithm iteratively finds the minimum distance between two surfaces. With N_I iterations, the overall complexity of ICP matching is $O(N_I N_S \log(N_T))$ [5]. Therefore, matching high-resolution images of both source and target surface leads to a heavy computational load. A fast ICP implementation is crucial for practical use in biometrics.

Nowadays, using shapes sensed by a 3D scanner is a major trend in biometrics [17, 18, 20, 51, 52, 74, 81]. Each scan yields a 3D shape that can be used as a representation of the subject. In this chapter, we illustrate our approach using both 3D ear and 3D face shapes. There are two types of image in a biometric application, gallery and probe. The gallery images are those that have been enrolled and are known to the system, while the probe images are those that need to be matched against the images in the gallery. In a recognition scenario, one probe is matched against all the images in the gallery, and the algorithm returns the best match with the minimum error distance. In a verification scenario, one probe is matched against just one gallery entry, the one enrolled for the claimed identity. In recognition or verification, enrollment occurs once and is followed by many instances of recognition. As we mentioned in the previous chapters, in developing 3D ear shape as a biometric, the gallery uses full resolution to achieve better recognition accuracy, which means that the computational expense depends heavily on the gallery data.

One special characteristic of this application is that all gallery images are already enrolled into the database before the matching takes place. Probe images are introduced into the system during the matching process. Taking advantage of the fact that the gallery images are enrolled prior to matching, we propose a novel method to accelerate the ICP matching. Our new method is called the “Pre-computed Voxel Closest Neighbor”. The idea is to voxelize a volume which can hold the 3D gallery surface, and for each voxel to pre-compute its distance to the 3D gallery surface and save this for future use.

In section 2 we review several fast ICP approaches. Then in section 3 we give details of our approach. Section 4 addresses the applicability of our approach by using the ear and face biometrics, and experimental results are presented and analyzed. Finally, section 5 discusses further refinements and possible future directions.

5.2 Literature Review

There have been a number of attempts to speed up ICP matching. One line of work is focused on fast algorithms for computing the nearest neighbor. The use of the k-d tree data structure appears to be the standard method in this area [5, 68]. Cleary and co-workers analyzed the “Elias” algorithm to search for a nearest neighbor in the n -dimensional Euclidean space [22]. They claimed that by using the “Elias” algorithm, the number of search points is independent of the total number of points on the surface. In [32], Greenspan *et al.* proposed a novel nearest neighbor algorithm for small point sets. They report that “Elias” is much faster than a plain k-d tree, and that the “spherical constraint” method improves the speed still further. Zinsser *et al.* analyzed the performance of the nearest

neighbor algorithm for ICP registration [85]. Their work is not limited to range images or triangular meshes, but also can be used with 3-D point sets generated by structure-from-motion techniques.

Benjiemaa [4] proposed a multi-z-buffer technique to accelerate the ICP algorithm. All points are projected in a z-buffer to perform the local search, and they claimed that this space partition speeds up the search for point-to-point correspondences. But in order for the multi-z-buffer technique to work properly, two surfaces need to be sampled with a high and uniform density.

Another line of work in this area looks at different sub-sample strategies to reduce computation time. One strategy is using multi-resolution approaches; that is, start with a coarse point set and use progressively finer point sets as the algorithm proceeds. The idea of the average distance between points in the current resolution in comparison to the average distance between matched points is the standard way to automate the switching between resolutions [50].

In [31], Gelfand *et al.* describe the importance of the quality of the point pairs. In the presence of noise or miscalibration in the input data, it is easy to create poor correspondences between pairs of points. Therefore, the least-squares technique might lead to wrong pose, or make it difficult for the algorithm to converge. They propose a technique to decide whether a pair of meshes has good quality by measuring the covariance matrix between two meshes which have been sparsely and uniformly sampled. This technique tries to avoid the unstable movement between two surfaces by sampling the features from the input data which are the best constraint for this kind of movement.

In [63], Rusinkiewicz and Levoy discussed the variants of ICP which affect all phases of the algorithm. They list most of these variants, and evaluate their effects

on the speed with which the correct alignment is reached. Also in the paper, they proposed a combination of ICP variants optimized for high speed.

Researchers have also looked at mixing the two lines of work, having some multi-resolution mixed with some constrained search for nearest neighbor. Jose and Hügli proposed a solution that combines a coarse to fine multi-resolution approach with the neighbor search [45]. The multi-resolution approach permits to successively improve the registration using finer levels of representation and the neighbor search algorithm speeds up the closest point search by using a heuristic approach. They claim this technique reduces the time complexity of searching from $O(N \log(n))$ to $O(N)$, while preserving the matching quality [44].

5.3 Fast ICP Matching for 3D shapes

The most time consuming part of the ICP algorithm is that for each point on the probe surface, the algorithm needs to find the closest point on the gallery surface. By using these pairs of corresponding points, the ICP algorithm iteratively refines the transforms between two surfaces, finding the translation and rotation to minimize the mis-match.

This search for a closest point on the gallery surface is initially done using a kd-tree, as described in [81], and each search takes $O(\log N_G)$, where N_G is the number of the points on the gallery surface. Our goal is to reduce this search time to a constant value. The main idea is that if we can pre-compute the distance from any point in the 3D space to the gallery surface, and use it when needed, then the search time for a closest point is a constant.

Our “precomputed voxel nearest neighbor” approach is illustrated on the application of matching 3D surfaces for biometric recognition. At the time of en-

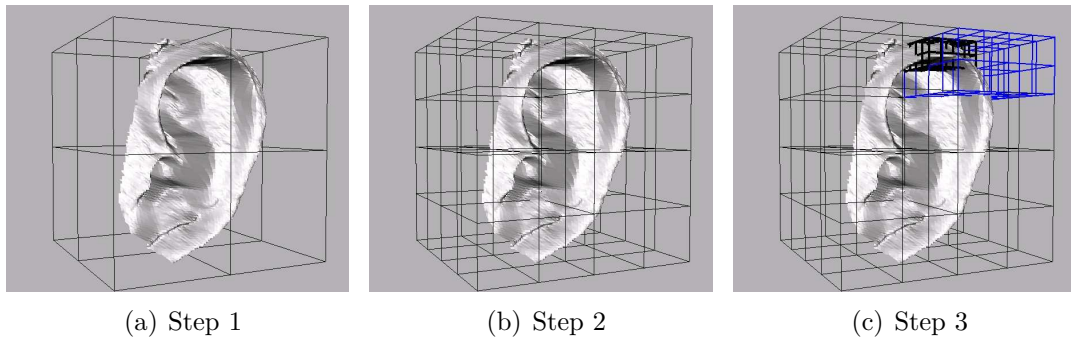


Figure 5.1: Voxelization of 3D Ear Data. In order to show it clearly, we present it from coarse to fine. In step 1 the volume is subdivided into 8 small voxels, and in step 2 each small voxel is subdivided into 8 even smaller voxels. And continue this subdivision until the size of each voxel is smaller than a threshold. (To implement this idea, we do not need to do it step by step, and instead we subdivide the volume once using a fixed voxel size).

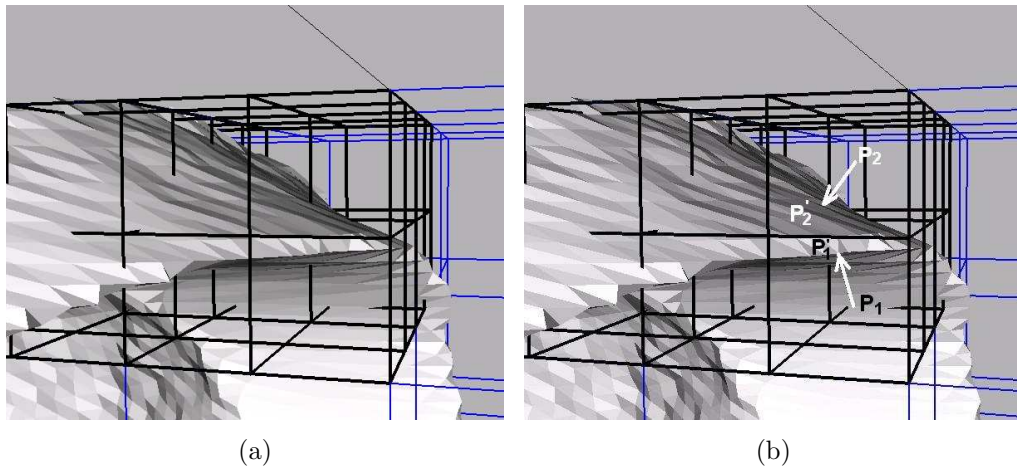


Figure 5.2: Close Look at Voxels and Example Distance Between Voxel and Gallery Surface. (P_1 is the center of the voxel 1 and the closest point on the gallery surface to P_1 is P_1' . P_2 is the center of the voxel 2 and the closest point on the gallery surface to P_2 is P_2').

rollment, the gallery 3D shape sits in a 3-D volume that we think of as a set of voxels. In our experiment, the volume size depends on the size of the biometric source, face or ear. The detailed explanation will be given in next section. Figure 5.1 illustrates how the voxelization is done.

Placing the enrolled 3D surface into this volume, each point on the gallery surface falls into a voxel. A given voxel can be empty or hold one or more points from the gallery surface. If a probe surface is placed into this volume, every point on the probe surface should also fall into some voxel if the volume size is big enough. Suppose that there is a point P_1 on the probe surface that lies in the voxel V_1 in the volume. P_2 which lies in voxel V_2 is the point on the gallery surface which is closest to P_1 . The distance between two points P_1 and P_2 can be approximated by the distance between the center of the two voxels V_1 and V_2 with the precision of the voxel size, shown in Figure 5.2.

In the ICP algorithm, given an enrolled surface in the volume, different probe surfaces attempt to find the minimum distance error to the enrolled surface. Here, the gallery surface is fixed, but the position of points on the probe surface varies within the volume from iteration to iteration. If all the points from the probe are within the volume which holds the gallery surface, each point should be in some voxel. For a given point P on the probe surface, suppose we know that its closest point on the gallery surface is P' and voxel V_p is the voxel this given point is in. The distance between P and P' is approximately equal to the distance between P' and the center of the voxel V_p , with the precision of the voxel size. Each voxel in the data structure can index a distance value pre-computed at enrollment of the surface. Therefore, given the position of one point, the index of the voxel can be calculated easily.

5.3.1 Volume Size

The initial experiments used a volume around the 3D shape corresponding to the max size of the object. For an ear, the volume size is set to 8 cm wide, 10 cm tall and 8 cm deep. For a face, the volume size is set as 10 cm wide, 14 cm tall and 7 cm deep. The voxel size is related to the precision of the 3D scanner. If the size for each voxel is $0.05\text{cm} \times 0.05\text{cm} \times 0.05\text{cm}$, we have $160 \times 200 \times 160 = 5.12M$ voxels per volume for an ear. The fixed volume size is usually larger than 3D objects in the volume. The reason that it has extra space is that we need to consider the orientation of the 3D objects. Even though the width of the ear is usually small, the overall crossing will be large if the ear is rotated along the Z axis instead of straight up. Unfortunately much space is wasted for fixed volume. In order to solve this problem, we reduce the volume size by applying principal component analysis (PCA) on the 3D data.

Principal Components Analysis is used for computing the dominant variances representing a given data set. As we apply PCA on the 3D data, it yields three eigenvectors. The first eigenvector is the direction of greatest variation in the data, the second eigenvector is the direction of second greatest variation, and the third eigenvector is the third greatest variation. And all eigenvectors are orthogonal to each other.

According to our 3D shape data, the greatest variation is related to the height of the 3D shape, the second one to the width, and third one to the depth of the data. After obtaining these three eigenvectors, a new coordinate system $[V_x^T \ V_y^T \ V_z^T]$ is generated, each V_i is a vector. If we project the old 3D points into the new system, the ranges along these three new axes represent the size of a box enclosing the 3D shape.

Figure 5.3 illustrates the steps of this procedure. When compared to the width, height, and depth in Figure 5.3(a) and Figure 5.3(c), the overall size of the bounding box of the new 3D shape is smaller than the size of the old one.

$$[X'Y'Z'] = [XYZ] * [V_x^T \ V_y^T \ V_z^T] \quad (5.1)$$

$$Width = max(X') - min(X') \quad (5.2)$$

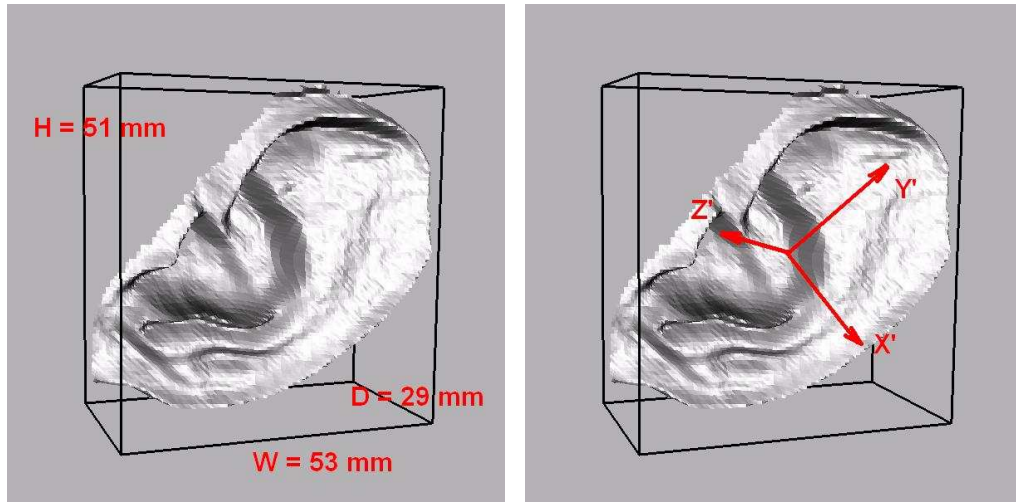
$$Height = max(Y') - min(Y') \quad (5.3)$$

$$Depth = max(Z') - min(Z') \quad (5.4)$$

For the ear experiment, the overall file size is reduced by 10 times after we used the PCA-based volume size instead of the fixed volume size. Also with a smaller file size to save the information, it requires less memory to build and read the data. Therefore this reduces the building time, and sometimes it also reduces the matching time when swapping is needed in the old approach.

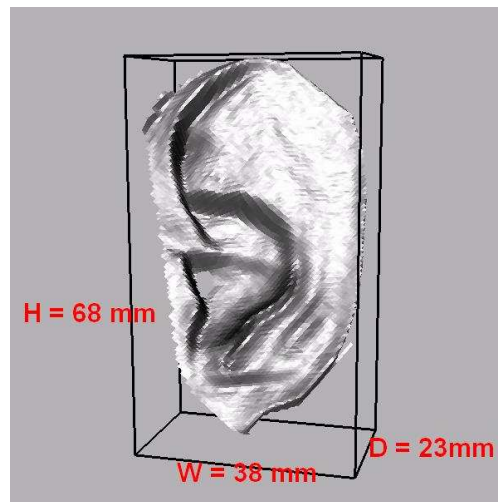
5.3.2 Implementation

To implement our strategy of pre-computed voxel nearest neighbor, we compute ahead of time for each voxel in the 3D space, it's closest point on the gallery 3D shape. The first step is to place the 3D surface into a volume whose center is the center of the 3D surface. Due to the noise in the 3D depth value, the position of the gallery surface center x , y and z are defined in different ways. $x_{center} = \frac{x_{max}+x_{min}}{2}$, $y_{center} = \frac{y_{max}+y_{min}}{2}$, $z_{center} = \frac{z_{max}+z_{min}}{2}$. Figure 5.4 shows a volume holding both gallery and probe. For each voxel element in the volume, we use a k-d tree to find the closest point on the gallery surface to that voxel's



(a) Original 3D Shape

(b) New Coordinate System



(c) New 3D Shape

Figure 5.3: Steps to Calculate the Volume Size. In (b), a new coordinate system is generated from eigenvectors of the covariance matrix, where Y' is according to the direction of the largest variance in the dataset, X' to the second, and Z' to the third. Part(c) shows the new 3D shape after projecting every old point onto new coordinate system.

center. Once the point is found, the index of the point is stored as the value of the voxel element. A data structure $VoxelElement[Width][Height][Depth]$ is used to represent the subdivision of 3D space into voxels, and the value of width, height and depth comes from 5.4. The value of $VoxelElement[x][y][z]$ is the index of the gallery point, which is closest to the point (x,y,z) . We store the index of the point instead of the point position to save space. Pre-computed results are saved to a file which can be read into memory when needed.

Then, computing the closest neighbor for a current position of the probe face is simply indexing into the voxel data structure. Thus, constant computational time instead of $O(\log N_G)$ is achieved. This is blazingly fast in comparison to any of the other nearest neighbor finding methods, but of course it is offset by the size of the storage required. Furthermore, since the access time is constant, we can use the finest resolution for the gallery image, which avoids the computation expense of using the point-to-triangle approach [20]. Figure 5.4(a) shows an example when both gallery and probe come into the voxel.

5.4 Experiments

In order to evaluate the efficiency of this method, we compare the recognition rate, space and running time between the original algorithm and our proposed approach.

