

# Human Ear Recognition in 3D

Hui Chen, *Student Member, IEEE*, and Bir Bhanu, *Fellow, IEEE*

**Abstract**—Human ear is a new class of relatively stable biometrics that has drawn researchers' attention recently. In this paper, we propose a complete human recognition system using 3D ear biometrics. The system consists of 3D ear detection, 3D ear identification, and 3D ear verification. For ear detection, we propose a new approach which uses a single reference 3D ear shape model and locates the ear helix and the antihelix parts in registered 2D color and 3D range images. For ear identification and verification using range images, two new representations are proposed. These include the ear helix/antihelix representation obtained from the detection algorithm and the local surface patch (LSP) representation computed at feature points. 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 angles between the normal of reference feature point and that of its neighbors. Both shape representations are used to estimate the initial rigid transformation between a gallery-probe pair. This transformation is applied to selected locations of ears in the gallery set and a modified Iterative Closest Point (ICP) algorithm is used to iteratively refine the transformation to bring the gallery ear and probe ear into the best alignment in the sense of the least root mean square error. The experimental results on the UCR data set of 155 subjects with 902 images under pose variations and the University of Notre Dame data set of 302 subjects with time-lapse gallery-probe pairs are presented to compare and demonstrate the effectiveness of the proposed algorithms and the system.

**Index Terms**—3D ear biometrics, 3D ear identification, 3D ear verification, range and color images, surface matching.

## 1 INTRODUCTION

BIOMETRICS deal with recognition of individuals based on their physiological or behavioral characteristics [1]. Researchers have done extensive studies on biometrics such as fingerprint, face, palm print, iris, and gait. Ear, a viable new class of biometrics, has certain advantages over face and fingerprint, which are the two most common biometrics in both academic research and industrial applications. For example, the ear is rich in features; it is a stable structure that does not change much with age [2] and it does not change its shape with facial expressions. Furthermore, ear is larger in size compared to fingerprints but smaller as compared to face and it can be easily captured from a distance without a fully cooperative subject although it can sometimes be hidden with hair, cap, turban, muffler, scarf, and earrings. The anatomical structure of the human ear is shown in Fig. 1. The ear is made up of standard features like the face. These include the outer rim (helix) and ridges (antihelix) parallel to the helix, the lobe, the concha (hollow part of ear), and the tragus (the small prominence of cartilage over the meatus). In this paper, we use the helix/antihelix for ear recognition.

Researchers have developed several biometrics techniques using the 2D intensity images [1, chapter 13], [3], [4], [5]. 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 pair. Fig. 2 shows an example of a range image and the registered color image acquired by the Minolta Vivid 300 camera. A range image is

relatively insensitive to illuminations and it 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 [6], [7], [8], [9], [10], [11], [12], [13]. More work on the 3D face biometrics can be found in [14], [15], [16], [17], [18], [19], [20].

In this paper, we propose a complete human recognition system using 3D ear biometrics. The system has two key components: 3D ear detection and 3D ear recognition. For ear detection, we propose a two-step approach using the registered 2D color and range images by locating the ear helix and the antihelix parts. In the first step, a skin color classifier is used to isolate the side face in an image by modeling the skin color and nonskin color distributions as a mixture of Gaussians [21]. The edges from the 2D color image are combined with the step edges from the range image to locate regions-of-interest (ROIs) which may contain an ear. In the second step, to locate the ear accurately, the reference 3D ear shape model, which is represented by a set of discrete 3D vertices on the ear helix and the antihelix parts, is adapted to individual ear images by following a new global-to-local registration procedure instead of training an active shape model [22] built from a large set of ears to learn the shape variation. The DARCES (data-aligned rigidity-constrained exhaustive search) algorithm [23], which can solve the 3D rigid registration problem efficiently and reliably, without any initial estimation, is used to perform the global registration. This is followed by the local deformation process where it is necessary to preserve the structure of the reference ear shape model since neighboring points cannot move independently under the deformation due to physical constraints. The bending energy of thin plate spline [24], a quantitative measure for nonrigid deformations, is incorporated into the proposed optimization formulation as a regularization term to preserve the topology of the ear shape model under the shape deformation. The optimization procedure drives the initial global registration toward the ear helix and the

• The authors are with the Center for Research in Intelligent Systems, University of California at Riverside, Riverside, CA 92521.  
E-mail: {hchen, bhanu}@cris.ucr.edu.

Manuscript received 2 Feb. 2006; revised 8 Aug. 2006; accepted 30 Oct. 2006; published online 18 Jan. 2007.

Recommended for acceptance by S. Prabhakar, J. Kittler, D. Maltoni, L. O'Gorman, and T. Tan

For information on obtaining reprints of this article, please send e-mail to: [tpami@computer.org](mailto:tpami@computer.org) and reference IEEECS Log Number TPAMISI-0116-0206. Digital Object Identifier no. 10.1109/TPAMI.2007.1005.

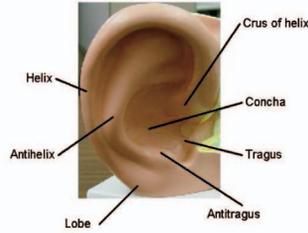


Fig. 1. The external ear and its anatomical parts.

antihelix parts, which results in the one-to-one correspondence of the ear helix and the antihelix between the reference ear shape model and the input image.

The approach for ear detection is followed to build a database of ears that belong to different people. For ear recognition, we present two representations: the ear helix/antihelix representation obtained from the detection algorithm and a new local surface patch representation computed at feature points to estimate the initial rigid transformation between a gallery-probe pair. For the ear helix/antihelix representation, the correspondence of ear helix and antihelix parts (available from the ear detection algorithm) between a gallery-probe ear pair is established and it is used to compute the initial rigid transformation. For the local surface patch (LSP) representation, a local surface descriptor is characterized by a centroid, a local surface type, and a 2D histogram. The local surface descriptors are computed for the feature points, which are defined as either the local minimum or the local maximum of shape indexes. By comparing the local surface patches for a gallery and a probe image, the potential corresponding local surface patches are established and then filtered by geometric constraints. Based on the filtered correspondences, the initial rigid transformation is estimated. Once this transformation is obtained using either of the two representations, it is then applied to randomly selected control points of the hypothesized gallery ear in the database. A modified Iterative Closest Point (ICP) algorithm is run to improve the transformation, which brings a gallery ear and a probe ear into the best alignment, for every gallery-probe pair. The root mean square (RMS) registration error is used as the matching error criterion. The subject in the gallery with the minimum RMS error is declared as the recognized person in the probe image.

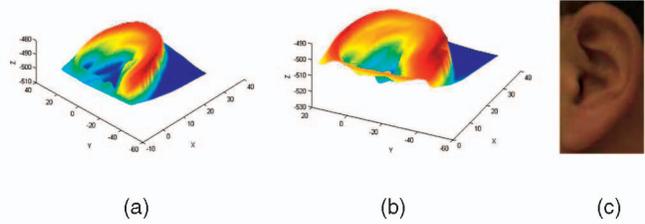


Fig. 2. Range image and color image captured by a Minolta Vivid 300 camera. In images (a) and (b), the range image of one ear is displayed as the shaded mesh from two viewpoints (the units of X, Y, and Z are mm). Image (c) shows the color image of the ear.

The rest of the paper is organized as follows: Section 2 introduces the related work and the contributions. Section 3 presents the ear recognition system. It presents the two-step approach to automatically detect ears using side face color and range images and describes the surface matching scheme to recognize ears using the ear helix/antihelix representation and the local surface patch representation. Section 4 develops a model to predict the performance of the ear recognition system in terms of cumulative match characteristic (CMC) curve. Section 5 gives the experimental results to demonstrate the effectiveness of the proposed approaches and the performance of the system. Finally, Section 6 provides the conclusions. Additional results are provided in the supplemental material [25].

## 2 RELATED WORK AND CONTRIBUTIONS

### 2.1 Related Work

Detection and recognition are the two major components of a biometrics system. Tables 1 and 2 provide a summary of object detection and recognition approaches for 3D biometrics. Table 3 compares our approach with Yan and Bowyer's work [8], [9], [10], [11], [12], [13] which is close to our research. As compared to all the work presented in Tables 1, 2, and 3, the detection approach proposed here uses a single reference ear shape model and exploits a two-step procedure that includes fusion of color and range data and the systematic adaptation of the shape model for local deformation. For the recognition, the most important difference between our approach and their work [8], [9], [10], [11], [12], [13] is that this paper proposes new ear helix/antihelix and

TABLE 1  
Object Detection: Summary of Approaches for 3D Biometrics

Authors	Technique	Biometrics
Tsalakanidou et al. [16]	Combines both depth and color data for localizing faces	Face
Boehnen and Russ[26]	Combination of the range and registered color images for automatically identifying facial features	Face
Lu et al. [27]	The shape index values used to identify the eye-points	Face
Kim et al. [28]	Combination of the range and color images for face detection	Face
Yan and Bowyer [13]	Active contour based segmentation (range and color images)	Ear
Chen and Bhanu [29]	Template matching based detection (range image only)	Ear
Chen and Bhanu [30]	Ear shape model-based registration (range image only)	Ear
This paper	Reference ear shape model and the global-to-local shape registration (range and color images)	Ear

TABLE 2  
Object Recognition: Summary of Approaches for 3D Biometrics

Authors	Technique	Biometrics
Lee and Milios [14]	Similarity computed by correlating extended Gaussian image (EGI)	Face
Gordon [15]	Face descriptors, using geometric features, calculated on salient surfaces and search the nearest neighbor in the scaled feature space	Face
Tsalakanidou et al. [16]	Embedded hidden Markov model applied to depth and color data	Face
Chua et al. [17]	Point signature of the 3D range data for the registration of two face surfaces with different facial expressions	Face
Bronstein et al. [18]	A representation of the facial surface invariant to isometric deformation	Face
Chang et al. [19]	Multimodal 2D+3D face recognition using principal component analysis	Face
Lu and Jain[20]	Integration of 3D face models with 2D appearance images	Face
Bhanu and Chen [6]	A local surface patch (LSP) descriptor for surface representation	Ear
Chen and Bhanu [7]	Two-step ICP algorithm for matching	Ear
Yan and Bowyer [8–13]	3D principal component analysis, 3D edge-based and ICP-based registration techniques	Ear
This paper	The ear helix/anti-helix and the local surface patch representations for the coarse alignment and the modified ICP algorithm for refinement	Ear

TABLE 3  
Comparison between Chen and Bhanu's and Yan and Bowyer's Approaches in Methodology

		Chen and Bhanu's approach (This paper)	Yan and Bowyer's approach [8–13]
Methodology	Detection	Fusion of range and color images, global-to-local registration for automatic ear extraction on all the datasets	Manual extraction on the UND dataset Collection F [9–11] and active contour method for automatic ear extraction on Collection G[13]
	Recognition	Ear helix/anti-helix and LSP representations for estimating the initial rotation and translation between a gallery-probe pair, then the modified ICP algorithm for refining the transformation	3D PCA, 3D edge-based and ICP-based approaches [8–13]. For ICP algorithm, only the translation between a gallery-probe pair is initialized.

the local surface patch (LSP) representations to estimate the initial rotation and translation between a gallery-probe pair. The initial transformation is critical for the success of the ICP algorithm and more details about it are given in Section 5.3.4.

## 2.2 Contributions of This Paper

The specific contributions of this paper are:

1. An automatic human recognition system using the 3D ear biometrics is developed.
2. A single reference ear shape model is adapted to incoming images by following a new global-to-local registration procedure with an associated optimization formulation. The bending energy of thin plate spline is incorporated into the optimization formulation as a regularization term to preserve the structure of the reference shape model under the deformation. This formulation can be applied to handle the nonrigid shape registration and it is demonstrated for the alignment of the hand and dude shapes.
3. The *helix/antihelix-based representation* and the *invariant local surface patch representation* for recognizing ears in 3D are proposed. These representations are evaluated for their pose invariance and robustness to the real world data.

4. A simple and effective surface matching scheme based on the modified ICP algorithm is introduced. ICP is initialized by a full 3D (translation and rotation) transformation.
5. A binomial model is presented for characterizing the performance of ear recognition.
6. The experimental results on two large ear databases (UCR and UND databases) of color and range images are presented and compared with the other published work [8], [9], [10], [11], [12], [13].

## 3 TECHNICAL APPROACH

The proposed human recognition system using 3D ear biometrics is illustrated in Fig. 3. It consists of two components: 3D ear detection and 3D ear recognition. By following the ear detection algorithm, ears are *automatically* extracted in side face range images. For the ear helix/antihelix representation, the ear gallery consists of 3D range data associated with the 3D coordinates of ear helix and antihelix parts obtained in detection. For the local surface patch (LSP) representation, the ear gallery consists of a set of descriptors associated with 3D range data. They are computed at selected feature points marked by plus signs in Fig. 3. In the following, we describe the details of the proposed ear detection and recognition algorithms and the system.

