1

Introduction

"The ear, thanks to these multiple small valleys and hills which furrow across it, is the most significant factor from the point of view of identification. Immutable in its form since birth, resistant to the influences of environment and education, this organ remains, during the entire life, like the intangible legacy of heredity and of the intrauterine life."

A. Bertillon, 1890 [5, 6].

1.1 Ear Biometrics and History

Biometrics deal with recognition of individuals based on their physiological or behavioral characteristics [7]. Since the characteristics for each human are distinct and unique, biometrics provide a more reliable system than the other traditional systems, such as identity cards, keys, and passwords. Researchers have done extensive biometric studies on fingerprints, faces, palm prints, irises and gaits. Ear biometrics, on the other hand, has received scant attention.

In claiming the ear as a class of biometrics, we have to show that it is viable as a class of biometrics (i.e., universality, uniqueness, and permanence [7]). After decades of research of anthropometric measurements of ear photographs of thousands of people, Iannarelli has found that no two ears are alike, even in the cases of identical and fraternal twins, triplets, and quadruplets [8]. He has also found that the structure of the ear does not change radically over time.

Iannarelli has appeared personally as an expert witness in many court cases [6]. In the preface of his book [8], Iannarelli says:

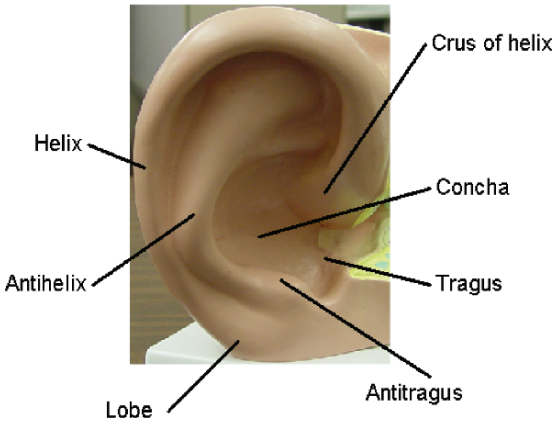
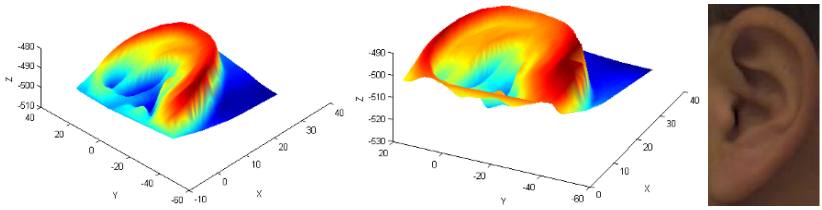


Fig. 1.1. The external ear and its anatomical parts.



(a) (b) (c)

Fig. 1.2. Range image and color image captured by a Minolta Vivid 300 camera. In images (a) and (b), the range image of one ear is displayed as the shaded mesh from two viewpoints (The units of x, y and z are in millimeters). Image (c) shows the color image of the ear.

Through 38 years of research and application in earology, the author has found that in literally thousands of ears that were examined by visual means, photographs, ear prints, and latent ear print impressions, no two ears were found to be identical—not even the ears of any one individual. This uniqueness held true in cases of identical and fraternal twins, triplets, and quadruplets.

By the statement "no two ears were found to be identical—not even the ears of any one individual," we understand and find based on our own analysis and results presented in this book, that the variations between the ear signatures of an individual are significantly less than the variations among different people. This is the scientific basis for considering the ear as a class of biometrics.

The anatomical structure of the human ear is shown in Figure 1.1. These include the outer rim (helix) and ridges (anti-helix) parallel to the helix, the lobe, the concha (hollow part of ear) and the tragus (the small prominence of cartilage over the meatus). The crus of helix is the beginning of helix.

1.2 2D and 3D Ear Biometrics

Researchers have developed several biometric techniques using the 2D intensity images of human ears [7, Chap. 13] [9–11]. The performance of these techniques is greatly affected by the pose variation and imaging conditions. However, an ear can be imaged in 3D using a range sensor which provides a registered color and range image. Figure 1.2 shows an example of a range image and the registered color image acquired by a Minolta Vivid 300 camera. A range image is relatively insensitive to illuminations and contains surface shape information related to the anatomical structure, which makes it possible to develop a robust 3D ear biometrics. Examples of ear recognition using 3D data are [1, 12–19].