We present results using 3D ear shape from 369 subjects and 3D face shape from 219 subjects. For each subject, the earlier 3D images are used for the gallery, and the later 3D images are used as probes. For the ear experiment, the gallery images use the full resolution, and the probes are subsampled by every 4 rows and every 4 columns. The average number of points is 5500 for a gallery ear shape, and

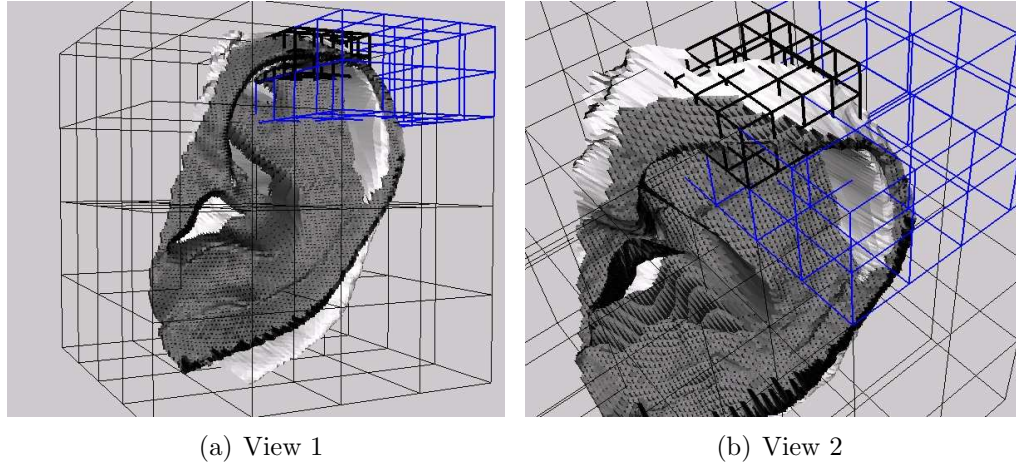


Figure 5.4: Gallery and Probe Images Show in the Same Volume. Each voxel in the volume corresponds to a point index on the gallery surface.

400 for a probe ear shape. And for the face experiments, both gallery and probe images are subsampled by every 4 rows and every 4 columns. The average number of points on a gallery and a probe surface are 4000 and 3000, respectively, for face shapes. In addition, different voxel sizes are tested, and comparison results are presented. The system runs on dual-processor Pentium Xeon 2.8GHz machines with 2GB RAM, and the implementation is written in C++.

5.4.1 Voxel Size

Three voxel sizes are examined using the same dataset for both ear and face biometrics. For the ear experiments, they are $1mm^3$, $0.5mm^3$ and $0.25mm^3$. For the face experiments, they are $2mm^3$, $1mm^3$ and $0.5mm^3$. The reason for using different voxel size for ear and face is because the gallery face images are subsampled by every 4 columns and rows. Before the matching procedure takes place, we build the volume for every gallery ear/face. For each voxel in the volume, a kd-tree structure is used to find the closest point on the gallery and we save the

TABLE 5.1

Ear Biometrics: Different Parameters Affected by Voxel Size. At 0.1 level of significance, there is no statistically significant difference between (1), (2) and (3) using both Z-test and Friedman test.

Voxel size	Building Time (1 gallery ear)	Reading Time (1 gallery ear)	Matching Time (1 against 369)	File Size (369 images)	Performance
1mm	10-50s	0s	15-25s	127 M	97.0% ¹
0.5mm	30-200s	1s	20-30s	1009M	97.3% ²
0.25mm	100-500s	1 - 5s	20-30s	7.8 G	97.3% ³

TABLE 5.2

Face biometrics: Different Parameters Affected by Voxel Size. At 0.1 level of significance, there is no statistically significant difference between (1), (2) and (3) using both Z-test and Friedman test.

Voxel size	Building Time (1 gallery ear)	Reading Time (1 gallery ear)	Matching Time (1 against 219)	File Size (219 images)	Performance
2mm	25-35s	0-1s	80-90s	983M	93.6% ¹
1mm	150-160s	5-6s	80-90s	15G	94.1% ²
0.5mm	1500-1600s	35-45s	80-95s	63 G	93.2% ³

results on the disk. In order to utilize our method, we read one voxelized gallery data structure into memory and match it against all the probes. Therefore, our recognition experiment has two processes, offline building and online matching. Tables 5.1 and 5.2 illustrate the time requirement for each process.

For the ear experiments, all the images are acquired using a Minolta Vivid 910 with the “Tele” lens, and the subject sat approximately 1.5 meters away from the sensor. Within that distance range, the sensor has a depth accuracy of approximately 0.55mm. According to our results, going to a finer voxel size, from 0.5mm to 0.25mm, does not yield much in term of increased accuracy, yet, it requires significantly more storage space and longer time to process. Even though the access time is a constant value, when the number of voxels is too big, it will exceed the size of available memory, and force the algorithm to use swap space, which will slow down the computation. If we increase the voxel size from 0.5mm to 1mm, the reading time drops, and matching time is at the same level, and the performance drops by around 0.3%, which is not statistically significantly different from the smaller voxel size.

For the face experiments, the image acquisition is the same. But since the gallery images are subsampled by 4 x 4, there is no statistically significant difference in performance for voxel size variations.

Table 5.3 compares the original ICP and our pre-computed ICP. The difference between the original ICP and our ‘Pre-computed Voxel Closest Neighbor’ ICP is obvious. The recognition performance of both methods is essentially the same, but for the time spent on the online matching, the pre-computed matching is significantly faster than the original one. For the regular ICP, the matching time for one probe against all the gallery images is around 106 seconds, while kd-tree

TABLE 5.3: Ear Experiments: Difference Between Original ICP and Pre-computed ICP. ** At 0.1 level of significance, there is no statistically significant difference between (1), (2), (3) and (4) using both Z-test and Friedman test.

Method	Building Time (1 gallery ear)	Running Time (1 against 369)	Perfor- mance
Original ICP	0	106	97.0% ¹
Pre-computed (voxel = 1 mm)	10- 50s	0.2+ (15 – 25)s	97.0% ²
Pre-computed (voxel = 0.5 mm)	30-200s	1+ (20 – 25)s	97.3% ³
Pre-computed (voxel = 0.25 mm)	100-500s	3 + (20 – 30)s	97.3% ⁴

building just takes less than one second. For the voxel ICP, the matching time is just around 25-30 seconds. But the file reading time takes 10 seconds when we use size of voxel as 0.25 mm. So the voxel matching time is about 1/3 to 1/4 of the time for the “regular” approach, but reading the data structure takes an extra amount of time.

Table 5.4 illustrates how running time increases when probe size gets larger for both original ICP and pre-computed ICP. When the probe size is small, there is no advantage to the voxel approach. However, for very large probe size the voxel approach yields an enormous improvement in speed. Here, we suppose all the gallery images can be kept in memory. In a real biometrics application, some or all of the gallery might be kept in memory all the time.

5.5 Improvement

As we stated in the previous section, the most time consuming part of the ICP algorithm is closest point searching. There are two common ways to find the closest point, point-to-point and point-to-surface. A detailed comparison between them for a biometric application can be found in [81]. The point-to-point approach is fast, and accurate when all the points on the probe surface can find a good closest

TABLE 5.4

Ear Biometric: Run Time vs. Probe Size for Both Original ICP and Pre-computed ICP. (voxel = 0.5 mm)

Probe size	Original ICP	Pre-computed ICP
10	5s	3s
50	20s	7s
100	35s	10s
200	75s	15s
300	106s	25s
369	150s	30s

point on the gallery surface. But if the gallery image is subsampled or coarse in the original, the point-to-point approach loses accuracy. On the other hand, the greatest advantage of the point-to-triangle approach is that it is accurate through all the different subsample combinations. But this behavior comes at a substantial computational expense.

Our voxel algorithm can shift the computation burden to offline, therefore if the gallery images are not in a fine resolution, the point-to-triangle method for pre-computed distance should be able to yield better performance without increasing the running time for the recognition. This is proved by the experimental results. By using point-to-triangle method for pre-computing, the ear recognition rate is improved from 98.3% to 98.7%, and the face recognition rate is improved from 94.1% to 96.4%. The improvement is more obvious in the face recognition experiment, which also demonstrates that the point-to-triangle method is more accurate than point-to-point method when the gallery images are coarse.

5.6 Summary And Discussion

The main contribution of this chapter is the “Pre-computed Voxel Closest Neighbor” strategy to improve the speed of the ICP algorithm for use in biometrics. This technique is aimed at a particular application in human identification. The idea is based on the possibility of computing the data structure before the matching procedure taking place.

Different voxel sizes are examined, and the performance and running time are compared with the results from the original ICP algorithm. Our experimental results verify the expected feature of our approach on our 369-subject dataset with ear biometric, and a 219-subject dataset with face biometric. The online matching time drops significantly when we use the pre-computed results from the enrolled 3D shape offline computation. Our results demonstrate that for very large galleries the voxel approach yields a dramatic improvement in speed.

CHAPTER 6

MULTI-BIOMETRICS ¹

Recently, multi-biometrics have been investigated by several researchers [9, 10, 41, 47]. Multi-biometrics can be divided into three simple classes, according to the method of combination. These are multi-modal, multi-algorithm and multi-instance. In general, multi-modal uses different modalities of biometrics, like face, voice, fingerprint, iris and ear of a same subject. Also we consider that for a given biometric, the data from different sensors are one kind of multi-modal, like 2D intensity data and 3D range data. Multi-algorithm uses different algorithms on the same data. For example, we can use both PCA and ICP on 3D ear data. Multi-instance has more than one representation for a given subject. For example, if we took three 2D ear images of the same person on different dates, then the three images together can be treated as a representation of this person. A complex combination can involve more than one kind of multi-biometrics. For example, we can combine 2D PCA and 3D ICP, which includes both multi-modal and multi-algorithm biometrics if we consider 2D intensity data and 3D range are different modalities.

As we first explored multi-modal biometrics using ear data, only a 202-subject dataset was available. As more image acquisitions were held, more images were

¹This section is based on the paper, “Multi-Biometrics 2D and 3D Ear Recognition”, presented at the *Audio- and Video-based Biometric Person Authentication Conference, 2005* [78].

accumulated, and in this chapter the largest dataset has 302 subjects. All the experiments run on the 202-subject dataset are rerun on the 302-subject dataset to validate our previous conclusions.

6.1 Fusion Levels and Score Normalization

Each simple biometric has four steps: (1) obtain the data from the sensor, (2) extract the interesting area or features from the raw data, (3) compare the data to a group of enrolled data to obtain the matching score and (4) determine the correct or incorrect match based on the matching score [61]. Based on these different steps, there are several possible fusion levels. Sensor level fusion combines the raw sensor outputs. Feature extraction level fusion combines multiple extracted features from each biometric. Matching score level fusion combines the matching scores from each biometric. Decision level fusion uses the results from different biometrics and makes the final decision based on all of them.

In our study, the fusion rules work at the matching score level. Since each simple biometric has a different meaning, range and distribution of matching scores, score normalization is required in order to combine them. In our experiments, min-max score normalization has been applied on all the results before we do the fusion: $s' = (s - min)/(max - min)$.

6.2 Multi-modal Biometrics

Multi-modal biometrics in this work refers to the combination of 2D intensity data and the 3D range data. There are three algorithms based on 3D range data, and one on 2D intensity data. Therefore, the combinations include 2D PCA with 3D ICP, 2D PCA with 3D PCA, and 2D PCA with 3D edge-based approach.

TABLE 6.1

Rank One Recognition for Fusion on Multiple Modalities (302 subjects).

Multi-modals	MIN	Simple SUM	Advanced Sum
2D PCA + 3D ICP	76.4%	81.1%	82.5%
2D PCA + 3D PCA	72.2%	78.8%	79.1%
2D PCA + 3D Edge	73.5%	80.5%	82.5%

First, two simple fusion rules are tried on all three combinations. As shown in Table 6.1, the sum rule performs much better than the min rule. This conclusion is similar to that in [15, 47]. Also, an advanced sum rule is tested. The rank one matching in each modality is given an additional weight, which measures the distance between itself and the rank two match. The advanced sum rule yields better results than the simple sum rule.

The sum rule adds individual matching scores from different matches. Equal weights are assigned to each modality. However, in general, some modalities have better performance than others. In order to show the relative strength of several modalities, different weights are assigned to individual modalities. We test the weight assignment by using 202 subjects on 2D PCA combined with 3D ICP. As shown in Table 6.2, the highest performance is 93.1%, obtained when the weight of ICP is 0.8, and the weight of PCA is 0.2.

Applying the same weighted sum rule to the other two combinations, the best performance is obtained when there is equal weight for each modality. This is because 2D PCA, 3D PCA and edge-based approaches have similar performance. The rank one recognition is 79.1% when combining 2D PCA and 3D PCA, and

TABLE 6.2

Different Weights for Fusing the ICP and PCA Results.

Weight	Weight	Performance	Performance
2D PCA	3D ICP	(202 Subjects)	(302 subjects)
1	0	71.4%	63.6%
0	1	85.1%	84.1%
0.9	0.1	73.3%	66.9%
0.8	0.2	76.7%	68.9%
0.7	0.3	78.2%	73.8%
0.6	0.4	81.7%	78.5%
0.5	0.5	84.2%	82.5%
0.4	0.6	86.6%	88.7%
0.3	0.7	89.1%	90.7%
0.2	0.8	93.1%	90.4%
0.1	0.9	91.6%	86.4%

TABLE 6.3

Fraction of the Correct Match in the Different Interval Level.

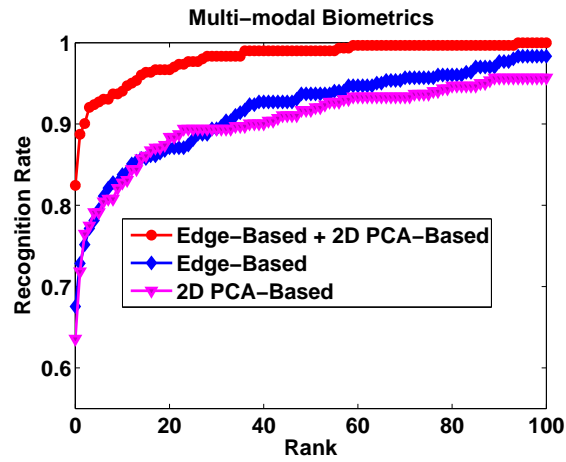
	1	2	3	4	5	6	7	8	9	10
PCA	0.4250	0.7429	0.8333	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ICP	0.2222	0.5319	0.9999	1.000	1.000	1.000	1.000	1.000	1.000	1.000

it is 82.5% when combining 2D PCA and 3D edge-based algorithm. The CMC curves are shown in Figure 6.1.

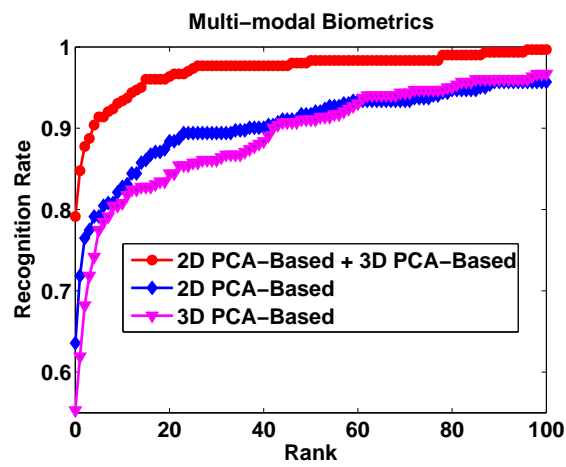
Our third combination rule is based on the analysis of the interval between rank 1 and rank 2 in both PCA and ICP results. Figure 6.2 shows that the overlap area between the correct matches and incorrect matches is much less in ICP than in PCA, which means that it is easier to use a threshold to separate the correct and incorrect matches in the ICP than in the PCA results.

Figure 6.3 shows the probability distribution of the different intervals between the correct matches and incorrect matches. In general, the greater the gap between the rank 1 and rank 2 matches, the higher the possibility that it is a correct match. When the interval in ICP is greater than 0.2, they are all correct matches. The corresponding value in PCA is 0.002. For both ICP and PCA, we split the interval range into 10 steps. All the interval values are placed into these 10 steps. The ratio of the correct over incorrect matches in each interval step is shown in Table 6.3.