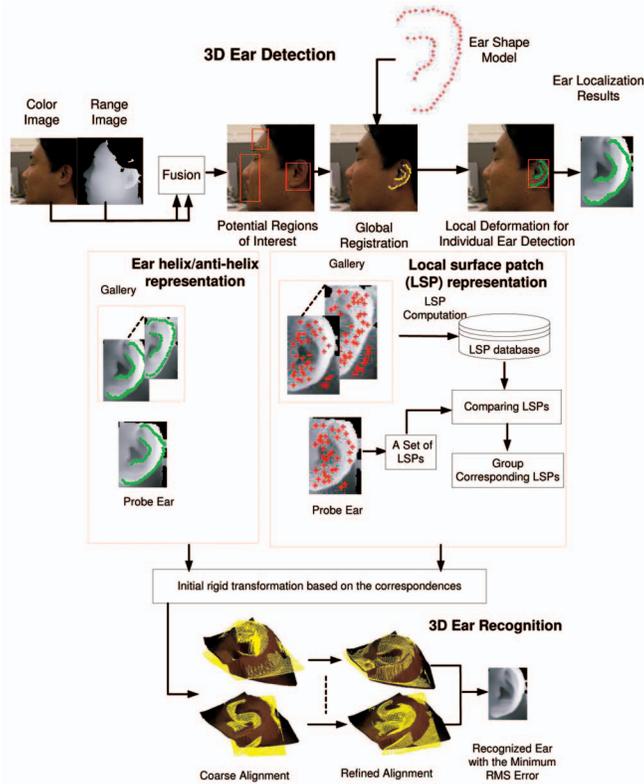


Fig. 3. The proposed human recognition system using 3D ear biometrics. The top part of the figure shows the ear detection module and the bottom shows the ear recognition module using the ear helix/antihelix and the local surface patch (LSP) representations.

### 3.1 Automatic Ear Extraction

The Minolta Vivid 300 and 910 range sensors used in this work provide a registered 3D range image and a 2D color image. The ear detection starts with the extraction of regions-of-interest (ROIs) using both the range and color images. Once the ROIs are located, the problem of identifying the ear helix and antihelix parts in a range image is converted to the alignment of the reference ear shape model with ROIs. The alignment follows a new global-to-local procedure: The global registration brings the reference shape model into coarse alignment with the ear helix and the antihelix parts; the local deformation driven by the optimization formulation drives the shape model more close to the ear helix and the antihelix parts.

#### 3.1.1 Regions-of-Interest (ROIs) Extraction

Since the images in two modalities (range and color) are registered, the ROIs can be localized in any one modality if they are known in the other modality.

- **Processing Color Images.** The processing consists of two major tasks:
  - **Skin Color Classification.** Skin color is a powerful cue for segmenting the exposed parts of the human body. Jones and Rehg [21] built a classifier by learning the distributions of skin and nonskin pixels from a data set of nearly one billion labeled pixels. The distributions are modeled as a mixture of Gaussians and their parameters are given in [21]. We use this method for finding skin regions.

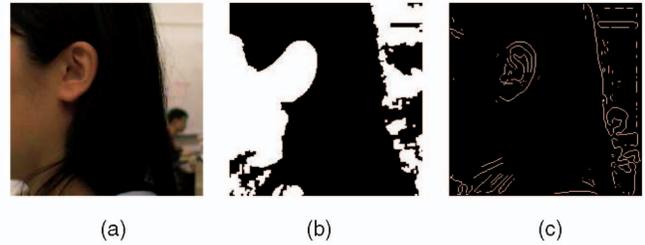


Fig. 4. One example of processing the color image. (a) Color image. (b) Skin color map. (c) Edge detection using a LOG edge detector.

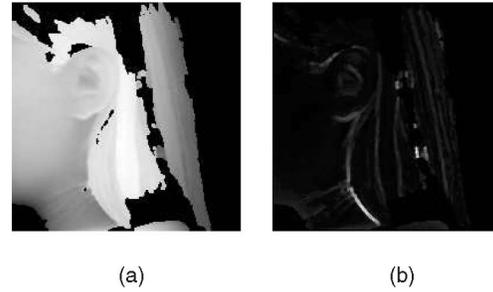


Fig. 5. One example of processing the range image. (a) Range image. (b) Step edge magnitude image. In image (a), the darker pixels are away from the camera and the lighter ones are closer. In image (b), the bright pixels denote large edge magnitude.

When a pixel  $p(R, G, B)$  is presented for the classification, we compute a posteriori probability  $P(\text{skin}|RGB)$  and  $P(\text{nonskin}|RGB)$  and make the classification using the Bayesian decision theory. Fig. 4a shows a color image and Fig. 4b shows the pixel classification result in which the skin pixels are shown as white. We observe that the large skin region containing the ear is roughly segmented.

- **Edge Extraction in Intensity Images:** There are edges, around the ear helix and antihelix parts, caused by a change in intensity. These are helpful for locating the ear region. The edges are extracted from 2D intensity images. The  $(R, G, B)$  color images are first converted to the gray-scale images (eliminating the hue and saturation information while retaining the luminance) and then edges are extracted by using the Laplacian of Gaussian (LOG) edge detector ( $13 \times 13$  window is used). Fig. 4c shows the edge detection result using the LOG detector.
- **Processing Range Images.** It can be clearly seen from Fig. 4 and Fig. 5 that there is a sharp change in depth around the ear helix part, which is helpful in identifying the ear region. Given a side face range image, the step edge magnitude, denoted by  $I_{step}$ , is calculated.  $I_{step}$  is defined by the maximum distance in depth between the center pixel and its neighbors in a  $w \times w$  window ( $w = 3$ ).  $I_{step}$  can be written as:

$$I_{step}(i, j) = \max_{\substack{- (w-1)/2 \leq k, l \leq (w-1)/2}} |z(i, j) - z(i+k, j+l)|, \quad (1)$$

where  $w$  is the width of the window and  $z(i, j)$  is the  $z$  coordinate of the point  $(i, j)$ . Fig. 5a shows a range

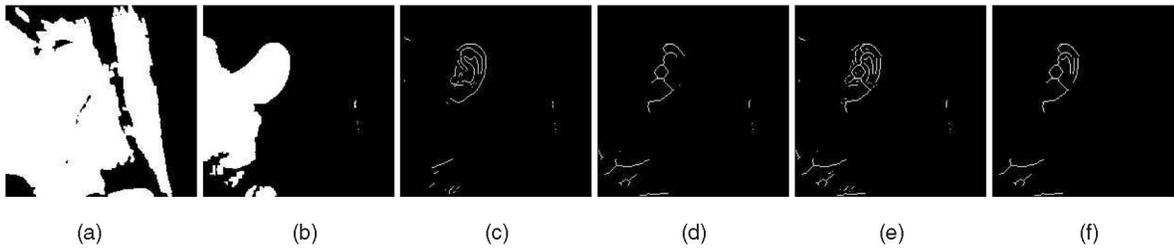


Fig. 6. Fusion of 2D color and 3D range images. (a) The range mask. (b) The final mask obtained by a combination of the range mask and the skin color map. (c) Edges in the intensity image after applying the final mask. (d) Edges in the range image after applying the final mask. (e) Combination of edges in both color and range images. (f) Edges after removal of small edge segments.

image in which the darker pixels are far away from the camera; Fig. 5b shows the step edge magnitude image in which the pixels with larger magnitudes are displayed as brighter pixels. We observe that the edge magnitude is large around the ear helix and the antihelix parts.

- **Fusion of Color and Range Images.** It works as follows:

- The range sensor provides a range mask indicating *valid pixels* (in white), which is shown in Fig. 6a.
- The range mask is combined with the skin color map to generate a final mask indicating the *valid skin pixels*, which is shown in Fig. 6b.
- The final mask is applied to edge pixels from the intensity image to remove some of the pixels which are *nonskin pixels* or *invalid pixels*. “Nonskin pixels” mean the pixels that are not on the skin and “invalid pixels” mean that the range sensor did not make measurements for these pixels. The edge pixels that are left over are shown in Fig. 6c.
- For the range image, the final mask is also applied to the step edge magnitude image. In order to get edges in the range image, the step edge magnitude image is thresholded. The selection of the threshold value is based on the cumulative histogram of the step edge magnitude image. Since we are interested in larger magnitudes, the top  $\eta$  percent ( $\eta = 3.5$ ) pixels with the largest magnitudes are selected as the edge pixels. The thresholded binary image is then dilated (using a  $3 \times 3$  square structuring element) and thinned (shrinking to a minimally connected stroke). The edges so obtained are shown in Fig. 6d.
- The edges from intensity image and range images are combined in the following manner: The final edge map that we expect to obtain is initialized to be the edge map of the range image (Fig. 6d); for each edge pixel in the intensity image (Fig. 6c), if none of its neighbors are edge pixels in the range image, then this edge pixel is added to the final edge map. An example of the final edge map is shown in Fig. 6e.
- The edge pixels are labeled by the connected component labeling algorithm and the small edge segments are removed (less than 10 pixels

in our experiments). The final left over edge segments are shown in Fig. 6f.

- **Clustering Edge Segments.** After edge segments are extracted, those close to each other are grouped into clusters. Each cluster is a region-of-interest. The clustering procedure works as follows:
  - while the number of edge segments  $> 0$ :
  - $i = 0$ .
  - Put the first edge segment  $e_i$  into a cluster  $C_i$ , calculate its centroid  $\{\mu_{xi}, \mu_{yi}\}$ .
  - For all the other edge segments  $e_j$ :
    - Calculate the centroid  $\{\mu_{xj}, \mu_{yj}\}$ .
    - If  $\max\{|\mu_{xj} - \mu_{xi}|, |\mu_{yj} - \mu_{yi}|\} \leq \epsilon$ , put  $e_j$  into the cluster  $C_i$ , remove  $e_j$  and update the cluster’s centroid.
  - $i = i + 1$  and relabel the edge segments.

In our experiments,  $\epsilon = 36$  pixels. The clustering examples are provided in [25].

### 3.1.2 Reference Ear Shape Model

In this paper, instead of training an active shape model to learn the shape variation, we adapt the reference ear shape model to input images by following a global-to-local procedure, described below, in which the topology of the ear shape model is preserved during the shape deformation. We build the reference ear shape model from an instance of an ear belonging to a person. The reference ear shape model  $s$  is defined by 3D coordinates  $\{x, y, z\}$  of  $n$  vertices which lie on the ear helix and the antihelix parts. The ear helix and the antihelix parts are manually marked for the reference shape model. The shape model  $s$  is represented by a  $3n \times 1$  vector  $(x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_n, y_n, z_n)^T$ . Fig. 7a shows the ear shape model  $s$  marked by the pluses (+). The corresponding color image is also shown in Fig. 7b.

### 3.1.3 Alignment of the Reference Ear Shape Model with ROI

Once an ROI is extracted, the ear helix and the antihelix parts are identified by the alignment of ROI with the ear shape model. Since the rigid registration cannot account for the local shape variation between ears, we develop a global-to-local procedure: The global registration brings the reference ear shape model into coarse alignment with the ear helix and the antihelix parts; the local deformation driven by the optimization formulation (given below) drives the reference ear shape model more close to the ear helix and the antihelix parts.

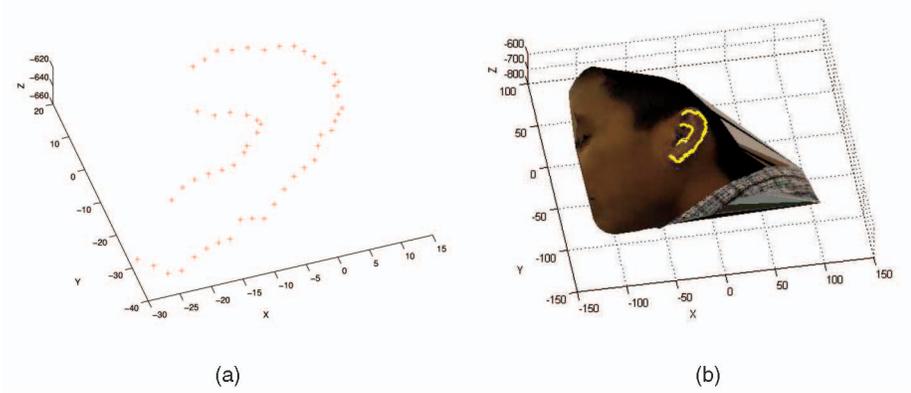


Fig. 7. The reference ear shape model. (a) The reference 3D ear shape model is displayed by the pluses (+). (b) The ear shape model is overlaid on the textured 3D face. The units of X, Y, and Z are mm.

- **Global Rigid Registration.** For the 3D registration problem, the Iterative Closest Point (ICP) algorithm [31] is widely used for matching points with unknown corresponding pairs. Although there are many variants of the ICP algorithm [32], [33], [34], basically it consists of two iterative steps: 1) identifying correspondences by finding the closest points and 2) computing the rigid transformation based on the corresponding pairs. The major drawback of an ICP-based algorithm is that it needs a good initial guess of the true transformation.

The RANSAC-based data-aligned rigidity-constrained exhaustive search algorithm (DARCES) [23] can solve the registration problem without any initial estimation by using rigidity constraints to find the corresponding points. First, three points (primary, secondary, and auxiliary) in the reference surface are selected; then, each point on the test surface is assumed to be in correspondence to the primary point and the other two corresponding points are found based on the rigidity constraints. For every corresponding triangle, a rigid transformation is computed and the transformation with the maximum number of overlapping points is chosen as the solution. Due to the exhaustive nature of the search, the solution it finds is the true one.