1.3 3D Ear Databases

The experiments were performed on the dataset collected by us (the UCR dataset) and the University of Notre Dame public dataset (the

UND dataset).[1] In the UCR dataset there is no time lapse between the gallery and probe for the same subject, while there is a time lapse of a few weeks (on the average) in the UND dataset.

1.3.1 The UCR Dataset

The data are captured by a Minolta Vivid 300 camera which is shown in Figure 1.3. This camera uses the light-stripe method to emit a horizontal stripe light to the object and the reflected light is then converted by triangulation into distance information. The camera outputs a range image and its registered color image in less than one second. The range image contains 200×200 grid points and each grid point has a 3D coordinate (x, y, z) and a set of color (r, g, b) values. During the acquisition, 155 subjects sit on a chair about 0.55–0.75 m from the camera in an indoor office environment. The first shot is taken when a subject's left-side face is approximately parallel to the image plane; two shots are taken when the subject is asked to rotate his/her head to the left and to the right side within $\pm 35^\circ$ with respect to his/her torso. During this process, there can be some face tilt as well, which is not measured. A total of six images per subject are recorded. A total of 902 shots are used for the experiments since some shots are not properly recorded. Every person has at least four shots. The average number of points on the side face scans is 23,205. There are there different poses in the collected data: frontal, left, and right. Among the total 155 subjects, there are 17 females. Among the 155 subjects, 6 subjects have earrings and 12 subjects have their ears partially occluded by hair (with less than 10% occlusion). Figure 1.4 shows side face range images and the corresponding color images of six people collected in our database. The pose variations, the earrings, and the hair occlusions can be clearly seen in this figure.

1.3.2 The UND Dataset

The data are acquired with a Minolta Vivid 910 camera. The camera outputs a 480×640 range image and its registered color image of the same size. During acquisition, the subject sits approximately 1.5 m away from the sensor with the left side of the face toward the camera.

[1] http://www.nd.edu/∼cvrl/UNDBiometricsDatabase.html, Collection F and a subset of Collection G

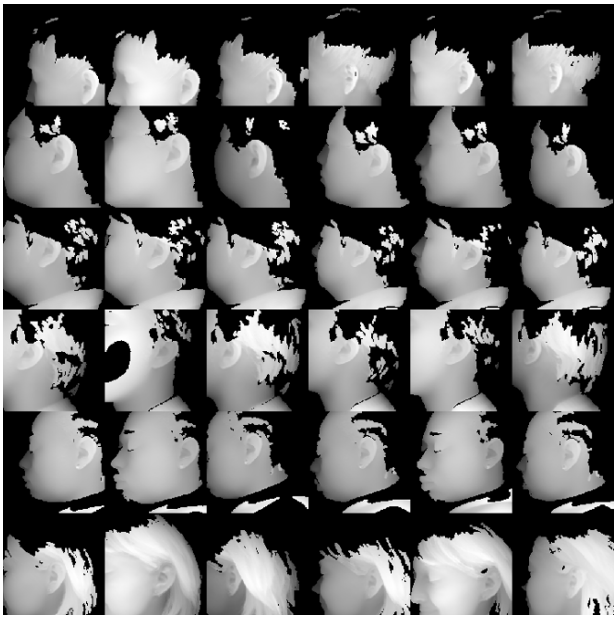
Fig. 1.3. The Minolta Vivid 300.

In Collection F, there are 302 subjects with 302 time-lapse gallery-probe pairs. Figure 1.5 shows side face range images and the corresponding color images of two people from this collection. Collection G contains 415 subjects of which 302 subjects are from Collection F. The most important part of Collection G is that it has 24 subjects with images taken at four different viewpoints. We perform experiments only on these 24 subjects and not the *entire* Collection G because Collection G became available only very recently and contains the entire Collection F that we have already used in our experiments.

1.4 Major Ideas Presented in the Book

This book presents the following important ideas in ear biometrics, especially 3D ear biometrics.

- Develops automatic human recognition systems using 3D ear biometrics.
- Presents a two-step procedure for the accurate localization of an ear in a given image using a single reference ear shape model and fusion of color and range images. Adapts the reference shape model



(a)

(b)

Fig. 1.4. Examples of side face range images and the corresponding color images for six people in our dataset. (a) Range images. (b) Color images. Note the pose variations, the earrings and the hair occlusions for the six shots.

(a)

(b)

Fig. 1.5. Examples of side face range images and the corresponding color images for two people in the UND dataset Collection F. (a) Range images. (b) Color images. In (a) and (b) the left column shows the gallery images and the right column shows the probe images.

to incoming images by following a new global-to-local registration procedure with an associated optimization formulation. Incorporates the bending energy of thin plate spline into the optimization formulation as a regularization term to preserve the structure of the reference shape model under the deformation.