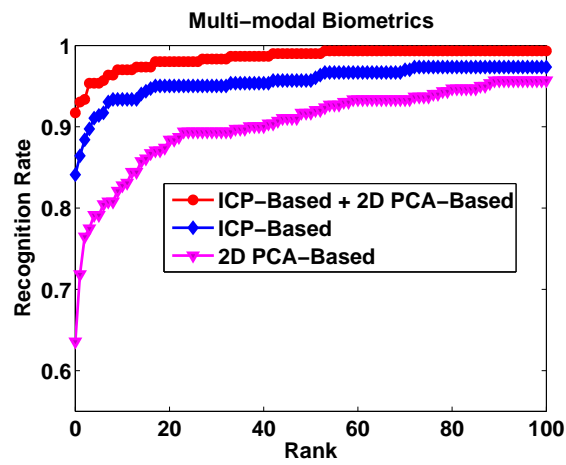
When an interval falls into a certain range, we can determine the possibility that it is a correct or incorrect match from Table 6.3. Using this information we can combine the PCA and ICP in a smarter way. Before the combination, the



(a)

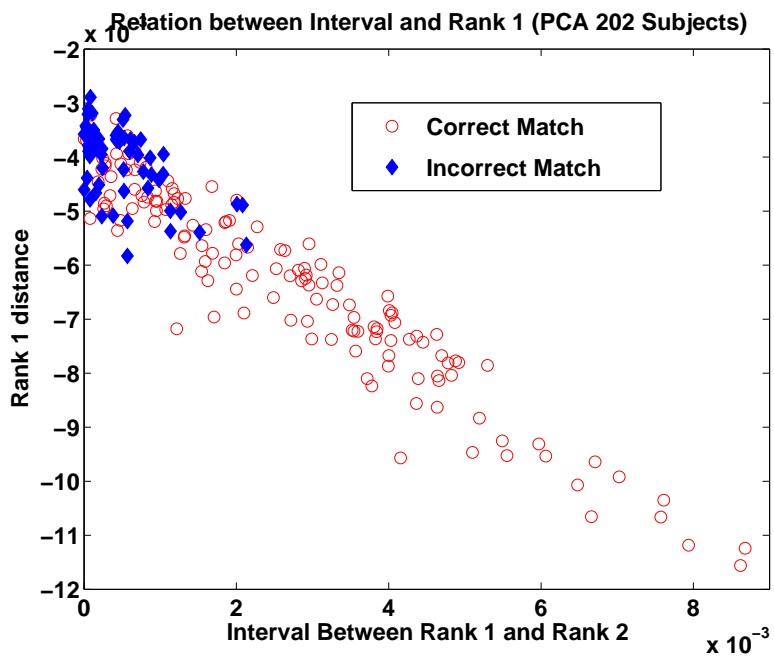


(b)

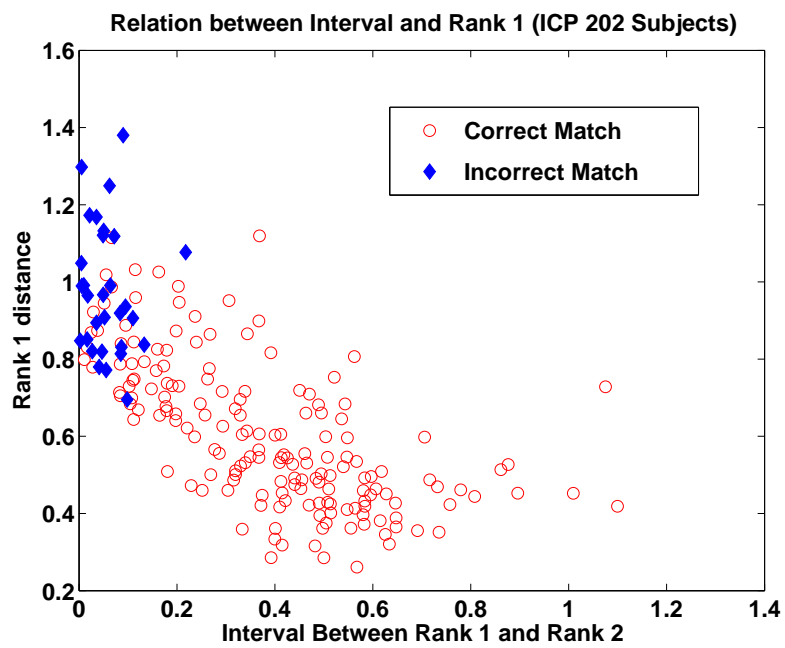


(c)

Figure 6.1. Multi-modal Biometrics Performance.



(a)



(b)

Figure 6.2. Relationship Between Correct Matches and Incorrect Matches.

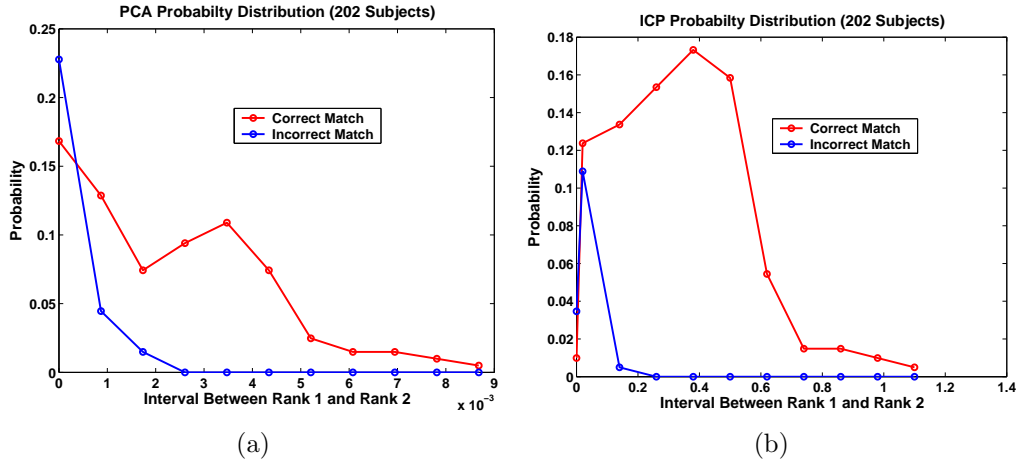


Figure 6.3. Interval Distribution Between Correct and Incorrect Matches Using Both PCA and ICP approaches.

interval between the rank 1 and rank 2 is computed for each comparison in the ICP and PCA. Then the corresponding percentage of the correct match and incorrect match is obtained according to Table 6.3. Using this strategy to combine the PCA and ICP results on 202 subjects, the rank one recognition rate is 93.1%, which is the same as the best result obtained from the simple weight scheme shown in Table 6.2.

Until now, all the results are calculated using the 202-subject dataset. Since the small dataset has a distribution similar to the larger dataset (302 subjects), we predict the distribution of the larger dataset by using the value in Table 6.3. The rank one recognition rate is 91.7%, which is even better than the results (90.1%) using simple weighted sum scheme. Thus it seems that performance may be increased by using a smart fusion step. However the increase is not statistically significant and this issue deserves further exploration.

TABLE 6.4

Multi-algorithm Biometrics Using Weighted Sum Rule.

	3DICP	3DPCA	3DEdge	Performance
ICP + PCA	0.90	0.10		87.70%
ICP + Edge	0.80		0.20	90.2%
PCA + Edge		0.40	0.60	69.9%

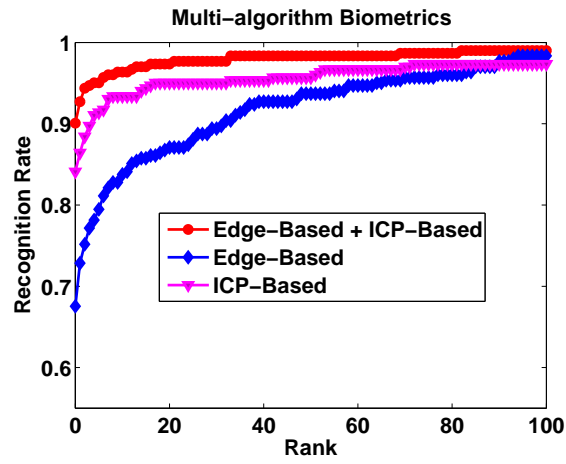
6.3 Multi-algorithm Biometrics

Three different algorithms have been developed to use on the 3D data. These are the ICP-based algorithm, PCA-based algorithm and edge-based algorithm. After score normalization, the weighted sum rule is used for combinations. Rank one recognition rates are shown in Table 6.4. The best performance is achieved when combining ICP and edge-based algorithm on the 3D data.

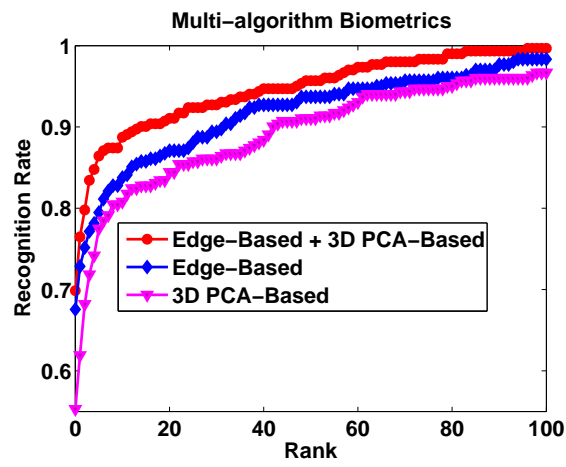
From Table 6.1, if we only consider those with not so good performance, like 2D PCA, 3D PCA and 3D edge-based approach, the multi-modal biometrics has better performance than the multi-algorithm.

6.4 Using Multiple Images to Represent a Person

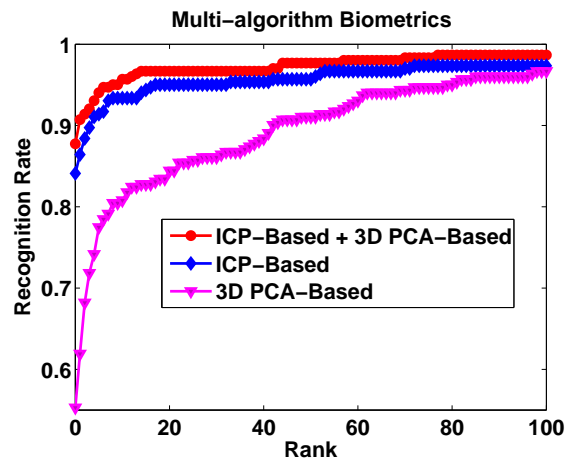
In general, approaches perform better with a multiple-sample representation of a person, and scale better to larger datasets. We have 169 subjects that have at least 4 good images in both 2D and 3D data. Each pair of 2D and 3D images were taken on a different date. In this section, we will concentrate on the multi-galleries or multi-probes using 2D PCA and 3D ICP algorithms.



(a)



(b)



(c)

Figure 6.4. Multi-algorithm Biometrics Performance.

TABLE 6.5

Fusion on Multiple Galleries and Probes (169 Subjects).

	2D PCA		3D ICP	
	MIN	SUM	MIN	<i>SUM</i>
1G1P	73.4%		81.7%	
1G3P	82.2%	83.4%	95.3%	81.1%
2G2P	84.0%	87.5%	97.0%	81.7%
3G1P	81.7%	80.5%	91.1%	81.7%

For each subject, there are four 2D and 3D images available. We consider three possible multiple-instance representations based on these images. These are (a) 1 in the gallery and the other 3 images in the probe, (b) 2 in the gallery and the other 2 in the probe, and (c) 3 in the gallery and the other 1 in the probe. Two fusion rules, min and sum, are explored to combine the results, shown in Table 6.5. It is interesting here that we have multi-instance better than multi-modal. Using one gallery and one probe for these 169 subjects, the rank one rate is 73.4% for 2D PCA, and 81.7% for 3D ICP. Combining the results of 2D PCA and 3D ICP, the best performance obtained is 88.2%.

In the multi-galleries and multi-probes experiments, the best performance is achieved when 2 images are put into the gallery and the other 2 put into the probe. This is true in both 2D PCA and 3D ICP algorithms. This combination gives us 4 matches, whereas the other combinations give 3 matches. Also we noticed that the min rule is much more powerful than the sum rule in the 3D ICP performance,

while it has similar performance to the sum rule in the 2D PCA performance. We attribute the performance of the min rule in the case of 3D ICP to the possibility of minimizing “outliers” in the 3D matching.

Matching of 3D ear images has many sources of “outliers”. There can be outlier noise in a given 3D image, such as a “spike” from 3D sensing. Also in matching one 3D image to another, incorrect point correspondences may arise, possibly due to points existing in one scan but not the other. Increasing the number of representations for a certain person in both the gallery and probe gives a better chance to find the correct correspondence between the points. Thus, the performance increases significantly in the ICP experiment.

6.5 Summary And Discussion

We find that multi-modal, multi-algorithm or multi-instance improve performance over a single biometric. The combination of the 2D PCA and 3D ICP gives the highest performance of any pairs of biometrics considered. Three different multi-biometric combinations were considered. All result in improvement over a single biometric. Among the four single modal ear biometrics, the ICP-based recognition outperforms the other three methods. And it is expected that the best combination includes the ICP as one of the components. Multi-modal with 2D PCA and 3D ICP gives the highest performance. As to the other three not-as-good methods, multi-modal biometrics turns out to have better performance than the multi-algorithm biometrics.

The fusion experiments on multi-modal, multi-algorithm and multi-instance biometrics yield different results. The sum rule outperforms the min rule on multi-modal and multi-algorithm biometrics, while the min rule performs well on

the multi-instance biometrics, especially when using the ICP algorithm. The min rule has the power to reduce the noise from the original data, which is suitable for the application to multi-instance biometrics. The new fusion rule we introduced in combining 2D PCA and 3D ICP is based on analyzing the interval between rank one and rank two. And the performance result is the best of the fusion rules we used.

The multi-modal 3D ICP plus 2D PCA recognition was 87.7% on the 302 person dataset, as listed in Table 6.4. It is useful to ask how a multi-modal result compares to the multi-instance results for the individual modes. The multi-modal approach represents a person by two images, in both the gallery and as a probe. If we look at the two-image representation in each of the individual imaging modes, we get 87.5% for 2D PCA and 97% for 3D ICP on the subset of 169 of the 302 persons, Table 6.5. The multi-modal result for this same subset of 169 persons is 88.2%. Thus we find that the multi-modal result barely improves over the two-image 3D ICP result, and that the four-image 3D ICP result for multi-instance is substantially better than the multi-modal result. This is different relative performance than found by Chang [16] in a study of multi-modal face recognition, where multi-modal 2D + 3D performance was greater than multi-image 2D or multi-image 3D. However, our work differs in several potentially important respects. One is of course that we study ear recognition rather than face recognition. But also, Chang used the same PCA-based approach for both the 2D face and the 3D face recognition, whereas we use an ICP approach for our 3D recognition. This is important because it appears in our results that the ICP-based approach is substantially more powerful than the PCA-based approach for 3D. Another potentially important difference is that in our multi-image results,

the two images used to represent a person are taken at different times, at least a week apart. Chang used images from the same acquisition session in his multi-image results. It is quite possible that images taken on different days give a more independent sample, and so give better performance.

6.6 Improved ICP Algorithm

The multi-biometric results presented in previous sections indicate that 3D shape matching with an ICP-based approach has strong potential for ear biometrics. Therefore, after the results in previous sections were completed, we considered various refinements to this approach, several of which were incorporated into an improved algorithm. The amount of the ear shape used in the gallery and probe representations was adjusted to reduce interference from the background. A step to remove outlier point matches was added to reduce the effects of incorrect correspondences. Our improved algorithm produces substantially better results. Using the 302-person dataset, with a single 3D ear scan as the gallery enrollment for a person, and a single 3D ear scan as the probe for a person, the new algorithm achieves 98.7% rank-one recognition [81]. This performance from a single modality and algorithm is high enough that a larger and more challenging data set is needed in order to experimentally evaluate its use in possible multi-biometric scenarios.

6.7 Multi-biometrics Using Ear and Face

So far, our multi-modal biometrics only considered ear images. It would be interesting to look at the combination of ear and face. Both of them have their pros and cons when used for recognition. The face has a larger area and therefore

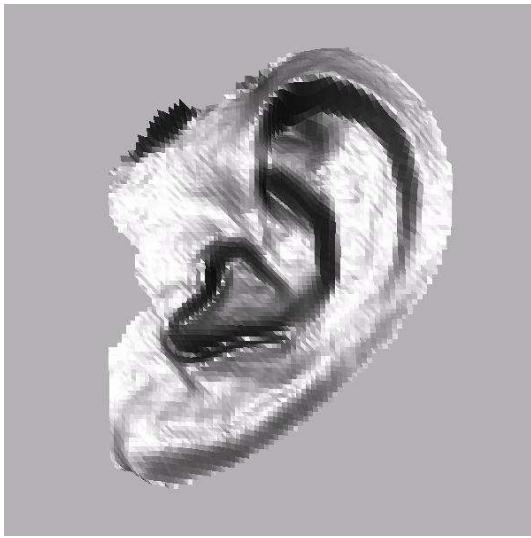
can be captured easily. However, the face shape can be distorted by expression change, eyeglasses, and heavy makeup. Conversely, ear shape is not affected by expression change, makeup or glasses. However, due to its relatively small size, it is more likely for the ear to have total occlusion than the face.

Studies [10, 16, 41, 49, 62, 73] have shown that combining different modalities usually increases the percentage of enrollment rate and improves the performance. When one biometric is not available, the other one can be used. Ear and face can be captured in the same manner, and they are physically close to each other. This provides the possibility and feasibility of having multi-modal biometrics using both the ear and face.

There are 174 subjects in our dataset that have two ear shapes and two face shapes. For each subject, we place one ear/face in the gallery and the other one in the probe. Our revised ICP algorithm is applied on both the ear and face datasets. The rank-one recognition rate is 97.7% for the ear shape matching, and 93.1% for face shape matching.

In order to combine these two modalities, we use the min-max normalization and both sum and interval-based fusion rules. Either fusion rule yields 100% rank-one recognition rate on these 174-subject dataset. The result is consistent with our conclusion about the efficiency of multi-modal biometrics, and also indicates the promising possibility of multi-modal biometrics using both 3D ear and face shapes.

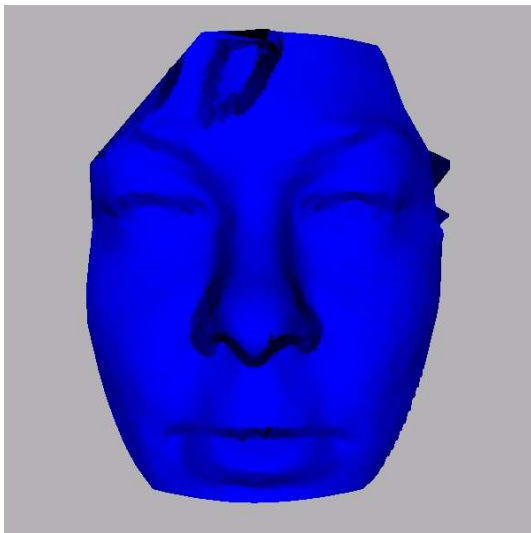
Figures 6.5 - 6.6 gives two examples of correctly matched subjects with multi-modal results, but that were originally mismatched with face or ear alone.