In our case, the 3D coordinates of the reference ear shape model are known. We use the DARCES algorithm to find the corresponding triangles (between the reference ear shape model and the ROI under consideration) and the initial transformation. The ICP algorithm is then used to refine the transformation. This process is repeated for each ROI and the ROI with the minimum registration error is passed to the local deformation stage.

- **Local Deformation.**

- **Thin Plate Spline Transformation.** The reference shape model (the ear helix and the antihelix parts) is deformed after it is globally aligned with an ROI. Thin plate spline (TPS) transformation is a powerful tool for modeling the shape deformation and is widely used in shape matching [24], [35], [36], [37]. The TPS  $\mathcal{R}^2 \rightarrow \mathcal{R}^2$  mapping function is defined by the following equation:

$$\begin{aligned} \mathbf{v} &= f(\mathbf{u}) = \begin{bmatrix} f^x(\mathbf{u}) \\ f^y(\mathbf{u}) \end{bmatrix} \\ &= A\mathbf{u} + \mathbf{t} + \sum_{i=1}^n \begin{bmatrix} w_i^x \\ w_i^y \end{bmatrix} \phi(|\mathbf{u} - \mathbf{u}_i|), \end{aligned} \quad (2)$$

where  $\phi(r) = r^2 \log r$ ,  $\mathbf{u} = [\hat{x}, \hat{y}]^T$ ,  $\mathbf{v} = [x, y]^T$ , and  $A$  and  $\mathbf{t}$  form an affine transformation given by

$$[A \quad \mathbf{t}] = \begin{bmatrix} a_{00} & a_{01} & t_0 \\ a_{10} & a_{11} & t_1 \end{bmatrix}.$$

The  $n \times 2$  matrix

$$W = \begin{bmatrix} w_1^x & w_1^y & \cdots & w_n^x \\ w_1^y & w_2^y & \cdots & w_n^y \end{bmatrix}^T$$

specifies the nonlinear warping where  $n$  is the number of landmark points. Given  $n$  landmark points  $\mathbf{u}(\hat{x}_i, \hat{y}_i)$  and their corresponding points  $\mathbf{v}(x_i, y_i)$ , (2) can be rewritten as  $2n$  linear equations. However, there are  $2n + 6$  unknown parameters to be solved. The following six constraints are added to make the spline function (2) have the square integrable second derivatives:

$$P^T [w_1^x, w_2^x, \dots, w_n^x]^T = 0, P^T [w_1^y, w_2^y, \dots, w_n^y]^T = 0, \quad (3)$$

where  $P$  is an  $n \times 3$  matrix which is defined by  $(\mathbf{1}, \hat{\mathbf{x}}, \hat{\mathbf{y}})$ ,  $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n)^T$ , and  $\hat{\mathbf{y}} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)^T$ . The  $2n + 6$  equations can be put into a compact matrix form:

$$\begin{bmatrix} \Phi & P \\ P^T & 0 \end{bmatrix} \begin{bmatrix} W \\ \mathbf{t}^T \\ A^T \end{bmatrix} = \begin{bmatrix} \mathbf{v} \\ \mathbf{0} \end{bmatrix}, \quad (4)$$

where the  $n \times n$  matrix  $\Phi_{ij} = \phi(\mathbf{u}_i - \mathbf{u}_j)$ ,  $\mathbf{v} = (\mathbf{x}, \mathbf{y})$ ,  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ , and  $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ . The TPS transformation minimizes the following bending energy function:

$$B_e = \int \int_{\mathcal{R}^2} (F(f^x) + F(f^y)) dx dy, \quad (5)$$

where  $F(g^*(x, y)) = (g_{xx}^2 + 2g_{xy}^2 + g_{yy}^2)^*$ ,  $*$  denotes the  $(x$  or  $y)$  under consideration and  $g_{xx}$ ,  $g_{xy}$ , and

$g_{yy}$  are second order derivatives. It can be shown that the value of bending energy is  $B_e = \frac{1}{8\pi} (\mathbf{x}^T K \mathbf{x} + \mathbf{y}^T K \mathbf{y})$ , where  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$  and  $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$  [24]. The matrix  $K$  is the  $n \times n$  upper left matrix of

$$\begin{bmatrix} \Phi & P \\ P^T & 0 \end{bmatrix}^{-1},$$

which only depends on the coordinates of the landmark points in  $\{\mathbf{u}\}$ . Therefore, the bending energy is determined by the coordinates of landmark points and their correspondences. Furthermore, the bending energy is a good measurement of the shape deformation. Since the coordinates of the reference ear shape model are known, the matrix  $K$  can be precomputed. The task is to drive the reference ear shape model toward the ROI ear such that the topology of the reference ear shape model is preserved. The bending energy is used to penalize the large shape deformation.

**Optimization Formulation.** In Section 3.1.1, we noted that there are strong step edge magnitudes in range images around the ear helix and the antihelix parts. After we bring the reference shape model into coarse alignment with the ear helix and the antihelix parts (in the ROI image) through the global rigid registration, we get the locations of the 3D coordinates of ear helix and antihelix parts in the 2D color image and perform the local deformation on the 2D image plane since the 2D color image is registered with the 3D range image. In other words, we would like to drive the reference ear shape model more close to the ear helix and the antihelix parts with the topology of the shape model preserved. We can achieve this task by minimizing the proposed new cost function:

$$\begin{aligned} E(\mathbf{x}, \mathbf{y}) &= E_{img}(\mathbf{x}, \mathbf{y}) + \gamma E_D(\mathbf{x}, \mathbf{y}) \\ &= \sum_{i=1}^n h(|\nabla I_{step}(x_i, y_i)|) \\ &\quad + \frac{1}{2} \gamma (\mathbf{x}^T K \mathbf{x} + \mathbf{y}^T K \mathbf{y}), \end{aligned} \quad (6)$$

where  $h(|\nabla I_{step}|) = 1/(1 + |\nabla I_{step}|)$ ,  $|\nabla I_{step}(x_i, y_i)|$  is the step edge magnitude of  $i$ th point of the shape model located in the 2D plane and  $\gamma$  is a positive regularization constant that controls the topology of the shape model. For example, increasing the magnitude of  $\gamma$  tends to keep the topology of the ear shape model unchanged. In (6), the step edge magnitude in range images is used for the term  $E_{img}$  since edges in range images are less sensitive to the change of viewpoint and illumination than those in color images. In (6), the first term  $E_{img}$  drives points  $(\mathbf{x}, \mathbf{y})$  toward the ear helix and the antihelix parts which have larger step edge magnitudes; the second term  $E_D$  is the bending energy that preserves the topology of the reference shape model under the shape deformation.

When we take the partial derivatives of (6) with respect to  $\mathbf{x}$  and  $\mathbf{y}$  and set them to zero, we have

$$\begin{aligned} \gamma K \mathbf{x} - \sum_{i=1}^n \frac{1}{(1 + |\nabla I_{step}(x_i, y_i)|)^2} \frac{\partial |\nabla I_{step}(x_i, y_i)|}{\partial \mathbf{x}} &= 0, \\ \gamma K \mathbf{y} - \sum_{i=1}^n \frac{1}{(1 + |\nabla I_{step}(x_i, y_i)|)^2} \frac{\partial |\nabla I_{step}(x_i, y_i)|}{\partial \mathbf{y}} &= 0. \end{aligned} \quad (7)$$

Since  $K$  is positive semidefinite, (7) can be solved *iteratively* by introducing a step size parameter  $\alpha$ , which is shown in (8) [38].

$$\begin{aligned} \gamma K \mathbf{x}_t + \alpha (\mathbf{x}_t - \mathbf{x}_{t-1}) - \mathcal{F}_{t-1}^{\mathbf{x}} &= 0, \\ \gamma K \mathbf{y}_t + \alpha (\mathbf{y}_t - \mathbf{y}_{t-1}) - \mathcal{F}_{t-1}^{\mathbf{y}} &= 0. \end{aligned} \quad (8)$$

The solutions can be obtained by matrix inversion, which is shown in (9), where  $I$  is the identity matrix. In (8) and (9),

$$\mathcal{F}_{t-1}^{\mathbf{x}} = \sum_{i=1}^n \frac{1}{(1 + |\nabla I_{step}^{t-1}(x_i, y_i)|)^2} \frac{\partial |\nabla I_{step}^{t-1}(x_i, y_i)|}{\partial \mathbf{x}}$$

and

$$\mathcal{F}_{t-1}^{\mathbf{y}} = \sum_{i=1}^n \frac{1}{(1 + |\nabla I_{step}^{t-1}(x_i, y_i)|)^2} \frac{\partial |\nabla I_{step}^{t-1}(x_i, y_i)|}{\partial \mathbf{y}},$$

and they are evaluated for all the coordinates  $(x_i, y_i)$  at the iteration  $t - 1$ .  $|\nabla I_{step}^{t-1}(x_i, y_i)|$  is the step edge magnitude at the location  $(x_i, y_i)$  at the iteration  $t - 1$ .

$$\begin{aligned} \mathbf{x}_t &= (\gamma K + \alpha I)^{-1} (\alpha \mathbf{x}_{t-1} + \mathcal{F}_{t-1}^{\mathbf{x}}), \\ \mathbf{y}_t &= (\gamma K + \alpha I)^{-1} (\alpha \mathbf{y}_{t-1} + \mathcal{F}_{t-1}^{\mathbf{y}}). \end{aligned} \quad (9)$$

We have used  $\alpha = 0.5$  and  $\gamma = 100$  in our experiments on ear detection described in Section 5.

## 3.2 Surface Matching for Ear Recognition Using the Ear Helix/Antihelix Representation

As shown in the ear recognition part of Fig. 3, the surface matching follows the coarse-to-fine strategy. Given a set of probe images, the ear and its helix and the antihelix parts are extracted by running the detection algorithm described above. The correspondence of the helix and the antihelix between the probe ear and the hypothesized gallery ear is used to compute the initial transformation that brings the hypothesized gallery ear into coarse alignment with the probe ear and then a modified ICP algorithm is run to refine the transformation to bring gallery-probe pairs into the best alignment.

### 3.2.1 Coarse Alignment

Given two corresponding sets of  $n$  3D vertices  $M$  and  $S$  on the helix and the antihelix parts, the initial rigid transformation,

which brings the gallery and the probe ears into coarse alignment, can be estimated by minimizing the sum of the squares of these errors ( $\Sigma = \frac{1}{n} \sum_{l=1}^n |S_l - R * M_l - T|^2$ ) with respect to the rotation matrix  $R$  and the translation vector  $T$ . The rotation matrix and translation vector are computed by using the quaternion representation [39].

### 3.2.2 Fine Alignment

Given the estimate of initial rigid transformation, the purpose of Iterative Closest Point (ICP) algorithm [31] is to determine if the match is good and to find a refined alignment between them. If the probe ear is really an instance of the gallery ear, the ICP algorithm will result in good registration and a large number of corresponding points between gallery and probe ear surfaces will be found. Since ICP algorithm requires that the probe be a subset of the gallery, a method to remove outliers based on the distance distribution is used [33]. The basic steps of the modified ICP algorithm are summarized below:

- **Input:** A 3D gallery ear range image, a 3D probe ear range image, and the initial transformation obtained from the coarse alignment.
- **Output:** The refined transformation between the two ears.
- **Procedure:**
  - a. Select control points ( $\sim 180$ ) in the gallery ear range image randomly and apply the initial transformation to the gallery ear image.
  - b. Find the closest points of the control points in the probe ear image and compute the statistics [33] of the distances between the corresponding pairs in the gallery and the probe images.
  - c. Discard some of the corresponding pairs by analyzing the statistics of the distances (a threshold is obtained based on the mean and standard deviation of distances) [33].
  - d. Compute the rigid transformation between the gallery and the probe ears based on the correspondences.
  - e. Apply the transformation to the gallery ear range image and repeat from Step b until convergence.

Starting with the initial transformation obtained from the coarse alignment, the modified ICP algorithm is run to refine the transformation by minimizing the distance between the control points of the gallery ear and their closest points of the probe ear. For each gallery ear in the database, the control points are randomly selected and the modified ICP is applied to those points. For a selected gallery ear, we repeat the same procedure 15 times and choose the rigid transformation with the minimum root mean square (RMS) error. The subject in the gallery set with the minimum RMS error is declared as the recognized person. In the modified ICP algorithm, the speed bottleneck is the nearest neighbor search. Therefore, the kd-tree structure is used in the implementation. Fig. 8a shows the coarse alignment after applying the initial rigid transformation; Fig. 8b shows the refined alignment after applying the modified ICP algorithm. In Fig. 8, the gallery ear represented by the mesh is overlaid on the textured 3D probe ear. We observe a better alignment after applying the modified ICP algorithm.

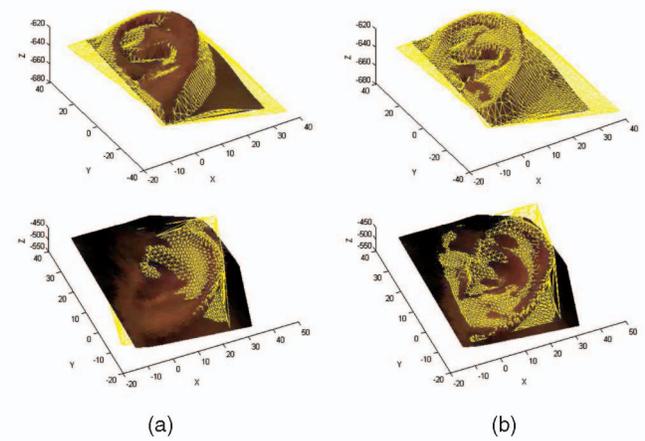


Fig. 8. Two examples of coarse and fine alignment. The gallery ear represented by the mesh is overlaid on the textured 3D probe ear. (a) Coarse alignment. (b) Fine alignment.