- Proposes two representations: the *helix/anti-helix based representation* and the *invariant local surface patch representation*, for recognizing ears in 3D. Evaluates these representations for their pose invariance and robustness to the real world data.
- Introduces a simple and effective surface matching scheme based on the modified iterative closest point (ICP) algorithm. Initializes ICP by a full 3D (translation and rotation) transformation.
- Introduces a novel computational framework for rapid recognition of 3D ears which combines the feature embedding and the support vector machine (SVM) technique to rank the hypotheses. No assumptions about the distributions of features is made and the performance of the proposed algorithm is scalable with the database size without sacrificing much accuracy.
- Presents a binomial model to predict ear recognition performance by modeling cumulative match characteristic (CMC) curve as a binomial distribution.
- Demonstrates the experimental results on the UCR dataset of 155 subjects with 902 images under pose variations and the University of Notre Dame dataset of 302 subjects with 604 images with time-lapse gallery-probe pairs. It also compares with the other published work [14–19]. In addition, it shows the results on 24 subjects (the UND dataset Collection G) taken from four different viewpoints.

1.5 Organization of the Book

The book consists of nine chapters.

In Chapter 2, we give a review of related work in the fields of ear biometrics and 3D object recognition.

In Chapter 3, we present three approaches for localizing the ear in side face range images: (1) template matching based–detection: the model template is represented by an averaged histogram of shape index. (2) shape model–based detection: we propose a shape model-based technique for locating human ears. The ear shape model is

represented by a set of discrete 3D vertices corresponding to ear helix and anti-helix parts. (3) global-to-local registration–based detection. We use a two-step approach for human ear detection by locating ear helix and anti-helix parts in registered 2D color and range images. Instead of training an active shape model to learn the shape variation, the ear shape model is adapted to images by following the global-to-local registration procedure.

In Chapter 4, we propose an ear helix/anti-helix representation for matching surfaces in 3D. The 3D coordinates of the ear helix/anti-helix are obtained from the detection algorithm described in Chapter 3. The correspondences of ear helix and anti-helix parts are used to obtain an initial transformation between every gallery-probe ear pair. This transformation is applied to randomly selected locations of gallery ears and a modified ICP algorithm is used to iteratively refine the transformation to bring the gallery ear and the probe ear into the best alignment. The root mean square (RMS) registration error is used as the matching error criterion. The subject in the gallery with the minimum RMS error is declared as the recognized person in the probe image.

In Chapter 5, we propose a novel surface descriptor, local surface patch (LSP) representation for recognizing human ears in 3D. A local surface descriptor is characterized by a centroid, a local surface type, and a 2D histogram. The 2D histogram shows the frequency of occurrence of shape index values versus the angle between the normal of reference feature point and that of its neighbors. Comparing local surface descriptors between a probe and a gallery image, an initial correspondence of local surface patches is established which is then filtered by geometric constraints. This generates a set of hypotheses for verification. The final verification is achieved by estimating transformations and aligning gallery images with the probe image based on the ICP algorithm.

In Chapter 6, we present a novel method for rapid recognition of 3D ears which combines the feature embedding for the fast retrieval of surface descriptors and the support vector machine (SVM) learning technique for ranking the hypotheses to generate a short list for the verification. The local surface patch (LSP) representation is used to find the correspondences between a gallery-probe pair. Due to the high dimensionality and the large number of LSP descriptors, the *FastMap* embedding algorithm maps the feature vectors to a low-dimensional

space where the distance relationships are preserved. A K-d tree algorithm works in the low-dimensional space for fast retrieval of similar descriptors. The initial correspondences are filtered and grouped to remove false correspondences using geometric constraints. Based on the correspondences, a set of features is computed as a measure of the similarity between the gallery-probe pair, and then the similarities are ranked using the support vector machines (SVM) rank learning algorithm to generate a short list of candidate models for verification.

In Chapter 7, we present a binomial model to predict ear recognition performance. Match and non-match distances are used to estimate the distributions. By modeling cumulative match characteristic (CMC) curve as a binomial distribution, the ear recognition performance is predicted on a larger gallery.

In Chapter 8, we apply the proposed techniques to the general 3D object recognition and non-rigid shape registration. The LSP representation proposed in Chapter 5 is general and can be used to recognize general 3D objects in range images. The global-to-local registration with an optimization formulation proposed in Chapter 3 is used to handle non-rigid 2D shape registration.

In Chapter 9, we present a summary of the book and the future work.