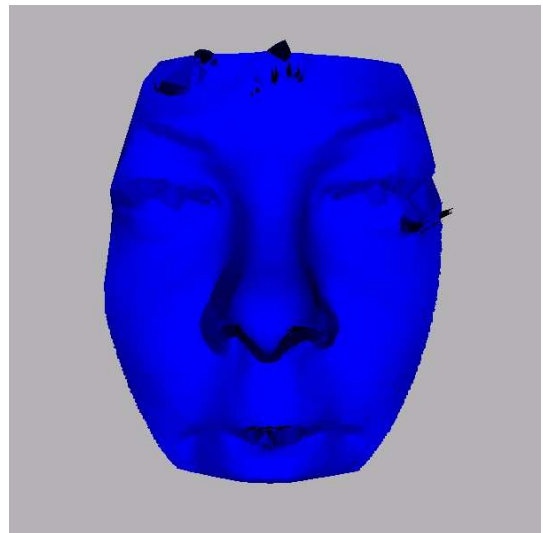
(a) Gallery Ear



(b) Mismatched Probe Ear



(c) Gallery Face



(d) Matched Probe Face

Figure 6.5. Multi-modal Biometrics using Ear and Face. The original matching score of ear alone does not return the correct match, but it is corrected by the face matching.

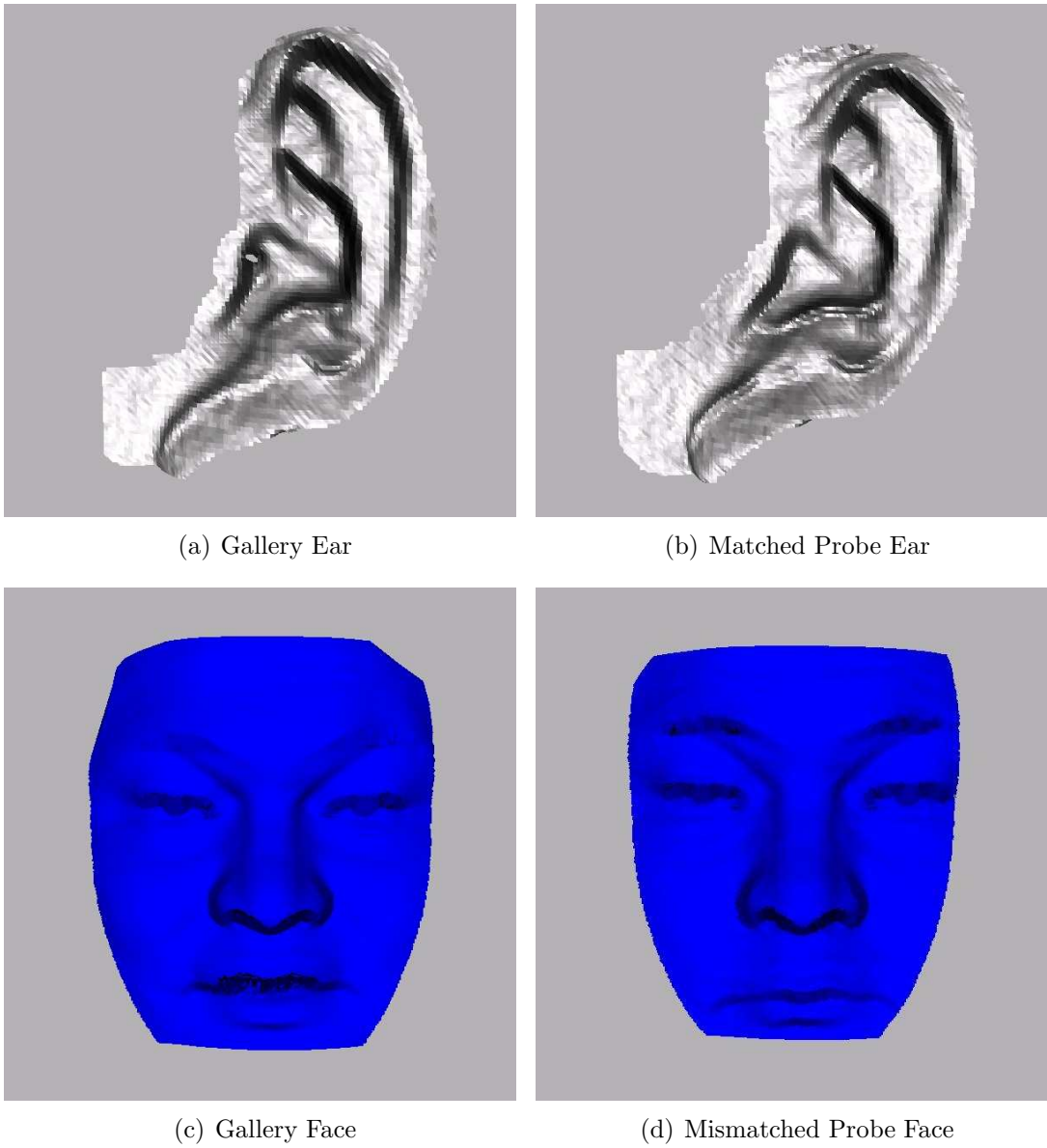


Figure 6.6. Multi-modal Biometrics using Ear and Face. The original matching score of face alone does not return the correct match, but it is corrected by the ear matching.

CHAPTER 7

AUTOMATIC EAR RECOGNITION ¹

In the earlier chapters, the pre-processing of ear images has had manual steps, and algorithms have not necessarily handled problems caused by hair and earrings. This chapter presents the first fully automated system for ear biometrics using 3D shape. There are two major parts to the system: automatic ear region segmentation and 3D ear shape matching. Starting with the multi-modal 3D+2D image acquired in a profile view, the system automatically finds the ear pit by using skin detection, curvature estimation and surface segmentation and classification. After the ear pit is detected, an active contour algorithm using both color and depth information is applied, and the contour expands to find the outline of the visible ear region.

We have found that the active contour (AC) algorithm is well suited for determining the ear region. The ear pit makes an ideal starting point. The AC algorithm grows a contour until it finds the ear edge, and is robust in its ability to exclude earrings and occluding hair. When the active contour finishes, the outlined shape is cropped from the 3D image, and the corresponding 3D data is then used as the ear image for matching. The matching algorithm achieves a rank-one recognition rate of 97.6% on a 415-subject dataset. In order to demonstrate the

¹This section is based on the paper, “An Automatic 3D Ear Recognition System”, presented at *Third International Symposium on 3D Data Processing, Visualization and Transmission, 2006*.

competence of our automatic ear segmentation, two other experiments are conducted on the same dataset. A comparable rank-one recognition rate using the old manually-assisted ear cropping that uses a fixed template [80] rather than the active contour is 96.4%. And a comparable rank-one recognition rate with everything being the same except the the location of the ear pit is manually marked is 97.1%.

This chapter is organized as follows. A review of related work is given in section 2. In section 3, we describe the experimental method and materials used in our work. Section 4 presents the detail of the automatic ear segmentation system. Section 5 describes an improved iterative closest point (ICP)-based approach for 3D ear shape matching. In section 6, we present the experimental results and analysis. Section 7 gives the summary and conclusions.

7.1 Segmenting the Ear Region from a Profile Image

Automatically extracting the ear region from a profile image is a key step in making a practical biometrics system. In order to locate the ear from the profile image, we need to have a robust feature extraction algorithm which is able to handle variation in ear location in the profile images. After we find the location of the ear, segmenting the ear from the surrounding is also important. Any extra surface region around the ear could affect the recognition performance. An active contour (“snakes”) approach is used for this purpose [24, 46, 75].

Initial empirical studies demonstrated that the ear pit is a good and stable candidate as a starting point for an active contour algorithm. When there is so much of the ear covered by hair that the pit is not visible, the segmentation will not be able to be initialized. But in such cases, there is not enough ear shape

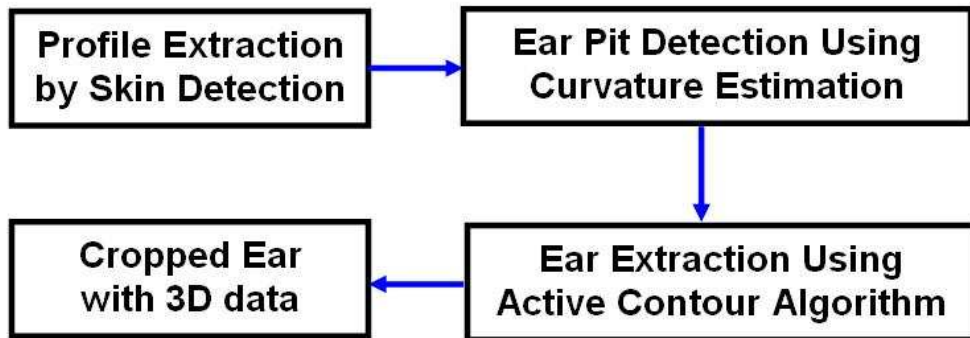


Figure 7.1. Data Flow of Automatic Ear Extraction.

visible to support reliable matching, anyway. From the profile image, we use skin detection, curvature estimation, and surface segmentation and classification to find the ear pit automatically. Figure 7.1 presents the steps that are involved in accomplishing the automatic ear extraction.

7.1.1 Ear Pit Detection

The first step is to find the starting point, the ear pit, for the active contour algorithm. Ear pit detection includes four steps: preprocessing, skin detection, curvature estimation, and surface segmentation and classification. We illustrate each step in following sections.

7.1.1.1 Preprocessing

We start with the binary image of valid depth values to find an approximate position of the nose tip. Given the depth values of a profile image, the face contour can be easily detected. An example of the depth image is shown in Figure 7.2(b).

A valid point has an (x, y, z) value reported by the sensor, and is shown as white in the binary image in Figure 7.2(c).

We find the X value along each row at which we first encounter a white pixel in the binary image, as shown in Figure 7.2(c). Using the median of the starting X values for each row, we find the approximate X value of the face contour. Within a 100-pixel range of X_{median} , the median value of the Y values for each row is an approximate Y position of the nose tip. Within a 120 pixel range of the Y_{median} , the valid point with minimum x value is the nose tip. The method avoids the possibility of locating hair or chin as the nose tip.

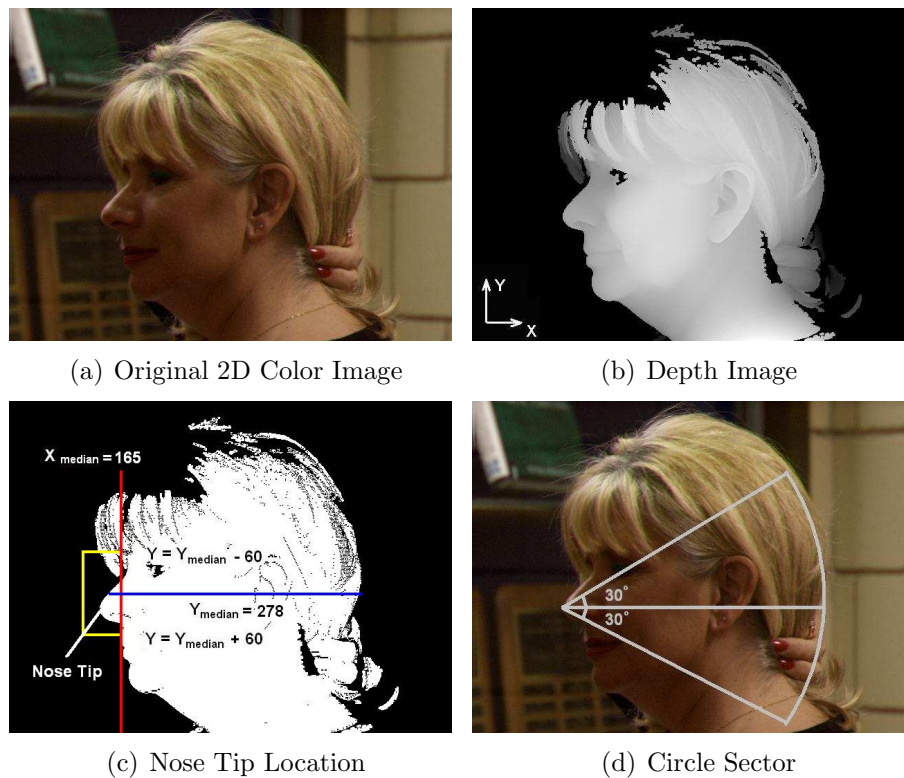


Figure 7.2. Using Nose Tip as Center to Generate Circle Sector.

Figure 7.2(c) illustrates the steps to find the nose tip. Using the point $P(X_{NoseTip}, Y_{NoseTip})$ as the center of a circle, we generate a sector spanning ± 30 degrees from the horizontal with a radius of 20 cm. One example is presented in Figure 7.2. With a high degree of confidence, the ear is included within the sector, along with some face, hair and shoulder. In all of the 830 images, none of the ears are outside the region that we look at, and the ± 30 degrees should cover people looking up or down by that much.

7.1.1.2 Skin Region Detection

Next, a skin detection method is applied to locate the skin region. This step is intended to isolate the face and ear region from the hair and clothes. Our skin detection method is based on the work of Hsu *et. al* [36]. The major obstacle of using color to detect skin region is that the appearance of skin-tone color can be affected by different lighting conditions. In their work, a light compensation technique is introduced to normalize the color appearance. In order to eliminate the dependence of skin-tone color on luminance, a nonlinear transformation on the YCbCr color space is applied to make the skin cluster luma-independent. With the transformation results, a parametric ellipse in the color space is used as a model of skin color, as described in [36]. Together with the preprocessing step, the skin region detection drops out some irrelevant regions, such as the shoulder or hair area, helping to reduce the computation time in later steps.

7.1.1.3 Surface Curvature Estimation

This section describes a method that can correctly detect the ear pit from the region obtained by previous steps. We know that the ear pit shows up in

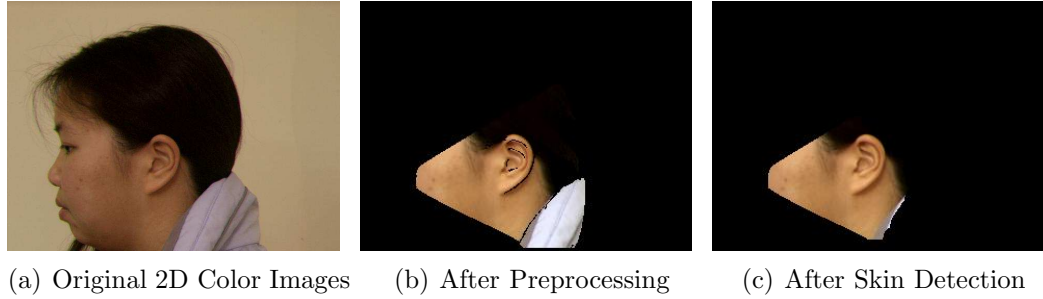


Figure 7.3. Ear Region with Skin Detection.

the 3D image as a pit in the curvature classification system [6, 28]. Flynn and Jain evaluated five curvature estimation methods and classified them into analytic estimation and discrete estimation [28]. The analytic estimation first fits a local surface around a point, then uses the parameters of the surface equation to determine the curvature value. Instead of fitting a surface, the discrete approach estimates either the curvature or the derivatives of the surface numerically. In our case, we use an analytic estimation approach with a local coordinate system determined by principal component analysis [28, 66].

In practice, the curvature estimation is sensitive to noise. For stable curvature measurement, we would like to smooth the surface without losing the ear pit feature. Since our goal at this step is only to find the ear pit, it is acceptable to smooth out other more finely detailed curvature information. Gaussian smoothing is applied on the data with an 11×11 window size. In addition, “spike” data points in 3D are dropped when an angle between the optical axis and a surface normal of observed points is greater than a threshold. (Here we set the threshold as 90 degrees.) Then for the (x,y,z) points within a 21×21 window around a given point P , we establish a local X,Y,Z coordinate system defined by PCA for that point P [28]. Using this local coordinate system, a quadratic surface is fit

to the (smoothed, de-spiked) points in the window. Once the coefficients of the quadratic form are obtained, their derivatives are used to estimate the Gaussian curvature and mean curvature.

7.1.1.4 Surface Segmentation and Classification

The Gaussian curvature (K) and mean curvature (H) are obtained for each point from the estimation step. The surface type at each point can be labeled based on H and K . See Figure 7.4 for the surface type definition. We group 3D points into regions with the same curvature label. In our experience, segmentation of the ear image by the sign of Gaussian and mean curvature is straightforward, and the ear pit can always be found in the ear region if it is not covered by hair or clothes.

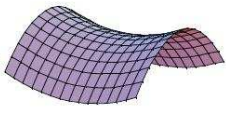
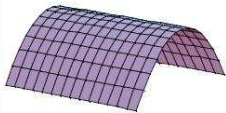

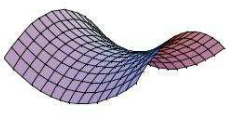

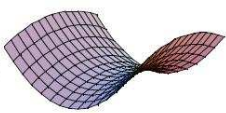

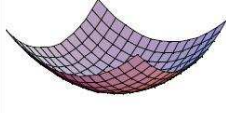
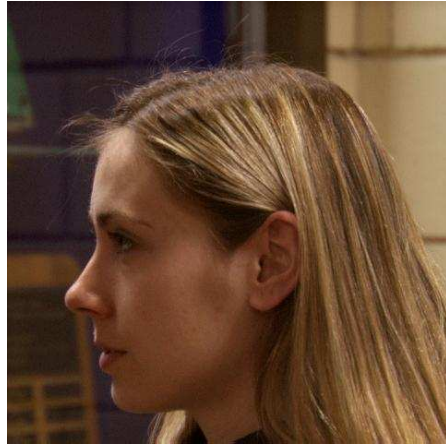
khs	$K < 0$	$K = 0$	$K > 0$
$H < 0$	 Saddle Ridge	 Ridge	 Peak
$H = 0$	 Minimal	 Flat	not possible
$H > 0$	 Saddle Valley	 Valley	 Pit

Figure 7.4. **H** and **K** Sign Test for Surface Classes [6].