## 3.3 Surface Matching for Ear Recognition Using Local Surface Patch (LSP) Representation

In 3D object recognition, the key problems are how to represent free-form surfaces effectively and how to match the surfaces using the selected representation. Researchers have proposed various surface signatures for recognizing 3D free-form objects which are reviewed in [6], [40]. In the following, we present a new surface representation, called the local surface patch (LSP), investigate its properties, and use it for ear recognition.

### 3.3.1 Local Surface Patch Representation (LSP)

We define a “local surface patch” as the region consisting of a feature point  $P$  and its neighbors  $N$ . The LSP representation includes feature point  $P$ , its surface type, centroid of the patch, and a histogram of shape index values versus dot product of the surface normal at point  $P$  and its neighbors. A local surface patch is shown in Fig. 10. The neighbors satisfy the following conditions,  $N = \{pixels \ N, ||N - P|| \leq \epsilon_1\}$  and  $acos(n_p \bullet n_n) < A$ , where  $\bullet$  denotes the dot product between the surface normal vectors  $n_p$  and  $n_n$  at point  $P$  and  $N$  and  $acos$  denotes the inverse cosine function. The two parameters  $\epsilon_1$  and  $A$  ( $\epsilon_1 = 5.8\text{mm}$ ,  $A = 0.5$ ) are important since they determine the descriptiveness of the local surface patch representation. A local surface patch is not computed at every pixel in a range image, but only at selected feature points. The feature points are defined as the local minimum and the maximum of shape indexes, which can be calculated from principal curvatures. In order to estimate curvatures, we fit a biquadratic surface  $f(x, y) = ax^2 + by^2 + cxy + dx + ey + f$  to a local window and use the least square method to estimate the parameters of the quadratic surface and then use differential geometry to calculate the surface normal, Gaussian, and mean curvatures and the principal curvatures [6], [41].

Shape index ( $S_i$ ), a quantitative measure of the shape of a surface at a point  $P$ , is defined by

$$S_i(P) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \frac{k_1(P) + k_2(P)}{k_1(P) - k_2(P)},$$

where  $k_1$  and  $k_2$  are maximum and minimum principal curvatures, respectively. With this definition, all shapes are

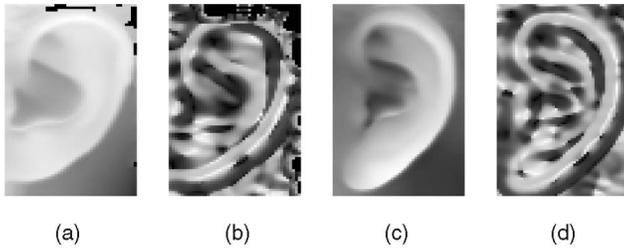


Fig. 9. Two examples of ear range images ((a), (c)) and their corresponding shape index images ((b), (d)). In images (a) and (c), the darker pixels are away from the camera and the lighter ones are closer. In images (b) and (d), the darker pixels correspond to concave surfaces and lighter ones correspond to convex surfaces.

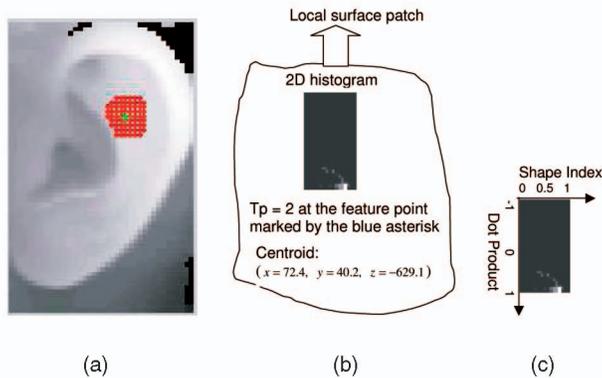


Fig. 10. Illustration of a local surface patch (LSP). (a) Feature point  $P$  is marked by the asterisk and its neighbors  $N$  are marked by the interconnected dots. (b) LSP representation includes a 2D histogram, a surface type, and centroid coordinates. (c) The 2D histogram is shown as a gray image in which the brighter areas correspond to bins a the high frequency of occurrence.

mapped into the interval  $[0, 1]$  [42]. The shape categories and corresponding shape index range are as follows: spherical cup  $[0, 1/16)$ , trough  $[1/16, 3/16)$ , rut  $[3/16, 5/16)$ , saddle rut  $[5/16, 7/16)$ , saddle  $[7/16, 9/16)$ , saddle ridge  $[9/16, 11/16)$ , ridge  $[11/16, 13/16)$ , dome  $[13/16, 15/16)$ , spherical cap  $[15/16, 1]$  [42], [43]. Fig. 9 shows original ear range images and their shape index images for two people. In this figure, the brighter pixels denote large shape index values which correspond to ridge and dome surfaces while the darker pixels denote small shape index values which correspond to valley and cup surfaces. Within a  $b \times b$  ( $b = 5$ ) window, the center point is marked as a feature point if its shape index is higher or lower than those of its neighbors. The results of feature points extraction are shown in Fig. 11, where the feature points are marked by red plus sign. In order to see the feature points' location, we enlarge the two images. We can clearly see that some feature points corresponding to the same physical area appear in both images.

For every local surface patch, we compute the shape indexes and normal angles between point  $P$  and its neighbors. Then, we form a 2D histogram by accumulating points in particular bins along the two axes. One axis of this histogram is the shape index, which is in the range  $[0, 1]$ ; the other is the dot product of surface normal vectors at  $P$  and  $N$ , which is in the range  $[-1, 1]$ . In order to reduce the effect of noise, we use bilinear interpolation when we calculate the 2D histogram. One example of 2D histogram is shown as a gray scale image in Fig. 10c; the brighter areas in the image

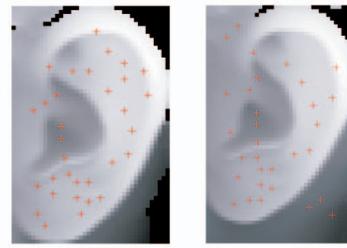


Fig. 11. Feature points location (+) in two range images shown as gray scale images of the same ear taken at different viewpoints.

correspond to bins with more points falling into them. In the implementation, the number of bins in the shape index axis is 17 and the number of bins in the other axis is 34.

We classify the surface shape of a local surface patch into three types: concave ( $T_p = 0$ ), saddle ( $T_p = 1$ ), and convex ( $T_p = 2$ ), based on the shape index value of the feature point. The shape index range and its corresponding surface type are as follows:  $[T_p = 0, S_i \in [0, 5/16)]$ ,  $[T_p = 1, S_i \in [5/16, 11/16)]$ , and  $[T_p = 2, S_i \in [11/16, 1)]$ . We also compute the centroid of a local surface patch. Note that a feature point and the centroid of a patch may not coincide.

In summary, every local surface patch is described by a 2D histogram, every local surface patch is described by a 2D histogram, surface type, and the centroid. The 2D histogram and surface type are used for comparison of LSPs and the centroid is used for computing the rigid transformation. The patch encodes the geometric information of a local surface.

It is to be noted that our LSP representation is different from the spin image representation [44]. Unlike the LSP representation, the spin image is a 2D histogram described by two parameters: the distance to the tangent plane of the oriented point from its neighbors and the distance to the normal vector of the oriented point. As described above, we compute LSPs for feature points, while the spin image is computed for every vertex on the surface of an object [44]. In Section 5, we provide a comparison of the LSP and the spin image representations on an ear data set.

### 3.3.2 Comparing Local Surface Patches

Given a probe range image, we extract feature points and get local surface patches. Considering the inaccuracy of feature points' location, we also extract local surface patches from the neighbors of feature points. Then, we compare them with all of the local surface patches saved in the gallery based on the surface type and histogram dissimilarity. We use a statistical method to assess the dissimilarity between the two probability density functions since a histogram can be thought of as an unnormalized approximation to it. The  $\chi^2$ -divergence is among the most prominent divergence used in statistics to assess the dissimilarity between two probability density functions. We use it to measure the dissimilarity between two observed histograms  $Q$  and  $V$ , as follows [45]:

$$\chi^2(Q, V) = \sum_i \frac{(q_i - v_i)^2}{q_i + v_i}. \quad (10)$$

From (10), we know the dissimilarity is between 0 and 2. If the two histograms are exactly the same, the dissimilarity

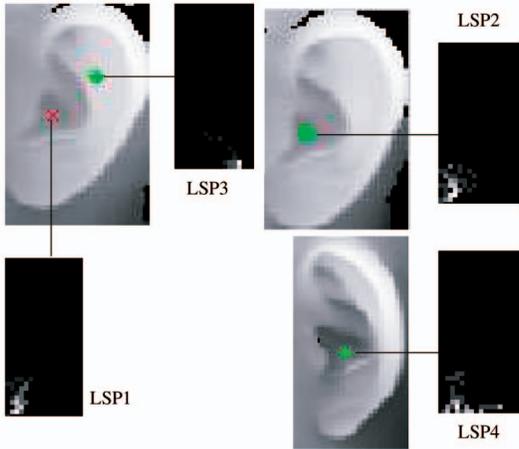


Fig. 12. The demonstration of the discriminatory power of local surface patches. The ear images in the first row are from the same person but with different viewpoints. The ear image shown in the second row is from a different person. The histograms of four local surface patches (LSP1 to LSP4) are also shown for the comparison.

will be zero. If the two histograms do not overlap with each other, it will achieve the maximum value of 2.

Fig. 12 shows an experimental validation that the local surface patch has the discriminative power to distinguish shapes. We do experiments under three cases: 1) a local surface patch (LSP1) generated for an ear is compared to another local surface patch (LSP2) corresponding to the same physical area of the same ear imaged from a different viewpoint; in this case, a low dissimilarity exists and both LSPs have the same surface type. 2) The LSP1 is compared to LSP3, which lies in a different area of the same ear; the dissimilarity is high and they have the different surface type. 3) The LSP1 is compared to LSP4, which lies in a similar area as the LSP1 but it is not the same ear; there exists a higher dissimilarity from the first case and they also

have different surface types. These experimental results suggest that the local surface patch provides distinguishable features and it can be used for differentiation between ears. Table 4 shows the comparison results.

### 3.3.3 Invariance of Local Surface Patches to Rigid Transformation

The LSP representation consists of a histogram of shape index and surface normal angle which are invariant to rigid transformation. To verify this, we compute the  $\chi^2$  dissimilarity between reference LSPs and their corresponding LSPs after rigid transformation. We synthetically generate range images at different views by applying 3D rigid transformation. Given a range image  $\{v = \{x, y, z\}\}$ , we apply the transformation ( $tv = R(v - v_0) + v_0$ ) to generate new views, where  $v_0$  is the centroid of 3D vertices in the original range image,  $R$  is the rotation matrix, and  $tv$  are new 3D coordinates after transformation. The rotation matrix  $R$  can be written as  $R = R_\phi * R_\beta * R_\alpha$ , where  $R_\phi$ ,  $R_\beta$ ,  $R_\alpha$  are rotation matrices along the  $x$ -axis,  $y$ -axis, and  $z$ -axis, respectively. We calculate shape index and surface normals for the synthetic range image, compute LSPs at the same location as the reference LSPs, and compute the  $\chi^2$  dissimilarity between the two corresponding LSPs for the extracted feature points. The surface type for the corresponding LSP is not changed by the rigid transformation. The dissimilarity distributions are shown in Fig. 13. Figs. 13a and 13b show the distributions for two different rotations. From this figure, we observe the dissimilarity does not change much and LSP representation is invariant to rigid transformation. Furthermore, as described in Section 5.3, we performed experiments on the UCR data set, which has pose variations ( $\pm 35^\circ$ ) for six different shots of the same subject, and on a subset of the UND data set Collection G, which has pose variations (up to  $45^\circ$ ) for four shots of the same subject. We achieved good performance. These results show the robustness and viewpoint invariance of the LSP representation.

TABLE 4  
Comparison Results for Four Local Surface Patches Shown in Fig. 12

Surface Type	LSP1	LSP2	LSP3	LSP4
	Tp=0	Tp=0	Tp=2	Tp=1
$\chi^2 - divergence$	$\chi^2(LSP1, LSP2)$		$\chi^2(LSP1, LSP3)$	$\chi^2(LSP1, LSP4)$
	0.479		1.99	0.984

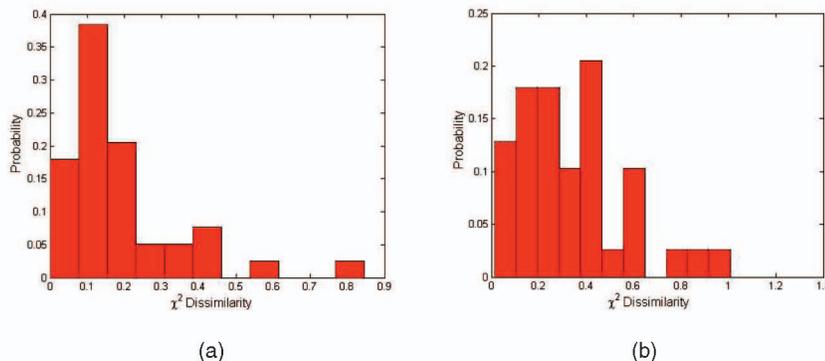


Fig. 13.  $\chi^2$  dissimilarity for corresponding LSPs with respect to the rigid transformation. (a)  $\phi = 5^\circ$ ,  $\beta = 10^\circ$ ,  $\alpha = 10^\circ$  and (b)  $\phi = 10^\circ$ ,  $\beta = 30^\circ$ ,  $\alpha = 30^\circ$ .