After segmentation, we expect that there is a pit region ($K > 0$ & $H > 0$) in the segmented image that corresponds to the actual ear pit. Due to numerical error and the sensitivity of curvature estimation, thresholds are required for H and K . Empirical evaluation showed that $T_K = 0.0009$ and $T_H = 0.00005$ provides good results. Figures 7.5(c) and 7.5(d) show an example of the face profile with curvature estimation and surface segmentation. Also, we find that the jawline close to the ear always appears as a wide valley region ($K \leq 0$ & $H > 0$), and is located to the left of the ear pit region.

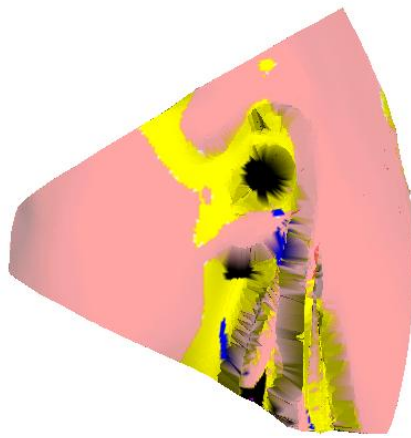
It is possible that there are multiple pit regions in the image, especially in the hair around the ear. A systematic voting method is developed to correctly find the ear pit. Three categories of information contribute to the final decision: the size of the pit region, the size of the wide valley region around the pit and how close the ear pit region is to the wide valley. Each category is given a score in the range of 0 to 10, and the score is simply calculated as the ratio of max area or distance at a scale of 10. For example, the largest pit region P_1 in the image has a score of 10, and the score of any other pit region P_2 is calculated as $Area(P_2)/Area(P_1) \times 10$. The pit with the highest average score is assumed to be the ear pit. The ear pit is correctly found in Figure 7.5(f). In order to validate our automatic ear extraction system, we compare the results $(X_{Auto_Ear_Pit}, Y_{Auto_Ear_Pit})$ with the manually marked ear pit $(X_{Manual_Ear_Pit}, Y_{Manual_Ear_Pit})$ for the 830 images. The maximum distance difference between $(X_{Auto_Ear_Pit}, Y_{Auto_Ear_Pit})$ and $(X_{Manual_Ear_Pit}, Y_{Manual_Ear_Pit})$ is 29 pixels. There are slightly different results from the active contour algorithm when using automatic ear pit finding and manual ear pit marking. But the difference does not cause problems for the active contour algorithm finding the ear



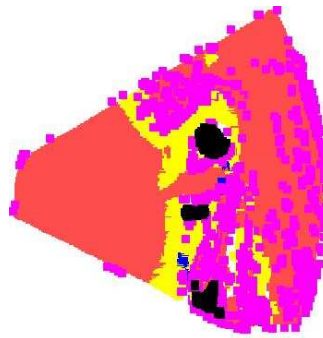
(a) Original 2D Color Images



(b) After Skin Detection



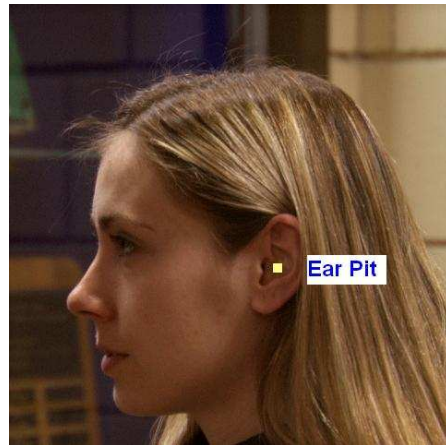
(c) Curvature Estim. & segm.



(d) 2D view of 7.5(c)



(e) Ear Pit Vote Results



(f) Ear Pit in Original Image

Figure 7.5. Steps of Finding the Ear Pit: (a) 2D/3D raw data, (b) skin detection, (c) curvature estimation, (d) surface segmentation, (e) region classification, ear pit detection.

region, at least on the 830 images considered here. Using a manual marking of the center of the ear pit rather than the automatically found center of ear pit results in minimal difference in rank-one recognition rate, 97.1% to 97.6%. The difference in performance is only a few cases and is not statistically significant. Therefore, as long as the starting point is near the ear pit, the active contour can always find the ear contour. After the ear pit is correctly located, the next step is to segment out the unoccluded portion of the ear from the image.

7.1.2 Ear Segmentation Using an Active Contour Algorithm

The 3D shape matching of the ear relies upon correct and accurate segmentation of the ear. Therefore, a robust and automatic ear segmentation algorithm is essential. Several factors contribute to the complexity of segmenting the ear out of the image. First, ear size and shape vary widely in different persons. Second, there is often hair touching or partially obscuring the ear. Third, if earrings are present, they overlap or touch the ear but should not be treated as a part of the ear. These characteristics make it hard to use a fixed template to crop the ear shape from the image. A bigger template will include too much hair, while a smaller template may lose shape information. Also it is hard to distinguish the ear from hair or earrings, especially when hair and earrings have similar color to skin or are very close to the ear.

Edges are usually defined as large magnitude changes in image gradient, which also indicates locations of intensity discontinuities. These intensity discontinuities are assumed to be in the same location as the geometric discontinuities. Even if the edges are correctly found by varying edge detection approaches, it is not clear how these edges can be connected to indicate a region of an object in the image.

The classical active contour function proposed by Kass, Witkin and Terzopoulos [46] is used to grow from the ear pit to the outline of the visible ear region.

$$E = \int_0^1 E_{int}(X(s)) + E_{ext}(X(s)) ds \quad (7.1)$$

$$E_{int} = \frac{1}{2}[\alpha|X'(s)|^2 + \beta|X''(s)|^2] \quad (7.2)$$

$$E_{ext} = E_{image} + E_{con} \quad (7.3)$$

$$E_{image} = \nabla Image(x, y) \quad (7.4)$$

$$E_{con} = -w_{con} \vec{n}(s) \quad (7.5)$$

The contour $X(s)$ starts from a closed curve within the region, and then grows under internal and external constraints, to move the curve toward local features (Equation 7.1). Following the description in [46], $X'(s)$ and $X''(s)$ denote the first and second derivative of the curve $X(s)$. α and β are weighting parameters for measuring the contour tension and rigidity, respectively. The internal function E_{int} restrains the curve from stretching or bending. The external function E_{ext} is derived from the image, so that it can drive the curve to areas with high image gradient, and lock on to close edges. It includes E_{image} and E_{con} . E_{image} is image energy, which is used to drive the curve to salient image features, such as lines, edges, terminations. In our case, we use edge feature as E_{image} . E_{con} represents the external constraints force that usually is a function of $X(s)$. Some human operators or initial constraints can be added to it. Usually E_{con} is set as zero.

The traditional active contour algorithm suffers from instability due to image force. When the initial curve is far away from image features, the curve is not attracted by the E_{image} , and would shrink into a point or a line depending on the

initial curve shape. Cohen proposed a balloon model to give more stable results [23]. The “pressure force” as E_{con} (Eq 7.5) is introduced, and when it is added to the previous curve, it pushes the curve outward so that it does not shrink to a point or a line. Here $\vec{n}(s_i)(x, y) = \frac{s_{i-1}(x,y) - s_{i+1}(x,y)}{Distance(s_{i-1}, s_{i+1})}$, s_i is the point i on curve s . Figure 7.6 shows how the active contour algorithm grows toward the image edge.

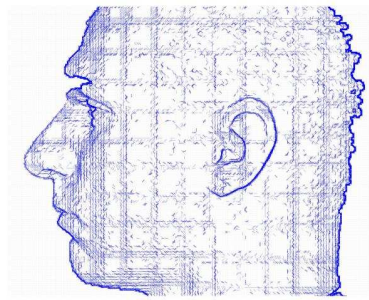
Our initial empirical studies demonstrated that the active contour algorithm is well suited for determining the ear region boundary. Starting with the ear pit determined in the previous step, the active contour algorithm grows until it finds the ear edge. Usually, there is either depth or color change, or both, along the ear edge. These attract the active contour to grow towards and stop at the ear boundary.

Initial experiments were conducted on the 3D depth image and 2D color images individually. For the 2D color images, three color spaces, RGB, HSV and YCbCr were examined. YCbCr’s Cr channel gave the best segmentation results. For the 3D images, the Z (depth) image is used. Results show that using color or depth information alone is not powerful enough for some situations, where the hair touches the ear and has similar color to skin.

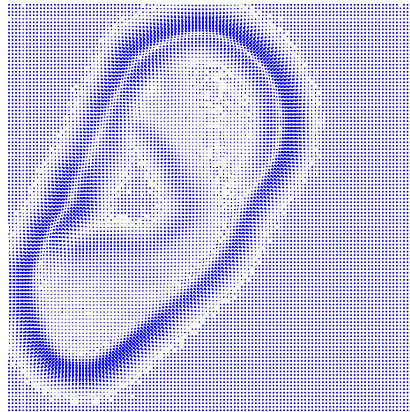
Figure 7.7 shows examples when only color or depth information is used for the active contour algorithm. In Figure 7.7(a), the hair touches the ear, but the hair color is completely different from skin color. Therefore, using only color information, the ear contour is correctly found. In Figure 7.7(b), the hair has similar color to skin, but there are large depth changes after the active contour reaches the ear edge, which helps the active contour stop growing. In these two cases, the ear edge is correctly found. However, when there is no clear color or depth change along the ear edge, it is hard for the algorithm to stop expanding.



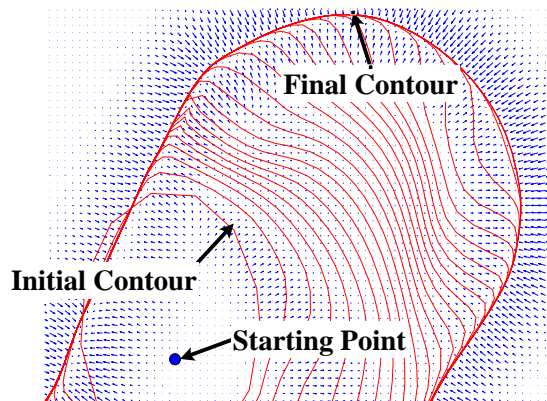
(a) Original Image



(b) Energy Map of 7.6(a)



(c) Energy Map of Ear



(d) Active Contours Growing

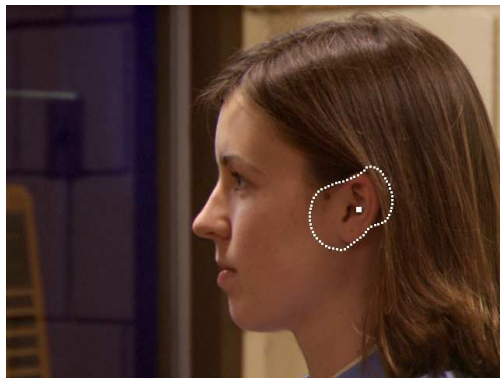
Figure 7.6. Active Contour Growing on Ear Image.
104



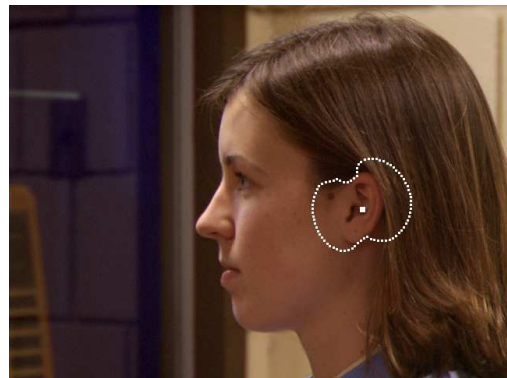
(a) Only Using Color



(b) Only Using Depth



(c) Only Using Color



(d) Only Using Depth

Figure 7.7. Active Contour Results using Only Color or Depth Information.

As shown in Figures 7.7(c) and 7.7(d), the active contour can keep growing after it reaches the ear contour.

The examples in Figure 7.7 imply that either color information or depth information alone is not adequate for finding the outline of the ear region. In order to improve the robustness of the algorithm, we combine both the color and 3D information in the active contour algorithm. To do this, the gradients of the depth images and the Cr channel from YCbCr together form the E_{image} . The E_{image} in the Equation (7.3) is replaced by the Equation (7.6). Consequently, the final energy E is represented by Equation (7.7).

$$E_{Image} = w_{depth} \nabla Image_{depth}(x, y) + w_{Cr} \nabla Image_{Cr}(x, y) \quad (7.6)$$

$$E = \int_0^1 \frac{1}{2} [\alpha |X'(s)|^2 + \beta |X''(s)|^2] \\ + w_{depth} \nabla Image_{depth}(x, y) + w_{Cr} \nabla Image_{Cr}(x, y) \\ - w_{con} \vec{n}(s) \quad (7.7)$$

In Figure 7.7, the active contour grows towards the side of the face more than expected. In order to correct this, we modify the internal energy of points to limit the expansion when there is no depth jump within a 3×5 window around the given point. The threshold for the maximum gradient within the window is set as 0.01. With these improvements, the new active contour algorithm works effectively in separating the ear from hair and earrings and the active contour stops at the jawline close to the ear.

The initial contour is an ellipse with ear pit as center, major axis as 30 pixels and minor axis as 20 pixels, and the major axis is vertical. Figure 7.8 illustrates the steps of active contour growing for a real image. And Figure 7.9 shows examples



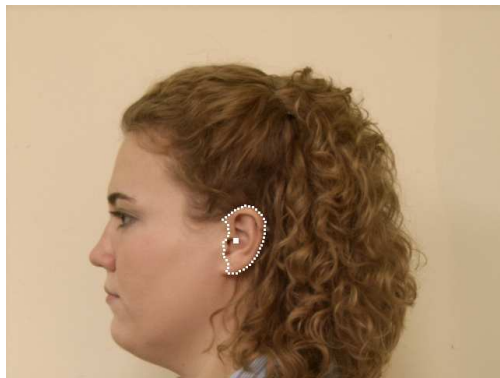
Figure 7.8. Active Contour Growing on A Real Image.



(a) Earring & Blonde Hair



(b) Blonde Hair



(c) Earring & Blonde Hair



(d) Earring



(e) Earring & Blonde Hair



(f) Earring & Blonde Hair

Figure 7.9. Effectiveness of Active Contour Algorithm when Dealing with Earring and Blonde Hair.

in which the active contour deals with hair and earrings. The 3D shape within the final contour is cropped out of the image for use in the matching algorithm.

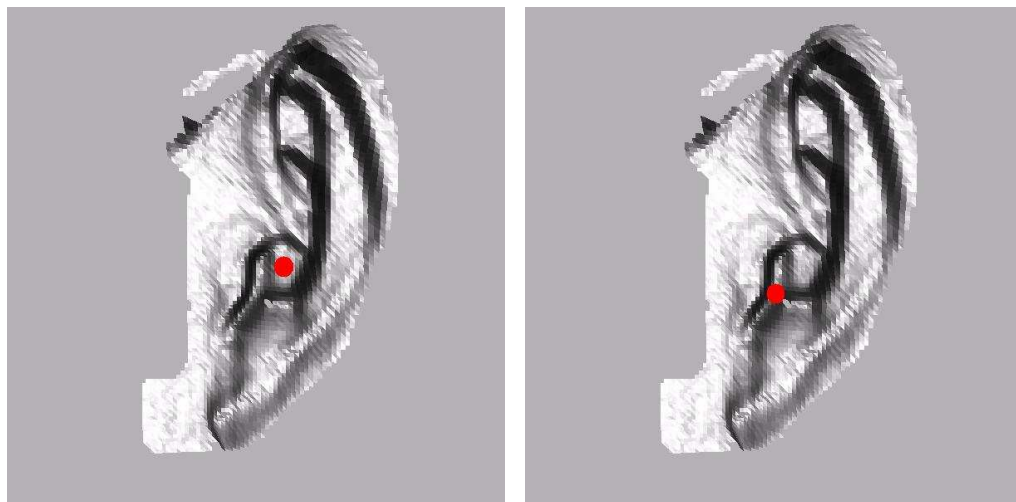
7.2 Matching 3D Ear Shape for Recognition

7.2.1 ICP-based Matching

We use ICP-based 3D shape matching to recognize the ear. Given a set of source points P and a set of model points X , the goal of ICP is to find the rigid transformation T that best aligns P with X . Beginning with a starting estimate T_0 , the algorithm iteratively calculates a sequence of transformations T_i until the registration converges. At each iteration, the algorithm computes correspondences by finding closest points, and then minimizes the mean square distance between the correspondences. A good initial estimation of the transformation is required, and all source points in P are assumed to have correspondences in the model X . The ear pit location from the automatic ear extraction is used to give the initial translation for the ICP algorithm. The following sections outline our refinements to improve the ICP algorithm for use in matching ear shapes.

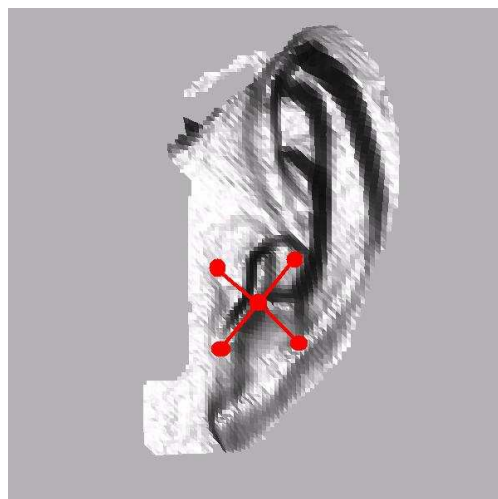
7.2.2 ICP - Perturbation of Starting Point

The initial transformation is crucial to the ICP algorithm. Given a starting transform, the ICP algorithm can always find the local minimal between the two surfaces, but it is not guaranteed that the global minimal will be returned. A good starting transform can help ICP to return the distance value which is close to the global minimum and is best to represent the similarity between two surfaces. Ear shape has rich curvature. Therefore small variations in the starting point may affect the performance.



(a) Manually Marked Ear Pit

(b) Automatically Marked Ear Pit



(c) Perturbation

Figure 7.10. Perturbation of Starting Point: The minimum error distance is 0.311 mm if using starting point from (1), the minimum error distance is 1.015 mm if using starting point from (b), and the the minimum error distance is 0.321 mm if using perturbation method, shown in (c).

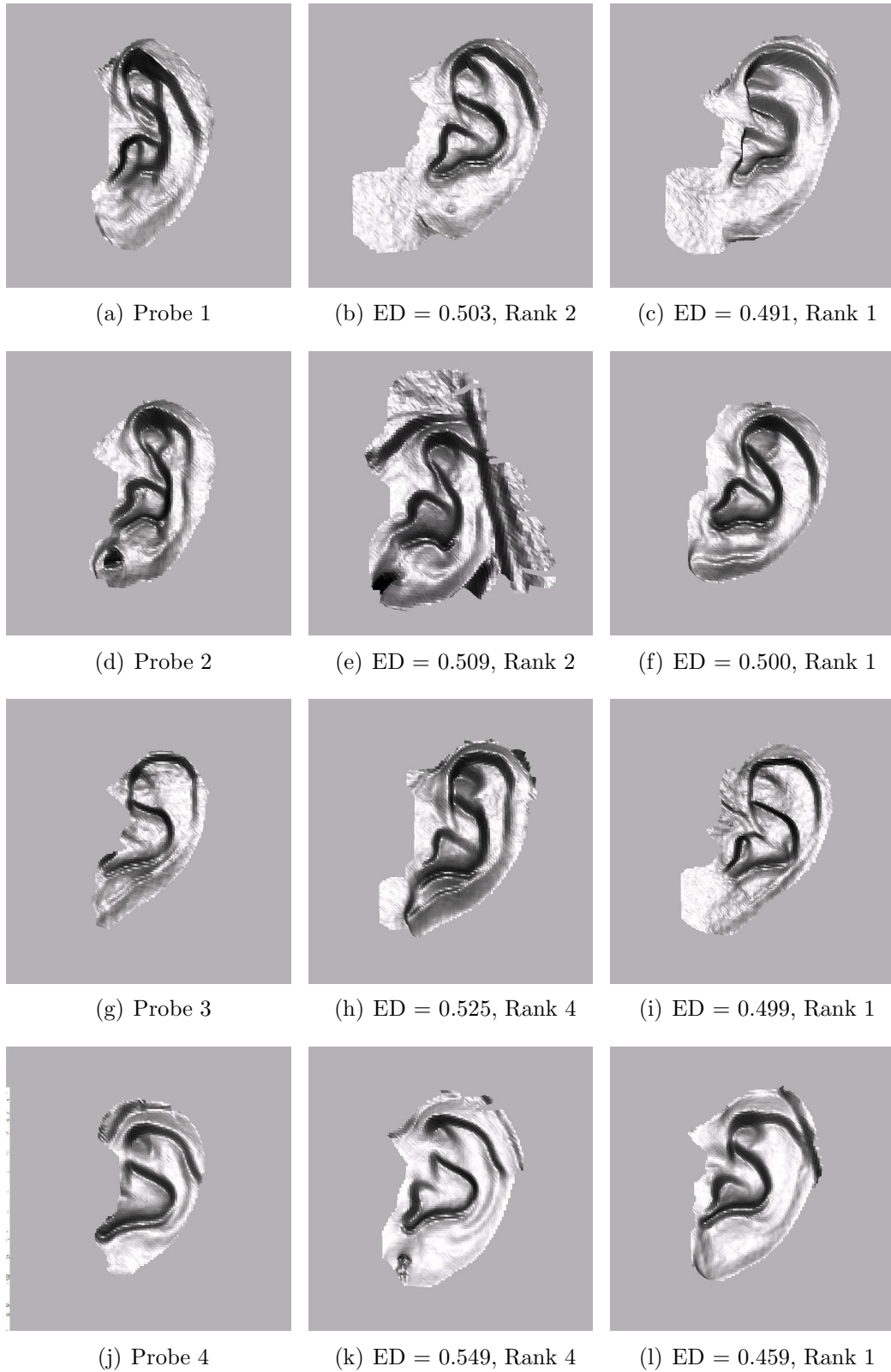


Figure 7.11. Continued on Next Page.

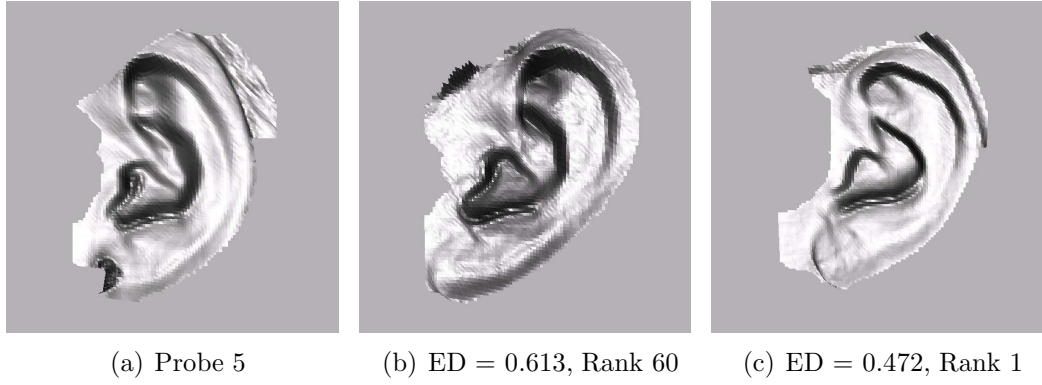


Figure 7.12. Mismatched Probe and Gallery. ED is error distance value returned from our ICP algorithm with perturbation. The first column is the probe images, the second column is the real gallery images, and the third column is the mismatched gallery images.

Figure 7.10 shows an example of mismatch due to a poor choice of starting point. In order to address this problem, we add perturbation into the matching. Our approach takes an initial point from the ear pit search, then a rectangle with width 10mm, and height 20mm is created with the initial point at its center. The algorithm examines 5 points, the initial point and each of the four corners of the square, and uses the best point as the starting point for the ICP process. The best point is selected in the following way. The five candidate points are ranked by their distance from the target surface. The distance is obtained by running the ICP algorithm for 10 iterations using the candidate point as the starting point. Various numbers of iterations for the perturbation ICP were checked and 10-iteration yields the best results. To make the difference clear, we call the ICP run in the perturbation process as perturbation ICP, and the run after perturbation and with the selected best point as starting point as registration ICP. For the old approach, the number of iterations for perturbation ICP is 0, and 40 for the registration ICP. In order to reduce the time complexity, the source surface is subsampled for

perturbation ICP, and the number of iterations is stopped when less than 40% of points find the closest point on the target surface within a threshold. Also we notice that with the perturbation, the overall iterations needed for registration ICP is reduced to 20 iterations instead of 40 iterations for the old approach. With this improvement based on considering perturbations of the starting point, the performance goes up from 97.1% to 98.8%, and the running time stays about the same as the old approach.

Figures 7.11 - 7.12 show the five cases out of 415 in which the probe image does not find the correct gallery image at the rank-one match.

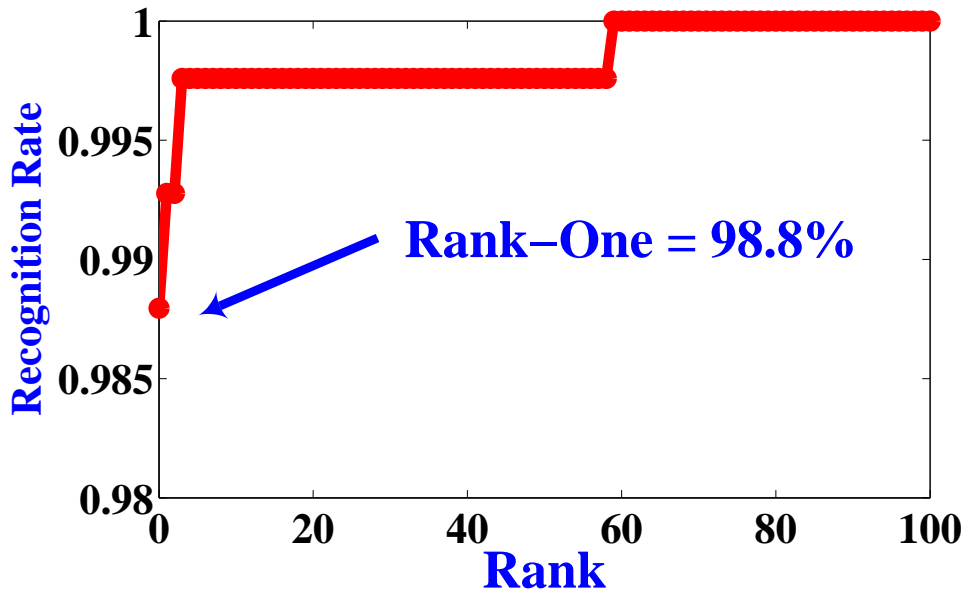
7.3 Experimental Results

In an identification scenario, our algorithm achieves a rank-one recognition rate of 98.8% on our 415-subject dataset. The CMC curve is shown in Figure 7.13(a). In a verification scenario, our algorithm achieves an equal-error rate (EER) of 0.012. The ROC curve is shown in Figure 7.13(b). These are excellent performance values in comparison to previous work in ear biometrics. Where higher performance values were reported, they were for much smaller datasets.

Also, the rank-one recognition is 94.2% (33 out of 35) for the 35 cases that involve earrings. This is a difference of just one of the 35 earring probes from the rank-one recognition rate for probes without earrings. Thus the presence of earrings in the image causes only a minimal loss in accuracy.

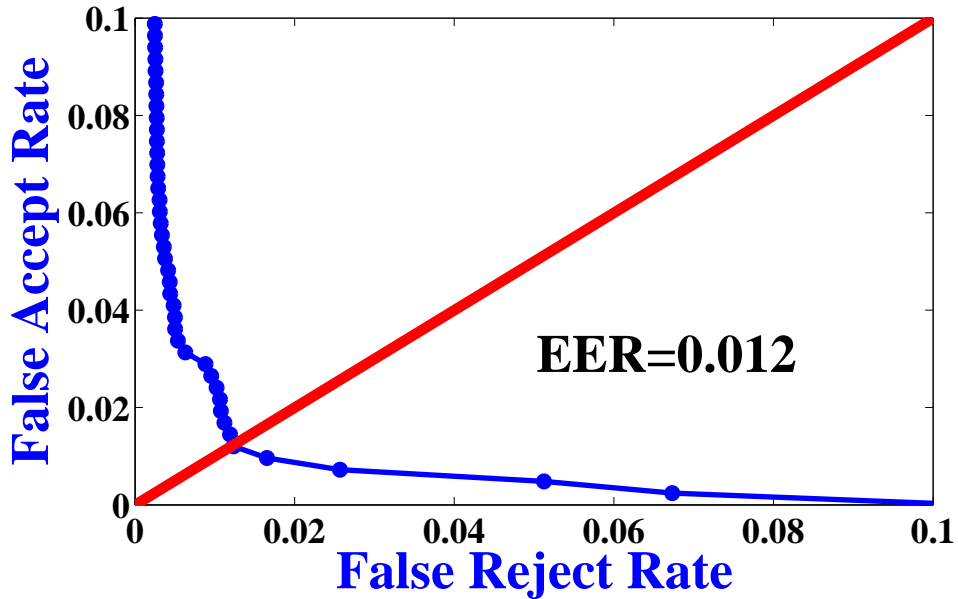
To further investigate the accuracy of our method of automatically finding the ear pit, we compared to recognition results using manually-marked ear pit location. We ran the same 415-subject person recognition experiment with everything the same except using a manually marked ear pit location rather than the auto-

Recognition Performance on 415 Subjects



(a) CMC Curve

Verification Performance on 415 Subjects



(b) ROC Curve

Figure 7.13. The Performance of Ear Recognition.

matically found location. The rank-one recognition rate resulting from using the manually-marked ear pit location is 97.1%. We can also compare to a previous algorithm that uses more detailed manual marking to extract a fixed-shape ear template [77]. Using this algorithm on the same 415-subject dataset, the rank-one recognition rate is 96.4%. Thus the performance of our automated approach seems as good as similar approaches that are assisted by manual location of landmark points.

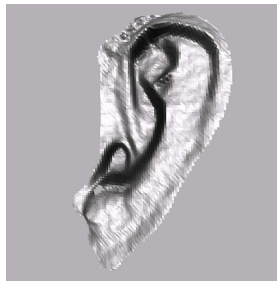
CHAPTER 8

OCCLUSION AND SYMMETRY EAR EXPERIMENTS

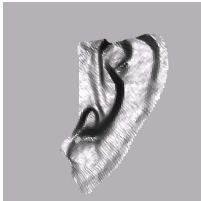
8.1 Partial Ear Experiments

So far, our experiments deal with some minor occlusion due to hair covering and earrings. And the results suggest that recognition performance is not noticeably affected by these noises. However it is still interesting to examine the algorithm's effectiveness in the presence of ear occlusion and analyze the significance of different portions of the ear. To set up the experiments, a partial ear shape is used for recognition. The images in the gallery are the same as used in the previous experiments. Portions of the probe ears are removed from different directions by a certain percentage of the horizontal or vertical distance of the boundary box. Examples are shown in Figure 8.1.

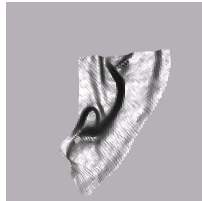
The experimental results in Table 8.1 demonstrate that the performance stays at the same level when the portion of the ear that is removed from top, bottom, left or right is in the 10 percent range, which indicates that minor hair cover will not affect the performance much, but large hair cover will certainly reduce the recognition rate. The results also suggest that the inner part of the ear contains the most useful information, and the top and right parts of the ear contain the least useful information. If we remove 10% portion of the ear from three directions: top, bottom and right section, the recognition rate is 98.3%.



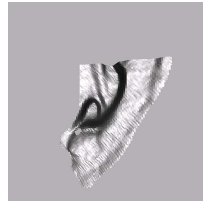
(a) Original Ear



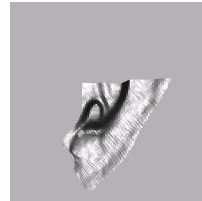
(b) Top 10%



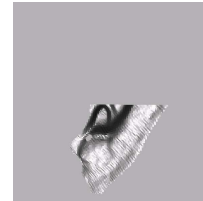
(c) Top 20%



(d) Top 30%



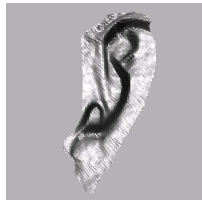
(e) Top 40%



(f) Top 50%



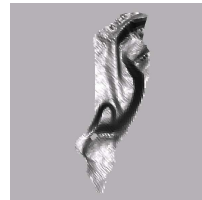
(g) Right 10%



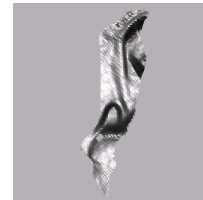
(h) Right 20%



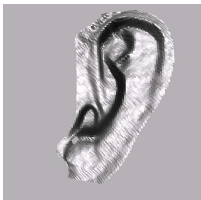
(i) Right 30%



(j) Right 40%



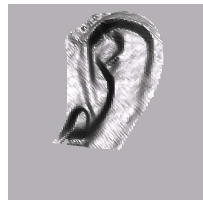
(k) Right 50%



(l) Bottom 10%



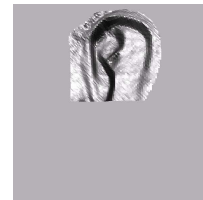
(m) Bottom 20%



(n) Bottom 30%



(o) Bottom 40%



(p) Bottom 50%



(q) Left 10%



(r) Left 20%



(s) Left 30%



(t) Left 40%



(u) Left 50%

Figure 8.1. Example of Partial Ear. The captions in part(b) through (u) indicate the percent of the ear removed.

In addition, we checked extreme cases where more than 30% of the ear is removed from different sections. The overall recognition rate is reduced by the removal of large sections. However, after removing 50% of the bottom or right portion of the ear image, the remaining shape still yields recognition rate of over 80%. This suggests that even in circumstances where much of the ear is occluded, it still has potential for recognition. Moreover, it lends support for using the ear as a biometric.

We notice that the performance increases from 97.6% to 98.3% after removing the top 20% of the ear. Figure 8.3 illustrates an example of this effectiveness on removing top 20% of the ear. It indicates that removing some top portion from the ear can actually decrease the effects of noise.

TABLE 8.1

Experiment Results Using Partial Ear with a 415-subject dataset: All the results are shown as rank-one recognition rate. The rank-one recognition rate is 97.6% by using the whole ear as probe.