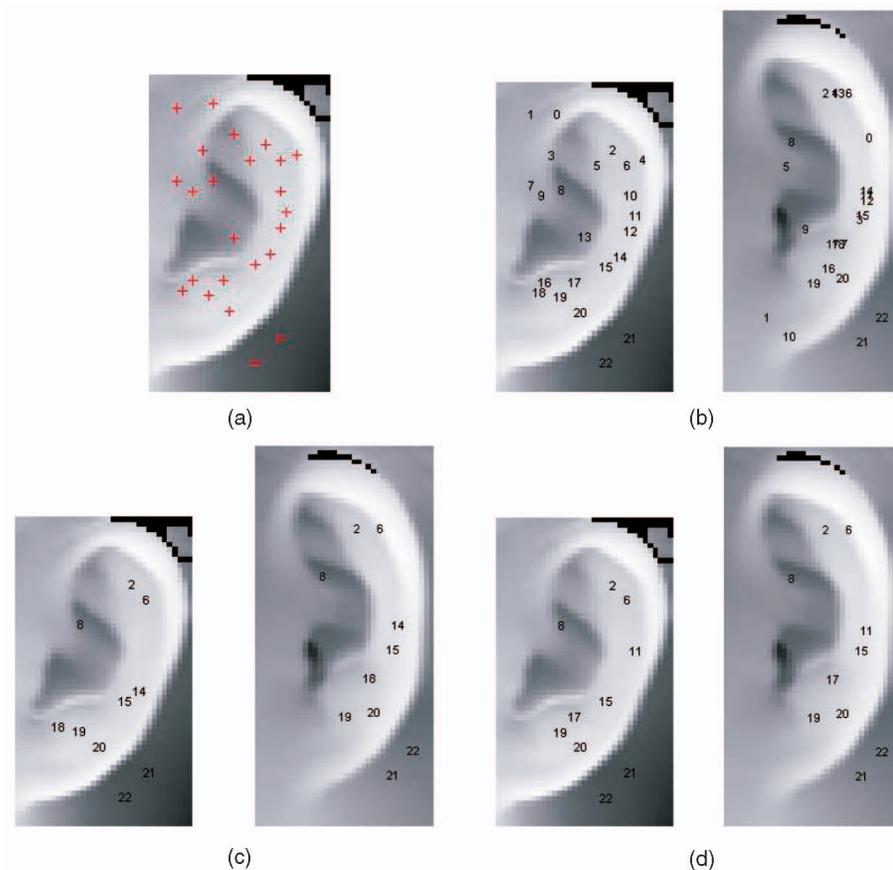


Fig. 14. Example of grouping corresponding LSPs for a pair of ears with a large pose variation. The probe ear is shown on the left image in (b), (c), and (d). (a) Feature points extraction from a probe ear. (b) Initial corresponding pairs. (c) Example of filtered corresponding pairs. (d) Example of filtered corresponding pairs.

### 3.3.4 Grouping Corresponding Pairs of Local Surface Patch

For every local surface patch from the probe ear, we choose the local surface patch from the database with minimum dissimilarity and the same surface type as the possible corresponding patch. We filter the possible corresponding pairs based on the geometric constraints given below:

$$d_{C_1, C_2} = |d_{S_1, S_2} - d_{M_1, M_2}| < \epsilon_2 \quad \max(d_{S_1, S_2}, d_{M_1, M_2}) > \epsilon_3, \quad (11)$$

where  $d_{S_1, S_2}$  and  $d_{M_1, M_2}$  are the Euclidean distance between centroids of two surface patches. The first constraint guarantees that distances  $d_{S_1, S_2}$  and  $d_{M_1, M_2}$  are consistent; the second constraint removes the correspondences which are too close. For two correspondences,  $C_1 = \{S_1, M_1\}$  and  $C_2 = \{S_2, M_2\}$ , where  $S$  is the probe surface patch and  $M$  is the gallery surface patch, they should satisfy (11) if they are consistent corresponding pairs. Therefore, we use simple geometric constraints to partition the potential corresponding pairs into different groups. The larger the group is, the more likely it contains the true corresponding pairs.

Given a list of corresponding pairs  $L = \{C_1, C_2, \dots, C_n\}$ , the grouping procedure for every pair in the list is as follows: Initialize each pair as a group. For every group, add other pairs to it if they satisfy (11). Repeat the same procedure for every group. Sort the groups in the ascending order based on the size of groups. Select the groups on the top of the list.

Fig. 14 shows one example of partitioning corresponding pairs into different groups. Fig. 14a shows the feature point extraction results for a probe ear. Comparing the local surface patches with LSPs on the gallery ear, the initial corresponding pairs are shown in Fig. 14b, in which every pair is represented by the same number superimposed on the probe and gallery images. We observe that both the true and false corresponding pairs are found. The examples of filtered groups after applying the two geometric constraints (11) are shown in Figs. 14c and 14d, respectively. We can see that the true corresponding pairs are obtained by comparing local surface patches and using the simple geometric constraints.

Once the corresponding LSPs between the gallery and probe are established, the initial rigid transformation is estimated and the coarse-to-fine surface matching strategy is followed (see Section 3.2). Note that, in the supplemental material, Section C [25], which can be found at <http://computer.org/tpami/archives.htm>, we provide a quantitative analysis of the effect of the generating parameters on the comparison of LSPs, the robustness of LSPs with respect to noise, and the impact of the number of LSPs in the probe image on the matching performance.

## 4 PERFORMANCE PREDICTION

The prediction of the performance of a biometrics system is an important consideration in the real-world applications.

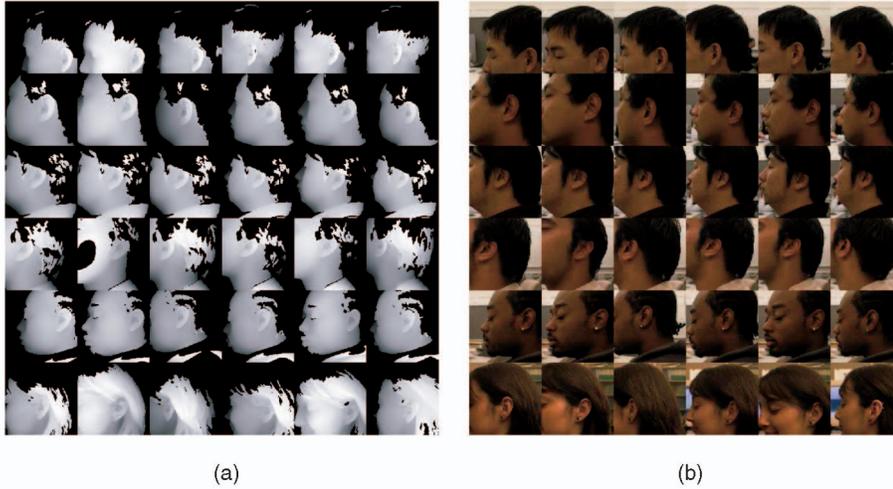


Fig. 15. Examples of side face range images and the corresponding color images for six people in the UCR data set. (a) Range images. (b) Color images. Note the pose variations, the earrings, and the hair occlusions for the six shots.

Our mathematical prediction model is based on the distribution of match and nonmatch scores [46], [47]. Let  $ms(x)$  and  $ns(x)$  denote the distributions of match and nonmatch scores in which the match score is the score computed from the true-matched pair and the nonmatch score is the score computed from the false-matched pair. If the match score is higher, the match is closer. The error occurs when any given match score is less than any of the nonmatch scores. The probability that the nonmatch score is greater than or equal to the match score  $x$  is  $NS(x)$ . It is given by  $NS(x) = \int_x^\infty ns(t)dt$ .

The probability that the match score  $x$  has rank  $r$  exactly is given by the binomial probability distribution  $C_{r-1}^{N-1}(1 - NS(x))^{N-r}NS(x)^{r-1}$ . By integrating over all the match scores, we get

$$\int_{-\infty}^{\infty} C_{r-1}^{N-1}(1 - NS(x))^{N-r}NS(x)^{r-1}ms(x)dx.$$

In theory, the match scores can be any value within  $(-\infty, \infty)$ . Therefore, the probability that the match score is within rank  $r$ , which is the definition of a cumulative match characteristic (CMC) curve, is

$$P(N, r) = \sum_{i=1}^r \int_{-\infty}^{\infty} C_{i-1}^{N-1}(1 - NS(x))^{N-i}NS(x)^{i-1}ms(x)dx. \quad (12)$$

In the above equations,  $N$  is the size of large population whose recognition performance needs to be estimated. Here, we assume that the match score and nonmatch score are independent and the match and nonmatch score distributions are the same for all the people. The small size gallery is used to estimate distributions of  $ms(x)$  and  $ns(x)$ .

For the ear recognition case, every 3D ear in the probe set is matched to every 3D ear in the gallery set and the RMS registration error is calculated using the procedure described in Section 3.2. This RMS registration error is used as the matching score criterion.

## 5 EXPERIMENTAL RESULTS

### 5.1 Data

The experiments are performed on the data set collected by us (UCR data set) and the University of Notre Dame public data set (UND data set).<sup>1</sup> In the UCR data set, 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 data set.

#### 5.1.1 UCR Data Set

The data are captured by Minolta Vivid 300 camera. 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  $200 \times 200$  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  $0.55 \sim 0.75$ 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 could be some face tilt as well, which is not measured. A total of six images per subject are recorded. In total, 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 three different poses in the collected data: frontal, left, and right. Among the total 155 subjects, there are 17 females; six subjects have earrings and 12 subjects have their ears partially occluded by hair (with less than 10 percent occlusion). Fig. 15 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. <http://www.nd.edu/~cvrl/UNDBiometricsDatabase.html>, Collection F and a subset of Collection G.

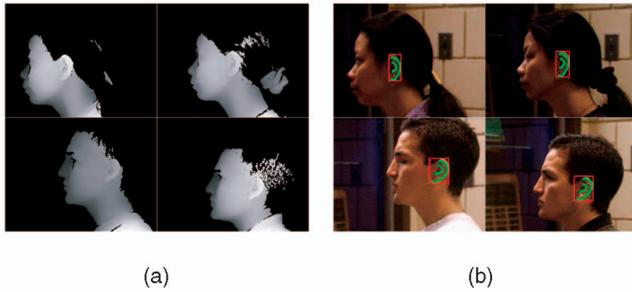


Fig. 16. Examples of side face range images and the corresponding color images for two people in the UND data set Collection F. (a) Range images. (b) Color images. In (a) and (b), the left column shows the gallery images and the right column shows the probe images. In (b), the helix and the antihelix parts are marked by the bright dots and the detected ear is bounded by a rectangular box.

### 5.1.2 UND Data Set

The data are acquired with a Minolta Vivid 910 camera. The camera outputs a  $480 \times 640$  range image and its registered color image of the same size. In Collection F, there are 302 subjects with 302 time-lapse gallery-probe pairs. Fig. 16 shows side face range images and the corresponding color images of two people from this collection. Collection G contains 415 subjects in which 302 subjects are from the Collection F. The most important part of Collection G is that it has 24 subjects with images taken at four different view points. We perform experiments only on these 24 subjects and not the *entire* Collection G because Collection G became available only very recently and it contains the entire Collection F. In this paper, we report the results on the entire Collection F. The identification results (rank-1 recognition rate) by Yan and Bowyer are similar on Collection F (98.7 percent) and Collection G (97.6 percent) [9], [13]. Therefore, we report and compare the rank-1 identification performance of algorithms on a subset of Collection G with pose variations.

## 5.2 Ear Detection Results

### 5.2.1 Ear Detection on UCR Data Set

The ear detection method, described in Section 3.1, combines the range image and the registered color image to locate the helix and the antihelix parts of the ear. The proposed automatic ear detection method is tested on 902 pairs of range and color images. Fig. 17 shows the effectiveness of the global-to-local registration procedure on three people. After the global registration, we get the positions of the 3D coordinates on the 2D image plane and their locations are marked by the bright dots which are shown in Fig. 17a. It can be seen that the shape model is roughly aligned with the ear helix and the antihelix parts. The ear shape model is then driven toward the ear helix and the antihelix parts by minimizing the cost function (6) and their locations are marked by the bright dots which are shown in Fig. 17b. It can be seen that the optimization formulation drives the shape model more close to the true positions with the topology of the reference ear shape model preserved. Fig. 17c shows that the cost function decreases with the number of iterations, which means the optimization formulation works. More examples of ear localization are shown in Fig. 18 in which the detected ear helix and the antihelix parts are shown by the dots superimposed on the 2D color images and the detected

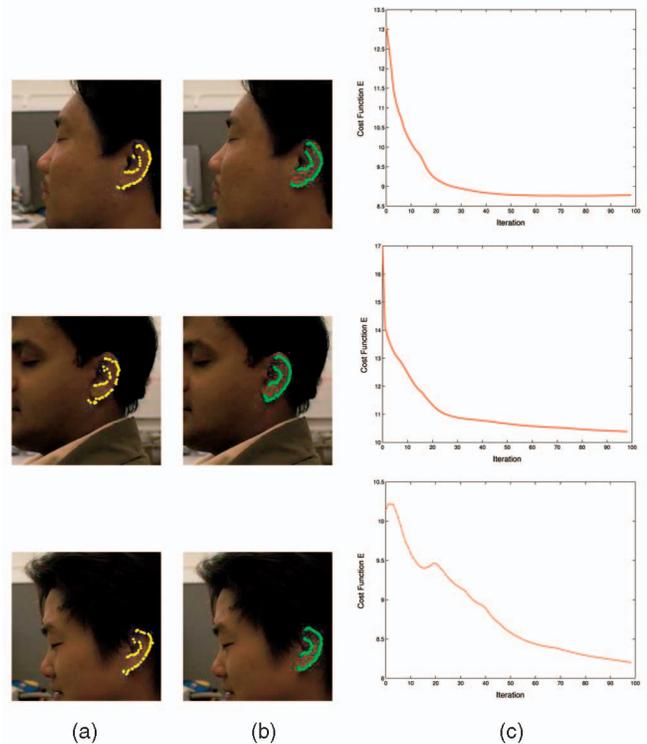


Fig. 17. Examples of global registration and local deformation. (a) Global registration results superimposed on the color images. (b) Local deformation results superimposed on the color images. (c) Cost function (6) versus iteration.

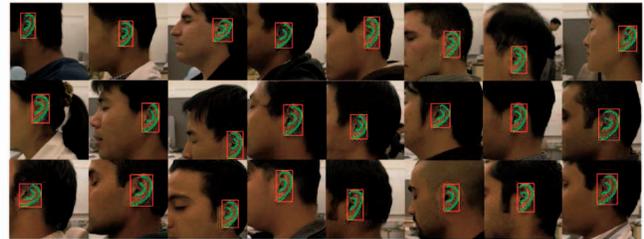


Fig. 18. Results of ear localization on the UCR data set. The helix and the antihelix parts are marked by the bright dots and the detected ear is bounded by a rectangular box.

ear is bounded by the rectangular box. We observe that the ears and their helix and antihelix parts are correctly detected.

In order to quantitatively evaluate the improvement of ear localization through the local deformation driven by the optimization formulation, we compute the error

$$\varepsilon = \frac{1}{N_m} \sum_{i=1}^{N_m} \left( \frac{1}{n} \sum_{j=1}^n \text{Dist}(v_{ij}, G_{ti}) \right)$$

for the global registration and the local deformation, where  $N_m$  is the number of side face range images ( $N_m = 208$  since we manually labeled 3D vertices on the ear helix and the antihelix parts for 208 images for evaluation purposes only),  $n$  is the number of points on the shape model,  $v_{ij}$  is the  $j$ th point on the shape model detected in the  $i$ th side face range image,  $G_{ti}$  is the set of manually labeled 3D points on the ear helix and the antihelix parts of the  $i$ th side face range image, and  $\text{Dist}(v_{ij}, G_{ti})$  is the distance between  $v_{ij}$  and its closest point in  $G_{ti}$ . The error  $\varepsilon$  for the global registration is

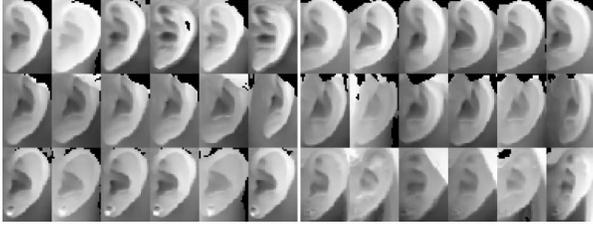


Fig. 19. Examples of extracted ears (from left to right and top to bottom) in the side face range images shown in Fig. 15.

5.4 mm; the error  $\varepsilon$  after the local deformation is 3.7 mm. Thus, the local deformation driven by the optimization formulation really improves the localization accuracy. The proposed global-to-local procedure with the optimization formulation can be applied to handle the nonrigid shape registration. This is shown in the supplemental material (Section A) associated with this paper [25], which can be found at <http://computer.org/tpami/archives.htm>.

Fig. 19 shows the extracted ears from the side face range images in Fig. 15. The average number of points on the ears extracted from 902 side face images is 2,797. The ear detection takes about 9.48 s with Matlab implementation on a 2.4 G Celeron CPU. If the reference ear shape model is aligned with the ear helix and the antihelix parts in a side face range image, we classify it as a positive detection; otherwise a false detection. On the 902 side face range images, we achieve 99.3 percent correct detection rate (896 out of 902).

### 5.2.2 Ear Detection on UND Data Set

Without changing the parameters of the ear detection algorithm on the UCR data set, the proposed automatic ear detection method is tested on 700 ( $302 \times 2 + 24 \times 4 = 700$ ) pairs of range and color images of the UND data set (Collections F and a subset of Collection G). We achieve 87.71 percent correct detection rate (614 out of 700). The average number of points (on 700 images) on the ears is 6,348. Fig. 16b shows the extracted ears from the side face range images in which the ear helix and the antihelix are marked by bright points and the extracted ear is bounded by a rectangular box.

## 5.3 Ear Recognition Results

In order to evaluate and compare the matching performance on the selected data sets, all the ears are correctly extracted. In these limited cases where the ears are not successfully detected in an automated manner, they are correctly extracted by human interaction. Note that we use LSP and helix/antihelix representations separately for the matching. In the UCR data set, there are 155 subjects with 902 images.

The data are split into a gallery set and a probe set. Each set has 155 subjects and every subject in the probe set has an instance in the gallery set. In order to evaluate the proposed surface matching schemes, we perform experiments under two scenarios: 1) One frontal ear of a subject is in the gallery set and another frontal ear of the same subject is in the probe set and 2) two frontal ears of a subject are in the gallery set and the rest of the ear images of the same subject are in the probe set. These two scenarios are denoted as  $ES_1$  and  $ES_2$ , respectively.  $ES_1$  is used for testing the performance of the system to recognize ears with the same pose;  $ES_2$  is used for testing the performance of the system to recognize ears with pose variations. In the UND data set Collection F, there are 302 subjects and each subject has two images. The gallery set has 302 images and the probe set has the corresponding 302 images. The experimental results on the UND data set Collection F are obtained using the same parameters of the ear recognition algorithm as those used on the UCR data set. Note that the resolution of the sensors for the UCR and UND data sets are different. We anticipate improvement in performance by fine tuning the parameters on the UND data set. However, these experiments are not performed since we wanted to keep the algorithm parameters fixed across data sets.

### 5.3.1 Identification Performance

Every probe is matched to every 3D ear in the gallery set and the RMS registration error is calculated using the procedure described in Section 3.2. The subject in the gallery set with the minimum error is declared as the recognized person in the probe image. The identification performance is evaluated by the cumulative match characteristics (CMC), which describes “is the right answer in the top rank- $r$  matches?” Table 5 shows the rank- $r$  recognition rates for the UCR data set and the UND data set Collection F using the ear helix/antihelix and the LSP representations. In Table 5, the numbers of images in the gallery and the probe sets are listed in the parenthesis following the name of the data set. Using the ear helix/antihelix representation, we achieve 96.77 percent rank-1 recognition rate (150 out of 155) on the UCR data set  $ES_1$  and 96.36 percent rank-1 recognition rate (291 out of 302) on the UND data set Collection F. As expected, the system performs better on  $ES_1$  with the same pose and the performance degrades slightly on  $ES_2$  with pose variations. The performance of the system using the LSP representation is better than the helix/antihelix representation on the UND data set but a little lower on the UCR data set. We observe that, without retuning the parameters of the proposed algorithm, we still achieved good recognition performance on the UND data set, which has several weeks of time lapse between the gallery and the probe. For the LSP representation, the average time to match a pair of ears, which includes computation of

TABLE 5  
Cumulative Matching Performance on the UCR Data Set and the UND Data Set Collection F

Dataset	Helix/anti-helix representation					LSP representation				
	Rank-1	Rank-2	Rank-3	Rank-4	Rank-5	Rank-1	Rank-2	Rank-3	Rank-4	Rank-5
UCR $ES_1$ (155, 155)	96.77%	98.06%	98.71%	98.71%	98.71%	94.84%	96.77%	96.77%	96.77%	96.77%
UCR $ES_2$ (310, 592)	94.43%	96.96%	97.80%	98.31%	98.31%	94.43%	96.96%	97.30%	97.64%	97.80%
UND(302, 302)	96.03%	96.69%	97.35%	97.68%	98.01%	96.36%	98.01%	98.34%	98.34%	98.34%

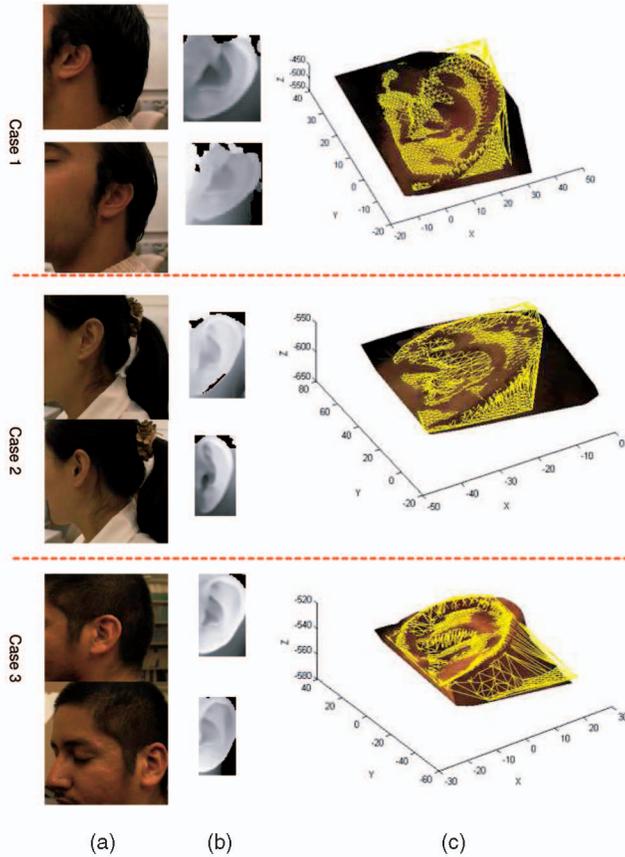


Fig. 20. UCR data set: Three cases of the correctly recognized gallery-probe ear pairs using the ear helix/antihelix representation with a large pose variation. (a) Side face color images. (b) Range images of the detected ears. In columns (a) and (b), the gallery image is shown first and the probe image is shown second. (c) The probe ear with the corresponding gallery ear after alignment. The gallery ear represented by the mesh is overlaid on the textured 3D probe ear. The units of X, Y, and Z are millimeters (mm). In Case 1, the rotation angle is  $33.5^\circ$  and the axis is  $[0.0099, 0.9969, 0.0778]^T$ . In Case 2, the rotation angle is  $-33.5^\circ$  and the axis is  $[-0.1162, 0.9932, 0.0044]^T$ . In Case 3, the rotation angle is  $32.9^\circ$  and the axis is  $[0.0002, 0.9998, 0.0197]^T$ .

LSPs and surface matching, is about 3.7 seconds with C++ implementation on a Linux machine with a AMD Opteron 1.8 GHz processor. For the helix/antihelix representation, the average time to match a pair of ears is about 1.1 seconds with C++ implementation on the same platform.

Fig. 20 shows three examples of the correctly recognized gallery-probe ear pairs with a large pose variation using the ear helix/antihelix representation. Fig. 20a shows the side face color images of the gallery and the probe alternately, Fig. 20b shows the range images of the ears that are automatically extracted and Fig. 20c shows the gallery ear represented by the mesh overlaid on the textured 3D probe ear images. We observe that the cases with a large pose variation are correctly handled.