Portion Removed	10%	20%	30%	40%	50%
Top	97.6%	98.3%	97.1%	82.4%	74.9%
Right	97.1%	97.1%	94.7%	90.1%	86.3%
Bottom	97.1%	96.6%	95.2%	87.4%	82.2%
Left	96.9%	94.2%	91.3%	68.4%	47.0%

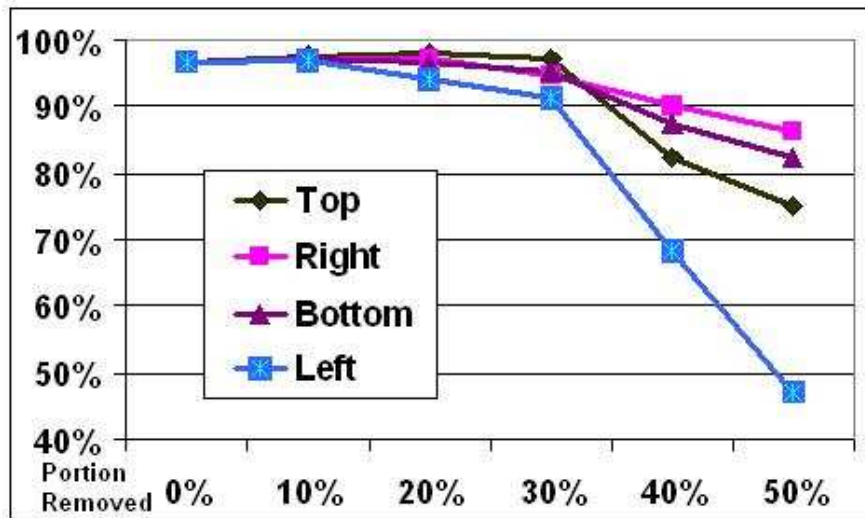


Figure 8.2. Partial Ear Experiment Results.

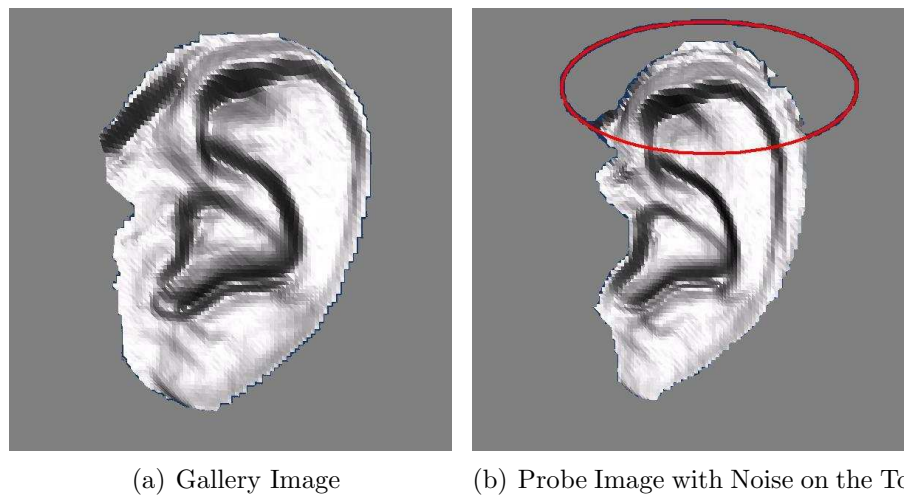


Figure 8.3. After Removing Top 20% of the Ear, the Recognition Rank Jumps From Rank 23 to Rank 1.

8.2 Ear Symmetry Experiment ¹

The ear data used in our experiments in previous chapters are gallery and probe images that are approximately straight-on views, of the same ear, acquired on different days. One interesting question to explore is the use of bilateral symmetry; that is, matching a mirrored left to the right ear of the same person. This means that for one subject we enroll his right ear and try to recognize using his mirrored left ear.

One example is shown in Figure 8.5(a) and 8.5(b). For our initial experiment to investigate this possibility, both ear images are taken on the same day. The rank-one recognition rates from matching a mirrored image of an ear are around 90%. By analyzing the results, we found that most people’s left and right ears are approximately bilaterally symmetric. But some people’s left and right ears have recognizably different shapes. Figure 8.5 shows an example of this. Thus it seems that symmetry-based ear recognition cannot be expected to be as accurate, in general, as matching two images of the same ear.

8.3 Off-angle Experiment

Another dimension of variability is the angle of view, which means that the enrolled gallery ear and the probe ear are from different angles of view. We enroll a straight-on right ear, and try to recognize a right ear viewed at some different amount of angle. The initial experiments only test on a small dataset. We have some observations, but cannot draw any strong conclusions until we have a larger dataset.

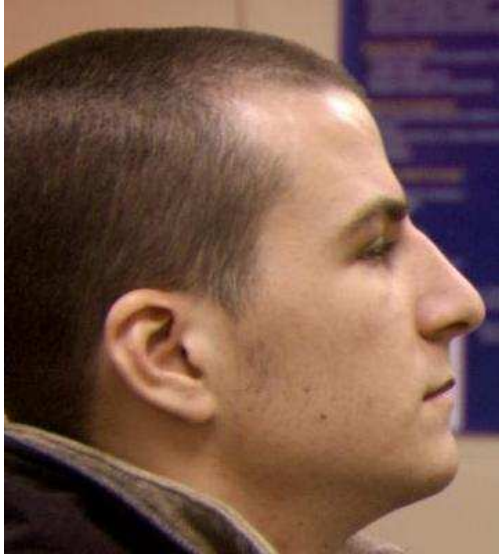
¹This section is based on the paper, “Empirical Evaluation of Advanced Ear Biometrics”, presented at the *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops: Empirical Evaluation Methods in Computer Vision, 2005* [80].



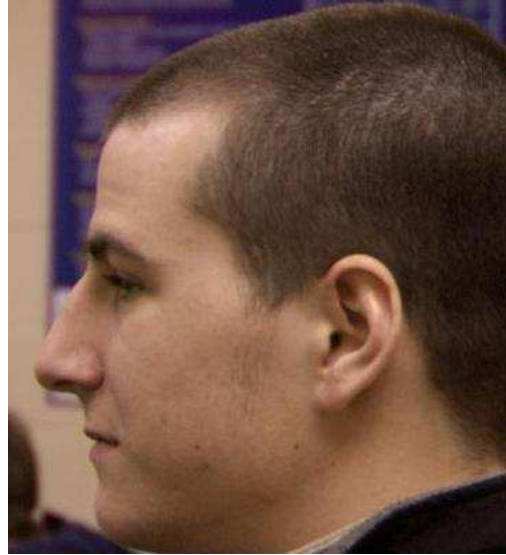
Figure 8.4. Image Acquired for Ear Symmetry Experiments.

In this experiment, there are 4 different angles of view for each ear: straight-on, 15 degrees off center, 30 degrees off center and 45 degrees off center, as shown in 8.6. The 45 degree images were taken on the first week. The 30 degree images were taken the second week. Finally, the 15 degrees and straight-on images were both taken on the third week. For each angle of ear image, we match it against all images in the different angle datasets.

There were only 24 subjects that showed up on every week. Therefore the initial experimental results, shown in Table 8.2, are obtained from a fairly small dataset. Two observations are drawn from the Table 8.2. The first is that 15 degrees and 30 degrees off center have better overall performance than the straight on and 45 degree off center. This observation makes sense, since there is more ear area exposed to the camera when the face is 15 degrees and 30 degrees off center. Also matching is generally good for 15 degrees off or less, but gets worse for more than 15 degrees. The second observation is that the elements in Table 8.2 are symmetric. A larger dataset is required if we want to verify our observation and draw any conclusions.



(a) Example 1



(b) Example 1



(c) Right ear

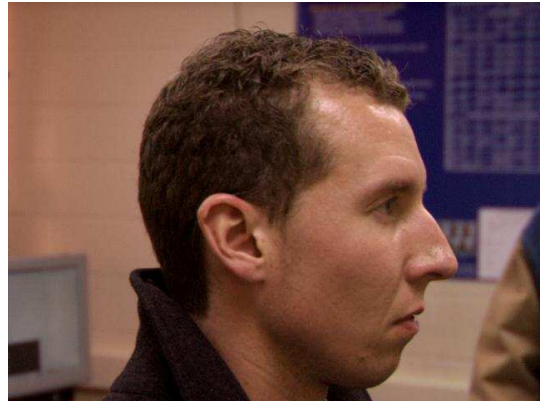


(d) Mirrored Left ear

Figure 8.5: Examples of Asymmetric Ears.



(a) Straight-on



(b) 15 Degrees off



(c) 30 Degrees off



(d) 45 Degrees off

Figure 8.6. Image acquired for Off Angle Experiments.

TABLE 8.2

Off Angle Experiments with a 24-Subject Dataset.

	Straight-on	15° off	30° off	45° off	Average
straight-on		100%	87.5%	70.8%	86.1%
15° off	100%		100%	87.5%	95.8%
30° off	87.5%	100%		95.8%	94.4%
45° off	79.2%	87.5%	100%		88.9%
Average	88.9%	95.8%	95.8%	84.7%	

CHAPTER 9

COMPARISON WITH OTHER BIOMETRICS

9.1 Comparison to 2D “Eigen-Ear” Results

In order to compare with the state of the art reported in the literature [14], we run the “eigen-ear” experiments on the subset of persons that have images usable for 2D eigen-ear. Chang obtained 73% rank-one recognition with 88 persons in the gallery and a single time-lapse probe image per person [14]. Our rank one recognition rate for PCA-based ear recognition using 2D intensity images for the first 88 persons in our 415-person dataset is 76.1%, which is similar to the result obtained by Chang, even though we used a completely different image data set acquired by a different sensor and used different landmark points. For the same 88 persons, the ICP-based ear recognition gives a 98.9% rank-one recognition rate, which is statistically significantly better than the 2D “Eigen-Ear” Results. This indicates that our algorithm represents a substantial advance relative to previous work in ear biometrics.

9.2 Comparison to Commercial Face Recognition Results

To relate the performance level achieved here to the commercial state of the art, we experimented with FaceIT version 6.1 [1]. We used frontal face images which were taken on the same date for the same subject as the ear images used in

the experiments. Only 409 out of the original 415 subjects also had good quality frontal face images. None of the 6 people who are missing in the face experiment has ear biometric recognition errors. The images are taken by Canon Power Shot G2 and Nikon D70, under controlled studio lighting and with no lighting condition change between gallery and probe [58]. Each image has resolution 1704x2272. With normal expression both in gallery and probe, the rank one recognition rate is 98.7%. Subjectively, our impression is that FaceIT version 6 has improved performance over version 5 in a noticeable way.

The difference in commercial face recognition performance and our ear biometrics performance is small, just 98.7% to 97.6%. This indicates that ear biometrics potentially has similar recognition power to face biometrics, and should motivate additional work in this area. Currently there has been far less research investment in ear biometrics than in face biometrics.

9.3 Scalability of Recognition with Dataset Size

Scaling of performance with dataset size is a critical issue in biometrics. For example, a decrease in recognition performance with increased dataset size was observed in FRVT 2002 for 2D face recognition:

For every doubling of database size, performance decreases by two to three overall percentage points. In mathematical terms, identification performance decreases linearly with respect to the logarithm of the database size. [57]

When the gallery size becomes bigger, the possibility of a false match increases. Usually, some techniques scale better to larger datasets than others. Figure 9.1 shows the scalability of our ear recognition algorithm. For a given data set size less than 415, the images are randomly selected from the 415 persons 100 times. The

blue line shows the average performance of these 100 random selected datasets. The red line shows the range of the performance at each dataset size. When the gallery size increases, the recognition performance is largely unchanged. At least within the range of dataset size considered here, it appears that doubling the dataset size incurs less of a performance decrease than found in FRVT 2002 for 2D face recognition. Going from 100 persons to 400 persons, doubling twice, reduces performance approximately 3%.

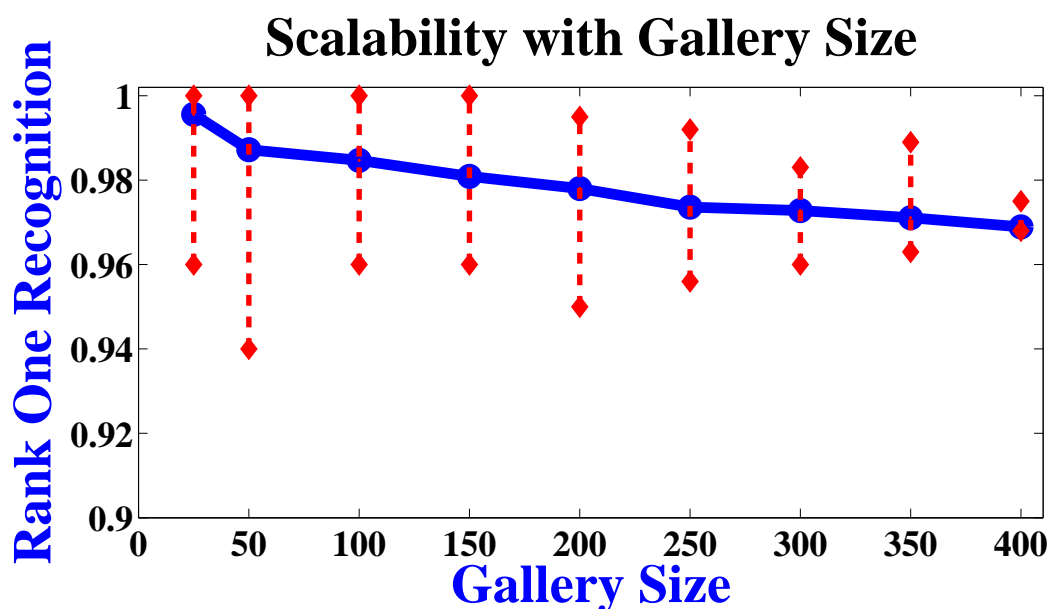


Figure 9.1. Scalability of Proposed Algorithm's Performance with Increased Dataset Size.

CHAPTER 10

CONCLUSIONS AND FUTURE WORK

10.1 Conclusions

We have presented experimental results for three different approaches to 3D ear recognition and a PCA-based approach to 2D ear recognition. Our results are based on the largest experimental dataset to date for ear biometrics, with 2D and 3D images acquired for over 400 persons on two different dates. This is the most comprehensive investigation of 3D ear recognition to be reported to date, the largest experimental evaluation of 2D ear recognition, and the first (only) comparison of 2D and 3D ear recognition.

Also we have presented a fully automatic ear recognition system using 2D and 3D information. The automatic ear extraction algorithm can crop the ear region from the profile image, separating ear from hair and earring. The recognition subsystem uses an ICP-based approach for 3D shape matching. The experimental results demonstrate the power of our automatic ear extraction algorithm and 3D shape matching when applied to biometric identification. The system has a 97.6% rank-one recognition rate on a time-lapse dataset of 415 persons. The system as outlined in this dissertation is a significant and important step beyond existing work in this area. The system is fully automatic, handling preprocessing, cropping and matching. The system addresses issues that plagued earlier attempts to use

3D ear images for recognition, specifically partial occlusion of the ear by hair and earrings.

The size of the dataset and the scalability tests demonstrate the applicability of our approach for 3D ear recognition. Comparisons to a 2D eigen-ear approach and to a commercial face recognition system further demonstrate the practical strengths of this work. Though there are certain issues that need to be addressed, such as the matching of even more occluded ear surfaces to full gallery images and the computational expense of the ICP algorithm, we are confident that the approach presented here represents a viable method of using the ear for human identification.

ICP in its current form is computationally expensive. We believe this is still a major challenge to its use in biometric identification. We are currently looking at ways of reducing the computational expense of ICP while maintaining a high level of accuracy. One effort toward this goal is presented in chapter 5. Additionally, noisy or incomplete data is inevitable in a production system. We are also examining the robustness and applicability of ICP to matching partial ear surfaces to full-ear gallery surface.

Our work extensively investigated the possibility of using ear as a biometric. The focus of the work is on the improvement of performance and reducing the computation time. The experimental results suggest the potential power of ear biometric. The automatic system proposed demonstrates its effectiveness and efficiency.

10.2 Future Work

There are several directions for future work. Our automatic ear extraction works well for ear segmentation by employing both color and depth information. However, the active contour algorithm may fail if there is no gradient change in either color or depth image. An improvement might focus on using shape and texture constraints to help the segmentation. It might be possible to build in some preferred shape, like an ellipse, or to penalize small irregular parts to the outline.

Even though our ear dataset is the largest to date, it is still too small to demonstrate practical use for many applications. A larger dataset is necessary to validate our work. In addition, it would be interesting to look further at partial ear occlusion and off-angle ear experiments. The initial experimental results allow us to make some observations, but a larger dataset is required if we want to verify our observations and draw any rigorous conclusions.

Speed and recognition accuracy remain important issues. We have proposed several enhancements to improve the speed of the algorithm, but the algorithm might benefit from adding feature classifiers. We have both 2D and 3D data and they are registered to each other, which should make it straightforward to test multi-modal algorithms. In our multi-modal biometric experiments, we considered the combination of 2D and 3D ear at the matching score level. It would be interesting to look at feature extraction level or decision level combination.

APPENDIX A

EAR BIOMETRICS DATA

Our experiments use the biometric database collected at the University of Notre Dame during academic year 2003 and 2005. The data is available under University of Notre Dame (UND) Biometrics Database distribution. Detailed information can be found at <http://www.nd.edu/~cvrl/UNDBiometricsDatabase.html>.

There are 1680 images enrolled based on use in this work, and each has an abs format (3D data) file and a ppm format (color image) file. In total, the images are about 3.9 G in size.

APPENDIX B
EARRING DATA

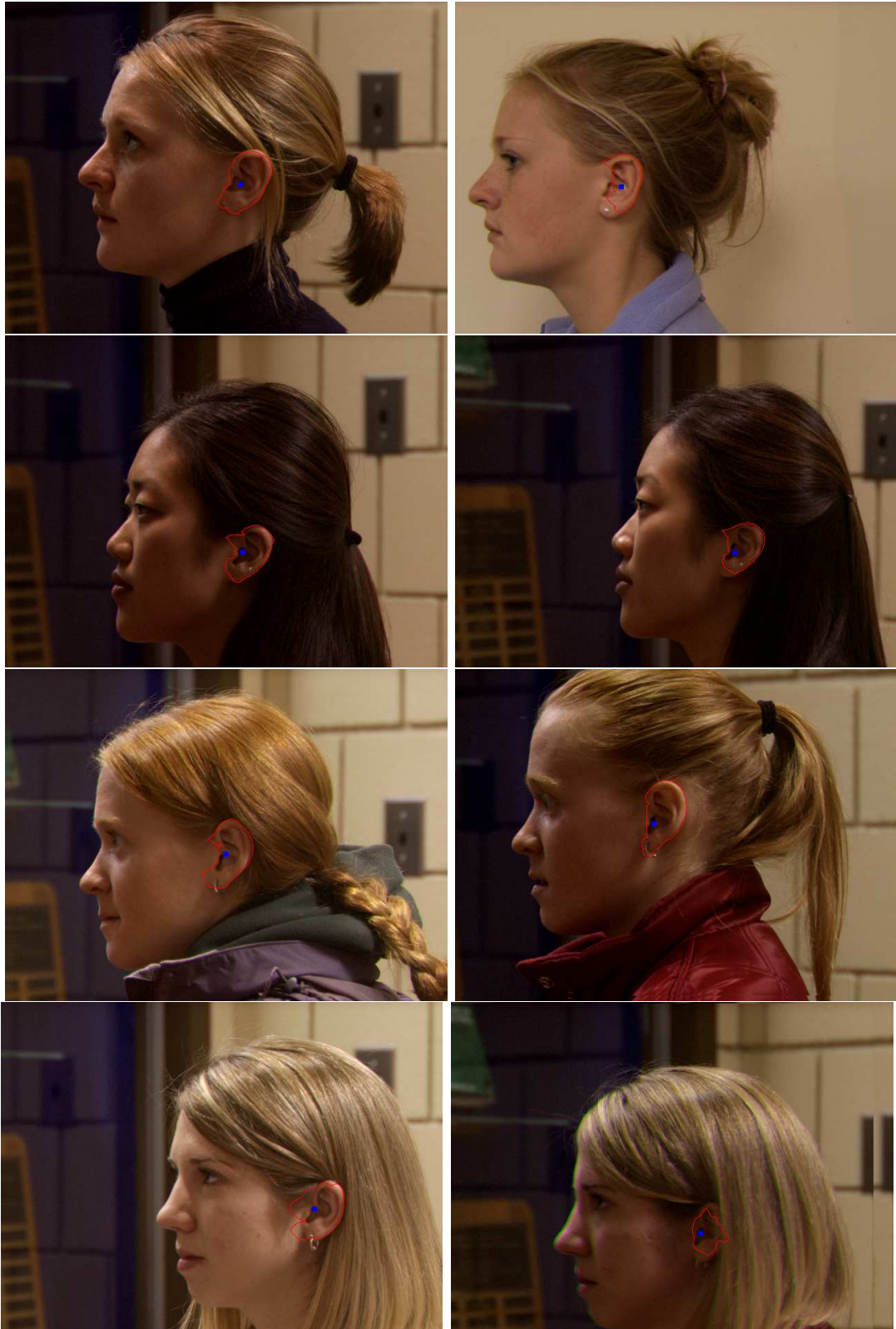


















APPENDIX C

MINOR HAIR COVERING











BIBLIOGRAPHY

1. http://www.identix.com/products/pro_faceit.html.
2. <http://konicaminolta.com/products/instruments/vivid/vivid910.html>.
3. B. Achermann and H. Bunke. Classifying range images of human faces with Hausdorff distance. In *International Conference on Image Processing*, pages 809–813, 2000.
4. R. Benjemaa and F. Schmitt. Fast global registration of 3D sampled surfaces using a multi-z-buffer technique. In *International Conference on Recent Advances in 3-D Digital Imaging and Modeling (3DIM '97)*, pages 113–119, 1997.
5. P. Besl and N. McKay. A method for registration of 3-D shapes. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 14:239–256, 1992.
6. P. J. Besl and R. C. Jain. Invariant surface characteristics for 3D object recognition in range images. In *Computer Vision Graphics Image Processing*, volume 33, pages 30–80, 1986.
7. R. Beveridge, K. She, B. Draper, and G. Givens. Evaluation of face recognition algorithm (release version 5.0). In *URL: www.cs.colostate.edu/evalfacerec/index.html*.
8. B. Bhanu and H. Chen. Human ear recognition in 3D. In *Workshop on Multimodal User Authentication*, pages 91–98, 2003.
9. E. Bigún, J. Bigún, and S. S. Fischer. Expert conciliation for multi modal person authentication systems using Bayesian statistics. In *Proceedings of the International Conference on Audio and Video-Based Biometric Person Authentication*, pages 291–300, 1997.
10. R. Brunelli and D. Falavigna. Person identification using multiple cues. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 12:955–966, 1995.

11. M. Burge and W. Burger. *Ear biometrics*, chapter 13, pages 273–286. Biometrics: Personal Identification in Networked Society. Kluwer Academic, 1999.
12. M. Burge and W. Burger. Ear biometrics in computer vision. In *15th International Conference of Pattern Recognition*, volume 2, pages 822–826, 2000.
13. J. Chang. *New multi-biometric approaches for improved person identification*. PhD thesis, Department of Computer Science and Engineering, University of Notre Dame, USA, 2004.
14. K. Chang, K. Bowyer, and V. Barnabas. Comparison and combination of ear and face images in appearance-based biometrics. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 25:1160–1165, 2003.
15. K. Chang, K. Bowyer, and P. Flynn. Face recognition using 2D and 3D facial data. In *Workshop on Multimodal User Authentication*, pages 25–32, 2003.
16. K. Chang, K. Bowyer, and P. Flynn. An evaluation of multi-modal 2D+3D face biometrics. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 27:619–624, 2005.
17. K. Chang, K. W. Bowyer, and P. J. Flynn. Effects of facial expression in 3D face recognition. In *Biometric Technology for Human Identification II, Proceedings of SPIE*, volume 5779, pages 132–143, March 2005.
18. H. Chen and B. Bhanu. Contour matching for 3D ear recognition. In *Seventh IEEE Workshop on Application of Computer Vision*, pages 123–128, 2005.
19. H. Chen and B. Bhanu. Human ear detection from side face range images. In *International Conference on Image Processing*, pages 574–577, 2004.
20. Y. Chen and G. Medioni. Object modeling by registration of multiple range images. *Image and Vision Computing*, 10:145–155, 1992.
21. M. Choraš. Ear biometrics based on geometrical feature extraction. *Electronic Letters on Computer Vision and Image Analysis*, 5:84–95, 2005.
22. J. G. Cleary. Analysis of an algorithm for finding nearest neighbors in Euclidean space. *ACM Transactions on Mathematical Software (TOMS)*, 5:183–192, 1979.
23. L. D. Cohen. On active contour models and balloons. *Computer Vision, Graphics, and Image Processing. Image Understanding*, 53(2):211–218, 1991. URL citeseer.ist.psu.edu/cohen91active.html.

24. D. Cremers. *Statistical Shape Knowledge in Variational Image Segmentation*. PhD thesis, Department of Mathematics and Computer Science, University of Mannheim, Germany, July 2002.
25. J. Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.
26. B. A. Draper, W. S. Yambor, and J. R. Beveridge. Analyzing PCA-based face recognition algorithms: Eigenvector selection and distance measures. In *Second Workshop on Empirical Evaluation in Computer Vision*, pages 39 – 60, 2000.
27. J. Feldmar, N. Ayache, and F. Betting. 3D-2D projective registration of free-form curves and surfaces. *Computer Vision and Image Understanding*, 65: 403–424, 1997.
28. P. Flynn and A. Jain. Surface classification: Hypothesis testing and parameter estimation. In *IEEE Conference on Computer Vision Pattern Recognition*, pages 261–267, 1988.
29. M. Friedman. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32:675–701, 1937.
30. M. Friedman. A comparison of alternative tests of significance for the problem of m rankings. *The Annals of Mathematical Statistics*, 11:86–92, 1940.
31. N. Gelfand, S. Rusinkiewicz, L. Ikemoto, and M. Levoy. Geometrically stable sampling for the ICP algorithm. In *Fourth International Conference on 3-D Digital Imaging and Modeling*, pages 260–267, October 06 - 10, 2003.
32. M. Greenspan and G. Godin. A nearest neighbor method for efficient ICP. In *Proceedings of the 3rd International Conference on 3-D Digital Imaging and Modeling*, pages 161–170, 2001.
33. W. L. Hays. *Statistics (4th Edition)*. Thomson Learning, 1996.
34. A. Hilton, A. J. Stoddart, J. Illingworth, and T. Winder. Marching triangles: Range image fusion from complex object modeling. In *International Conference on Image Processing*, volume 2, pages 381 – 384, 1996.
35. A. Hoogstrate and E. H. Van den Heuvel. Ear identification based on surveillance camera’s images. In <http://www.forensic-evidence.com/site/ID/IDearCamera.html>, 2000.

36. R.-L. Hsu, M. Abdel-Mottaleb, and A. Jain. Face detection in color images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 24:696–706, 2002.
37. D. Hurley, M. Nixon, and J. Carter. Force field energy functionals for image feature extraction. *Image and Vision Computing Journal*, 20:429–432, 2002.
38. D. Hurley, M. Nixon, and J. Carter. Force field energy functionals for ear biometrics. *Computer Vision and Image Understanding*, 98:491–512, 2005.
39. D. Huttenlocher, G. Klanderman, and W. Rucklidge. Comparing images using the Hausdorff distance. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 15:850–863, 1993.
40. A. Iannarelli. *Ear Identification*. Paramount Publishing Company, 1989.
41. A. Jain and A. Ross. Multibiometric system. *Communications of the ACM*, 47:34–40, 2004.
42. A. E. Johnson. <http://www-2.cs.cmu.edu/vmr/software/meshtoolbox>.
43. A. E. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3D scene. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 21:433–449, 1999.
44. T. Jost and H. Hügli. Multi-resolution ICP with heuristic closest point search for fast and robust 3D registration of range images. In *Fourth International Conference on 3-D Digital Imaging and Modeling*, pages 427–433, October 06 - 10, 2003.
45. T. Jost and H. Hügli. Fast ICP algorithms for shape registration. *Lecture Notes in Computer Science*, 2449:91–99, 2002.
46. M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1:321–331, 1987.
47. J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 20:226–239, 1998.
48. J. Koenderink and A. van Doorn. Surface shape and curvature scales. *Image and Vision Computing*, 10:557–565, 1992.
49. A. Kumar, D. C. M. Wong, H. C. Shen, and A. K. Jain. Personal verification using palmprint and hand geometry biometric. In *Proceedings of Fourth International Conference on AVBPA*, pages 668–678, 2003.

50. K.-L. Low and A. Lastra. Reliable and rapidly-converging ICP algorithm using multiresolution smoothing. In *Fourth International Conference on 3-D Digital Imaging and Modeling*, pages 171–178, October 06 - 10, 2003.
51. X. Lu, D. Colbry, and A. K. Jain. Three-dimensional model based face recognition. In *7th IEEE Workshop on Applications of Computer Vision*, pages 156–163, 2005.
52. G. Medioni and R. Waupotitsch. Face recognition and modeling in 3D. In *IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2003)*, pages 232 – 233, 2003.
53. K. Messer, J. Matas, J. Kittler, J. Luetttin, and G. Maitre. XM2VTSDB: The extended M2VTS Database. In *Audio- and Video-based Biometric Person Authentication*, pages 72–77, 1999.
54. B. Moreno, A. Sanchez, and J. Velez. On the use of outer ear images for personal identification in security applications. In *IEEE International Carnaham Conference on Security Technology*, pages 469–476, 1999.
55. P. B. Nemenyi. *Distribution-free multiple comparisons*. PhD thesis, Princeton University, 1963.
56. P. Phillips, H. Moon, S. Rizvi, and P. Rauss. The FERET evaluation methodology for face recognition algorithms. volume 22(10), pages 1090–1104, 2000.
57. P. Phillips, P. Grother, R. Micheals, D. Blackburn, E. Tabassi, and J. Bone. FRVT 2002: Overview and summary. In <http://www.frvt.org/FRVT2002/default.htm>, 2003.
58. P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *IEEE Conference on Computer Vision Pattern Recognition*, volume 1, pages 947–954, 2005.
59. K. Pulli. Multiview registration for large data sets. In *Second International Conference on 3-D Imaging and Modeling (3DIM '99)*, pages 160–168, October, 1999.
60. K. Pun and Y. Moon. Recent advances in ear biometrics. In *Proceedings of the Sixth International Conference on Automatic Face and Gesture Recognition*, pages 164–169, May, 2004.
61. A. Ross and A. Jain. Information fusion in biometrics. *Pattern Recognition Letters*, 24:2115–2125, 2003.

62. A. Ross and A. K. Jain. Multimodal biometrics: An overview. In *Proc. of 12th European Signal Processing Conference (EUSIPCO)*, pages 1221–1224, 2004.
63. S. Rusinkiewicz and M. Levoy. Efficient variants of the ICP algorithm. In *Third International Conference on 3D Digital Imaging and Modeling (3DIM)*, pages 145–152, June 2001.
64. S. L. Salzberg. One comparing classifiers: Pitfalls to avoid and a recommended approach. *Data Mining and Knowledge Discovery*, 1:317–328, 1997.
65. W. Schroeder, K. Martin, and B. Lorensen. *The Visualization Toolkit, An Object-Oriented Approach To 3D Graphics, 3rd edition*. Kitware, Inc., 2003.
66. H.-Y. Shum, M. Hebert, K. Ikeuchi, and R. Reddy. An integral approach to free-form object modeling. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 19:1366 – 1370, 1997.
67. D. Simon. *Fast and Accurate Shape-Based Registration*. PhD thesis, Robotics Institute, Carnegie Mellon University, December, 1996.
68. D. Simon, M. Hebert, and T. Kanade. Techniques for fast and accurate intra-surgical registration. In *First International Symposium on Medical Robotics and Computer Assisted Surgery*, pages 90–97, September 1994.
69. L. Sirovich and M. Kirby. Low dimensional procedure for characterization of human faces. *Journal of the Optical Society of America*, 4:519–524, March 1987.
70. G. Turk and M. Levoy. Zippered polygon meshes from range images. In *Proceeding of ACM Siggraph 94*, pages 311 – 318, 1994.
71. M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3:71–86, 1991.
72. B. Victor, K. Bowyer, and S. Sarkar. An evaluation of face and ear biometrics. In *16th International Conference of Pattern Recognition*, pages 429–432, 2002.
73. Y. Wang, T. Tan, and A. K. Jain. Combining face and iris biometrics for identity verification. In *Proceedings of Fourth International Conference on AVBPA*, pages 805–813, 2003.
74. D. L. Woodard and P. J. Flynn. Finger surface as a biometric identifier. *Computer Vision and Image Understanding*, 100:357–384, 2005.
75. C. Xu and J. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transaction on Image Process*, 7:359–369, 1998.

76. P. Yan. Edge detector performance evaluation using pixel-level ground truth and using an object recognition task. Master's thesis, Department of Computer Science and Engineering, University of Notre Dame, USA, Nov. 2003.
77. P. Yan and K. W. Bowyer. Ear biometrics using 2D and 3D images. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, page 121, 2005.
78. P. Yan and K. W. Bowyer. Multi-Biometrics 2D and 3D ear recognition. In *Audio- and Video-based Biometric Person Authentication*, pages 503–512, 2005.
79. P. Yan and K. W. Bowyer. A fast algorithm for ICP-based 3D shape biometrics. In *Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID 2005)*, pages 213–218, October 2005, Buffalo, New York.
80. P. Yan and K. W. Bowyer. Empirical evaluation of advanced ear biometrics. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, page 41, 2005.
81. P. Yan and K. W. Bowyer. ICP-based approaches for 3D ear recognition. In *Biometric Technology for Human Identification II, Proceedings of SPIE*, volume 5779, pages 282–291, 2005.
82. T. Yuizono, Y. Wang, K. Satoh, and S. Nakayama. Study on individual recognition for ear images by using genetic local search. In *Proceedings of the 2002 Congress on Evolutionary Computation*, pages 237–242, 2002.
83. D. Zhang and M. Hebert. Harmonic maps and their applications in surface matching. In *IEEE Conference on Computer Vision Pattern Recognition*, volume 2, pages 524–530, 1999.
84. Z. Zhang. Iterative point matching for registration of freeform curves and surfaces. *International Journal of Computer Vision*, 13:119–152, 1994.
85. T. Zinsser, J. Schmidt, and H. Niemann. Performance analysis of nearest neighbor algorithms for ICP registration of 3-D point sets. In *8th International Fall Workshop of Vision, Modeling, and Visualization*, November 19 - 21, 2003.

<p><i>This document was prepared & typeset with L^AT_EX 2_ε, and formatted with NDDiss2_ε classfile (v1.0[2004/06/15]) provided by Sameer Vijay.</i></p>
--