We show four special cases of correctly recognizing gallery-probe ear pairs using the LSP representation in Fig. 21. In this figure, each probe ear is rendered as a textured 3D surface and each gallery ear is displayed as a mesh. In order to examine the results visually, we display the prealigned gallery ear and the probe ear in the same image (Fig. 21b) and also the postaligned (transformed) gallery and the probe ear in the same image (Fig. 21c). From Fig. 21, we

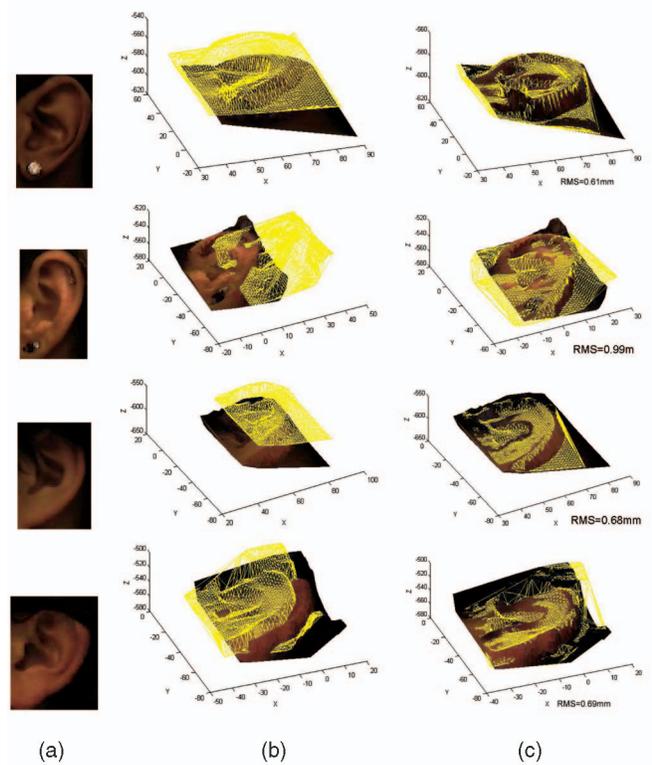


Fig. 21. UCR data set: Four examples of the correctly recognized gallery-probe pairs using the LSP representation. Two ears have earrings and the other two ears are partially occluded by the hair. Images in (a) show color images of ears. Images in (b) and (c) show the probe ear with the corresponding gallery ear before the alignment and after the alignment, respectively. The gallery ears represented by the mesh are overlaid on the textured 3D probe ears. The units of X, Y, and Z are millimeters (mm).

observe that the ear recognition system can handle partial occlusion. Twelve more examples of correctly recognized gallery-probe ear pairs (using both the helix/antihelix and the LSP representations) are shown in the supplemental material (Section C) that accompanies this paper [25], which can be found at <http://computer.org/tpami/archives.htm>.

During the recognition, some errors are made and the four error cases are illustrated in Fig. 22. Figs. 22a and 22b show the color images of two visually similar probe and gallery ears that belong to different subjects; Fig. 22c shows the true gallery ear overlaid on the textured 3D probe ear after registration; Fig. 22d shows the falsely recognized gallery ear overlaid on the textured 3D probe ear after alignment. In Fig. 22d, the root mean square error for the falsely recognized ear is smaller than the error for the correct ear in Fig. 22c. In this figure, we obtain good alignment between the gallery and probe ears from different people since these ears are quite similar in 3D.

### 5.3.2 Verification Performance

The verification performance of the proposed system is evaluated in terms of the two popular methods, the receiver operating characteristic (ROC) curve and the equal error rate (EER). The ROC curve is the plot of the genuine acceptance rate (GAR) versus the corresponding false acceptance rate (FAR). GAR is defined as the percentage of the occurrences where an authorized user is correctly accepted by the system, while FAR is defined as the

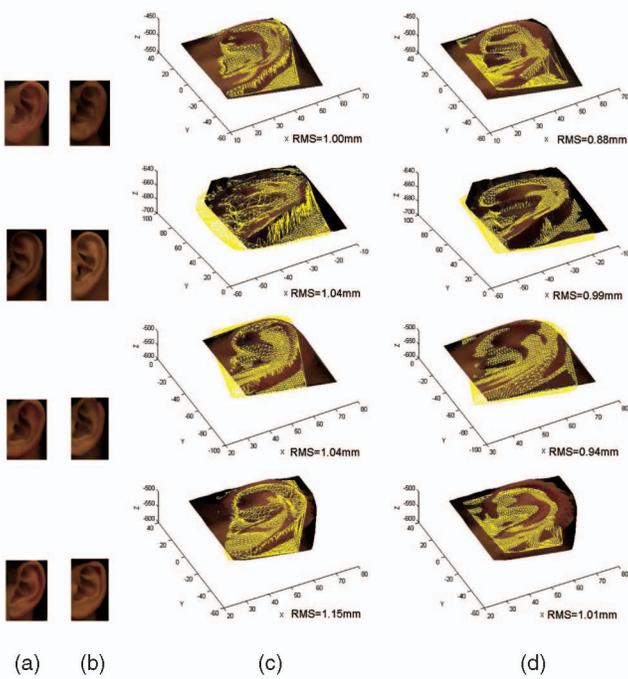


Fig. 22. UCR data set: Four cases of incorrectly recognized gallery-probe pairs using the ear helix/antihelix representation. Each row shows one case. The gallery ears represented by the mesh are overlaid on the textured 3D probe ears. The units of X, Y, and Z are millimeters (mm). (a) Color images of the probe ears. (b) Color images of falsely recognized gallery ears. (c) True gallery ears after alignment are overlaid on the textured 3D probe ears. (d) The falsely recognized gallery ears after alignment are overlaid on the textured 3D probe ears. Note that, for the incorrect matches, the gallery ears (d) achieve a smaller value of RMS error than the gallery ears in (c).

percentage of the occurrences where a nonauthorized user is falsely accepted by the system. The EER, which, indicates the rate at which the false rejection rate ( $FRR = 1 - GAR$ ) and the false acceptance rate are equal, is a threshold independent performance measure.

During the verification, the RMS distance is computed from matching the gallery ears to the probe ears and it is then compared to a threshold to determine if the probe is an authorized user or an imposter. By varying the threshold, FAR and GAR values are computed and plotted in Fig. 23.

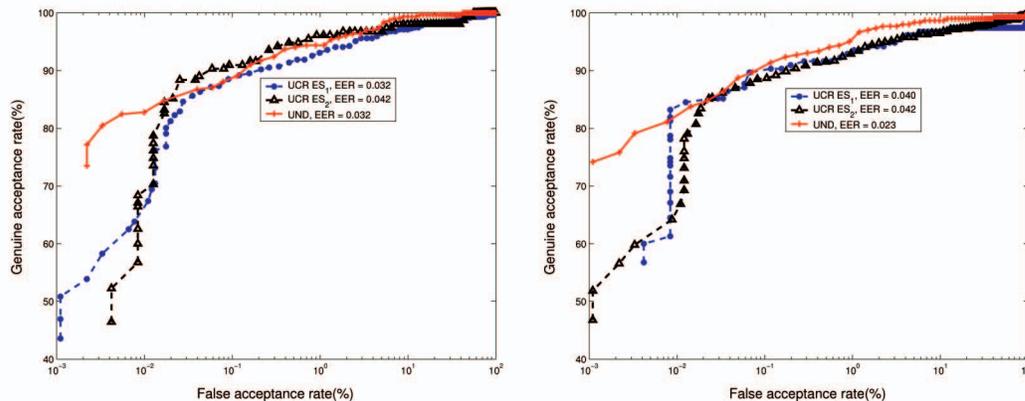


Fig. 23. UCR data set and UND data set Collection F: Verification performance as an ROC curve. (a) ROC curves on the UCR data set  $ES_1$ ,  $ES_2$ , and the UND data set using the ear helix/antihelix representation. (b) ROC curves on the UCR data set  $ES_1$ ,  $ES_2$  and the UND data set using the LSP representation.

Fig. 23a shows the ROC curves on the UCR and the UND data set Collection F using the ear helix/antihelix representation for surface matching; Fig. 23b shows the ROC curves on the UCR and the UND data set Collection F using the LSP representation for surface matching. As expected, the system performs better on  $ES_1$  than on  $ES_2$  using the ear helix/antihelix and the LSP representations. We obtain the best performance with a 0.023 EER on the UND data set using the LSP representation. It is clearly seen that, without retuning the parameters of the proposed algorithms, we achieved good verification performance on the UND data set.

Based on the ROC curve, we can select a threshold which satisfies the user's requirement for the false alarm rate or false rejection rate. This threshold can be used to reject the unauthorized users.

### 5.3.3 Evaluation of the Verification Performance

We discuss and evaluate the accuracy of the ear verification system following the method in [16], [48], [49]. As described in Section 5.1, the UCR data set has 155 subjects. There are 155 probes providing 155 user claims and 23,870 ( $155 \times 154$ ) imposter claims for the UCR data set  $ES_1$ . For the UCR data set  $ES_2$ , there are 592 probes providing 592 user claims and 91,168 ( $592 \times 154$ ) imposter claims. The UND data set Collection F has 302 subjects. There are 302 pairs of images providing 302 user claims and 90,902 ( $302 \times 301$ ) imposter claims.

We calculate the number of user claims and imposter claims that gives statistically significant results. Let  $\zeta_\mu$  denote the  $\mu$ -percentile of the standard Gaussian distribution with zero mean and unit variance. Since the verification tests can be thought of as Bernoulli trials, we can assert that, with confidence  $\delta = 1 - \mu$ , the minimum number of user claims which ensures that the expected value of FRR ( $p_{FRR}$ ) and the empirical value ( $\hat{p}_{FRR}$ ) are related by  $|p_{FRR} - \hat{p}_{FRR}| \leq \xi$  is given by

$$v_C = \left( \frac{\zeta_{\mu/2}}{\xi} \right)^2 p_{FRR}(1 - p_{FRR}). \quad (13)$$

The number  $v_I$  of the imposter claims [49] which is sufficient to ensure that the expected value of FAR ( $p_{FAR}$ ) and the empirical value ( $\hat{p}_{FAR}$ ) are related by  $|p_{FAR} - \hat{p}_{FAR}| \leq \xi$  is given by:

TABLE 6

Comparison between Chen and Bhanu's Approach and Yan and Bowyer's Approach on a Subset of Collection G

Probe\Gallery	Straight-on	15° off	30° off	45° off	Average
<b>Straight-on</b>		[100%, 100%], 100%	[91.7%, 87.5%], 87.5%	[87.5%, 83.3%], 70.8%	[93.1%, 90.3%], 86.1%
15° off	[100%, 100%], 100%		[100%, 100%], 100%	[87.5%, 91.7%], 87.5%	[95.8%, 97.2%], 95.8%
30° off	[91.7%, 91.7%], 87.5%	[100%, 100%], 100%		[95.8%, 91.7%], 95.8%	[95.8%, 94.4%], 94.4%
45° off	[87.5%, 87.5%], 79.2%	[91.7%, 87.5%], 87.5%	[95.8%, 87.5%], 100%		[91.7%, 87.5%], 88.9%
<b>Average</b>	[93.1%, 93.1%], 88.9%	[97.2%, 95.8%], 95.8%	[95.8%, 91.7%], 95.8%	[90.3%, 88.9%], 84.7%	[94.1%, 92.4%], 91.3%

Our rank-1 identification results are put inside the brackets. The first number is obtained using the helix/antihelix representation and the second one is obtained using the LSP representation. The number outside the bracket is obtained from [50, chapter 8.3].

TABLE 7

Comparison of Results between Chen and Bhanu's and Yan and Bowyer's Approaches

		Chen and Bhanu's approach (This paper)	Yan and Bowyer's approach [8–13]
Results	Identification	96.4% rank-1, 98.0% rank-2 recognition rate on Collection F, 94.4% rank-1 recognition rate on the UCR dataset $ES_2$	98.7% rank-1 recognition rate on Collection F, 97.6% rank-1 recognition rate on Collection G, no results on the UCR dataset
	Verification	EER = 0.023 on Collection F and EER = 0.042 on the UCR dataset $ES_2$	EER = 0.012 on Collection G, EER on Collection F is not known and no results on the UCR dataset
	Transformation	Transformation matrix for some gallery-probe pairs are shown	No
	Visualization	Matching of some gallery-probe pairs are shown	No
	Prediction	Recognition performance prediction on large datasets	No

$$v_I = \left( \frac{\zeta_{\mu/2}}{\xi'} \right)^2 p_o (1 - p_o), \quad \xi' = \frac{\xi}{k}, \quad (1 - p_o)^k = 1 - p_{FA}, \quad (14)$$

where  $p_o$  is the probability that one imposter is falsely accepted as an authorized user and  $k$  is the number of imposter claims ( $k = 154$  for the UCR data set and  $k = 301$  for the UND data set Collection F). By setting the desired EER,  $\delta$  and  $\xi$ , we can compute  $v_C$  and  $v_I$ . For the UCR data set  $ES_1$ , we find  $v_C = 149$  and  $v_I = 24,759$  with  $EER = 5\%$ ,  $\delta = 95\%$ , and  $\xi = 3.5\%$ . For the UCR data set  $ES_2$ , we find  $v_C = 456$  and  $v_I = 75,826$  with  $EER = 5\%$ ,  $\delta = 95\%$ , and  $\xi = 2\%$ . For the UND data set Collection F, we find  $v_C = 292$  and  $v_I = 94,874$  with  $EER = 5\%$ ,  $\delta = 95\%$ , and  $\xi = 2.5\%$ . Note that the order of magnitude of these numbers are the same as those provided by the test scenarios. The values of  $\xi$  on the UCR data set  $ES_2$  ( $\xi = 2\%$ ) and the UND data set Collection F ( $\xi = 2.5\%$ ) are smaller than the value on the UCR data set  $ES_1$  ( $\xi = 3.5\%$ ) since the UCR data set  $ES_2$  and the UND data set Collection F are larger in size than the size of the UCR data set  $ES_1$ .

### 5.3.4 Comparison between Chen and Bhanu's Approach and Yan and Bowyer's Approach

In the UND data set Collection G, there are 24 subjects whose images are taken at four different poses, straight-on, 15° off center, 30° off center, and 45° off center. For each angle of an ear image, we match it against all the images at different angles. This is the same experiment performed in [50, chapter 8.3]. The results are given in Table 6. We observe that the results obtained by the helix/antihelix representation are better than Yan and Bowyer's results [50, chapter 8.3] for all the cases except the case (45° probe against 30° gallery). The results obtained by the LSP representation outperform the results obtained in [50, chapter 8.3] for the cases with large

pose variations (45° and 30° probes against straight-on gallery, straight-on, and 15° probes against 45° gallery). From Table 6, we see that our representations can reasonably handle the pose variation up to 45°. In Table 7, we provide a comparison of the experimental results for the two approaches. As discussed in Section 2.1, the most important difference between the two approaches is that, in our system, the ear helix/antihelix and the local surface patch (LSP) representations are used (independently) to estimate the initial 3D transformation, which is critical for the success of ICP algorithm. The Yan and Bowyer's approach uses the ICP algorithm directly with the initialization of translation only (no rotation).

### 5.3.5 Comparison of LSP and Spin Image Representations

In order to compare the distinctive power of the LSP and the spin image representations, we follow the same procedures as described in this paper to recognize ears using the spin image representation. In the experiments, the size of the spin image is  $15 \times 15$ . We perform experiments on the UCR data set  $ES_1$  (155 shots in the gallery and 155 shots in the probe) to compare the performance of these two representations in terms of the CMC and the ROC curves. Table 8 shows the CMC values using the LSP and the spin image representations for ear recognition. Fig. 24 shows the ROC curves using the LSP and the spin image representations for matching ears. From Table 8 and Fig. 24, we observe that the LSP representation achieved a slightly better performance than the spin image representation.

## 5.4 Prediction of the Recognition Performance

We predict the CMC performance based on the matching distance distributions obtained on the UCR data set  $ES_1$

TABLE 8  
Cumulative Matching Performance on the UCR Data Set  $ES_1$   
Using the LSP and the Spin Image Representations

Representation	Rank-1	Rank-2	Rank-3	Rank-4	Rank-5
LSP	94.84%	96.77%	96.77%	96.77%	96.77%
Spin Image	92.90%	95.48%	95.48%	95.48%	95.48%

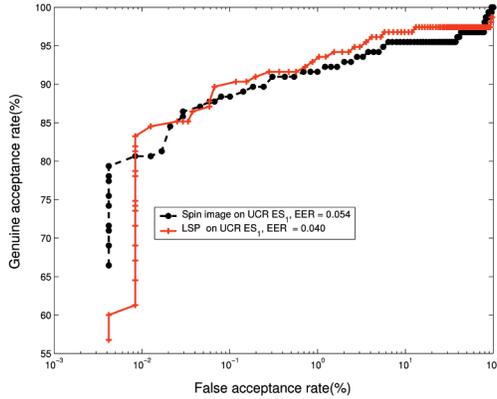


Fig. 24. Verification performance as ROC curves using the LSP and the spin image representations on the UCR data set  $ES_1$ .

using the ear helix and antihelix representation. Every 3D ear image in the probe set is matched to every 3D ear in the gallery set and the RMS registration error is calculated using the procedure described in Section 3.2. The RMS registration error is used as the matching distance. Therefore, 155 true-match distances and 11,935 ( $C_2^{155}$ ) nonmatch distances are obtained. The matching distance distributions for the true-match and the nonmatch are shown in Fig. 25. Note that there is an overlap between the two distributions, which accounts for false matches. Based on the distributions, the CMC curve  $P(N, r)$  can be predicted where  $r = 1, 2, 3, 4, 5$ , and  $N$  is the database size. The results of the directly calculated CMC curve and the predicted CMC curve are shown in Fig. 26. Fig. 26 shows that the predicted CMC values on 155 subjects are close to the real CMC values, which demonstrates the effectiveness of the prediction model. The predicted CMC curves for larger data sets of 300 and 500 subjects are also provided in Fig. 26, which shows that the recognition performance degrades slowly with the increase in database size. This suggests the scalability of the proposed system.

## 6 CONCLUSIONS

We presented a human recognition system using 3D ear biometrics, which performed 3D ear detection, 3D ear identification, and 3D ear verification. We proposed a novel ear detection approach that uses both color and range images to localize the ear region accurately by following a global-to-local registration procedure. This procedure with its optimization formulation can be applied to the nonrigid shape registration and the detection of boundary of similar objects. We demonstrated it for the alignment of the simple shapes (in the supplemental material Section A, which can be found at <http://computer.org/tpami/archives.htm>) [25] and the localization of ear helix and the antihelix parts.

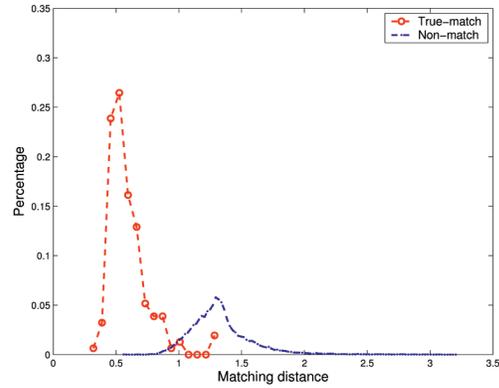


Fig. 25. Matching distance distribution for match and nonmatch pairs on the UCR data set  $ES_1$ .

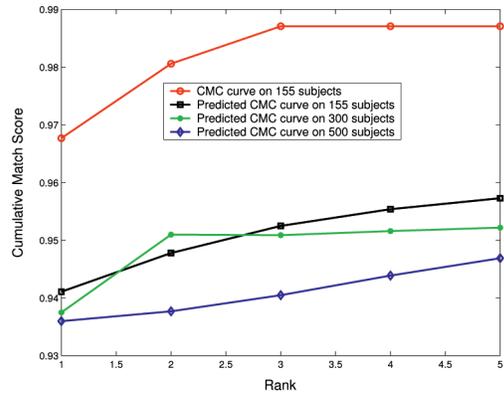


Fig. 26. Real and predicted CMC curves on the UCR data set  $ES_1$ .

After the ear is automatically extracted from the side face range images, we proposed two different representations for surface matching. The first representation is the ear helix/antihelix, whose 3D coordinates are obtained from the detection algorithm; the second representation is the local surface patch (LSP), which is invariant to rotation and translation. We used these representations for finding initial correspondences between a gallery-probe pair. Then, a modified Iterative Closest Point (ICP) algorithm iteratively refined the transformation, which brings the hypothesized gallery and a probe image into the best alignment. The root mean square (RMS) registration error is used as the matching error criterion. The experimental results on two real ear range and color image data sets demonstrated the potential of the proposed algorithms for robust ear recognition in 3D. Extensive experiments are performed on the UCR data set (155 subjects with 902 images under pose variations), the UND data set Collection F (302 subjects with 302 time-lapse gallery-probe pairs), and a subset of the UND data set G for evaluating the performance with respect to pose variations without retuning the parameters of the proposed algorithms. These results showed that the proposed ear recognition system is capable of recognizing ears under pose variations, partial occlusions, and time lapse effects. The proposed representations are less sensitive to pose variations. We also provided a comparison of the LSP representation with the spin image representation for identification and verification. This comparison showed that the LSP representation achieved a slightly better performance than the spin image

representation. Furthermore, the performance prediction model showed the scalability of the proposed system with increased database size.

Our experimental results show that ear biometrics has the potential to be used in the real-world applications to identify/authenticate humans by their ears. It can be used in both the low and high security applications and in combination with other biometrics such as face. With the decreasing cost (few thousand dollars) and size of a 3D scanner and the increased performance, we believe that 3D ear biometrics will be highly useful in many real-world applications in the future.

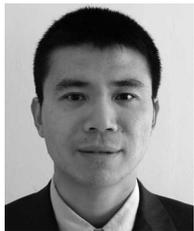
## ACKNOWLEDGMENTS

The authors would like to thank the computer vision research laboratory at the University of Notre Dame for providing their public biometrics database Collections F and G that are used in this paper. The authors would like to thank Dr. Yan and Dr. Bowyer for providing their publications and some useful discussions.

## REFERENCES

- [1] *BIOMETRICS: Personal Identification in Network Society*, A. Jain et al., eds. Kluwer Academic, 1999.
- [2] A. Iannarelli, *Ear Identification*. Paramont Publishing, 1989.
- [3] M. Burge and W. Burger, "Ear Biometrics in Computer Vision," *Proc. Int'l Conf. Pattern Recognition*, vol. 2, pp. 822-826, 2000.
- [4] D. Hurlley, M. Nixon, and J. Carter, "Force Field Feature Extraction for Ear Biometrics," *Computer Vision and Image Understanding*, vol. 98, no. 3, pp. 491-512, 2005.
- [5] K. Chang, K.W. Bowyer, S. Sarkar, and B. Victor, "Comparison and Combination of Ear and Face Images in Appearance-Based Biometrics," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1160-1165, Sept. 2003.
- [6] B. Bhanu and H. Chen, "Human Ear Recognition in 3D," *Proc. Workshop Multimodal User Authentication*, pp. 91-98, 2003.
- [7] H. Chen and B. Bhanu, "Contour Matching for 3D Ear Recognition," *Proc. Seventh IEEE Workshop Application of Computer Vision*, vol. 1, pp. 123-128, 2005.
- [8] P. Yan and K.W. Bowyer, "Multi-Biometrics 2D and 3D Ear Recognition," *Proc. Audio and Video-Based Biometric Person Authentication*, pp. 503-512, 2005.
- [9] P. Yan and K.W. Bowyer, "Empirical Evaluation of Advanced Ear Biometrics," *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshop Empirical Evaluation Methods in Computer Vision*, 2005.
- [10] P. Yan and K.W. Bowyer, "Ear Biometrics Using 2D and 3D Images," *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshop Advanced 3D Imaging for Safety and Security*, 2005.
- [11] P. Yan and K.W. Bowyer, "ICP-Based Approaches for 3D Ear Recognition," *Proc. SPIE Conf. Biometric Technology for Human Identification*, pp. 282-291, 2005.
- [12] P. Yan and K.W. Bowyer, "A Fast Algorithm for ICP-Based 3D Shape Biometrics," *Proc. IEEE Workshop Automatic Identification Advanced Technologies*, pp. 213-218, 2005.
- [13] P. Yan and K.W. Bowyer, "An Automatic 3D Ear Recognition System," *Proc. Third Int'l Symp. 3D Data Processing, Visualization, and Transmission*, pp. 213-218, 2006.
- [14] J.C. Lee and E. Miliotis, "Matching Range Images of Human Faces," *Proc. Int'l Conf. Computer Vision*, pp. 722-726, 1990.
- [15] G.G. Gordon, "Face Recognition Based on Depth and Curvature Features," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 808-810, 1992.
- [16] F. Tsalkanidou, S. Malasiotis, and M.G. Strintzis, "Face Localization and Authentication Using Color and Depth Images," *IEEE Trans. Image Processing*, vol. 14, no. 2, pp. 152-168, 2005.
- [17] C.S. Chua, F. Han, and Y. Ho, "3D Human Face Recognition Using Point Signatures," *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pp. 233-238, 2000.
- [18] A. Bronstein, M. Bronstein, and R. Kimmel, "Expression-Invariant 3D Face Recognition," *Proc. Audio and Video-Based Biometric Person Authentication*, pp. 62-70, 2003.
- [19] K. Chang, K.W. Bowyer, and P. Flynn, "An Evaluation of Multimodal 2D+3D Face Biometrics," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 619-624, Apr. 2005.
- [20] X. Lu and A.K. Jain, "Integrating Range and Texture Information for 3D Face Recognition," *Proc. Seventh IEEE Workshop Application of Computer Vision*, vol. 1, pp. 156-163, 2005.
- [21] M.J. Jones and J.M. Rehg, "Statistical Color Models with Application to Skin Detection," *Int'l J. Computer Vision*, vol. 46, no. 1, pp. 81-96, 2002.
- [22] T.F. Cootes, D. Cooper, C.J. Taylor, and J. Graham, "Active Shape Models—Their Training and Application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38-59, 1995.
- [23] C.S. Chen, Y.P. Hung, and J.B. Cheng, "RANSAC-Based DARCES: A New Approach to Fast Automatic Registration of Partially Overlapping Range Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 11, pp. 1229-1234, Nov. 1999.
- [24] F.L. Bookstein, "Principal Warps: Thin-Plate Splines and the Decomposition of Deformations," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 6, pp. 567-585, June 1989.
- [25] H. Chen and B. Bhanu, "Supplemental Material to Accompany 'Human Ear Recognition in 3D'," <http://computer.org/tpami/archives.htm>, 2007.
- [26] C. Boehnen and T. Russ, "A Fast Multi-Modal Approach to Facial Feature Detection," *Proc. Seventh IEEE Workshops Application of Computer Vision*, vol. 1, pp. 135-142, 2005.
- [27] X. Lu, D. Colbry, and A.K. Jain, "Three-Dimensional Model Based Face Recognition," *Proc. Int'l Conf. Pattern Recognition*, vol. 1, pp. 362-366, 2004.
- [28] S.H. Kim, N.K. Kim, S.C. Ahn, and H.G. Kim, "Object Oriented Face Detection Using Range and Color Information," *Proc. Int'l Conf. Face and Gesture Recognition*, pp. 76-81, 1998.
- [29] H. Chen and B. Bhanu, "Human Ear Detection from Side Face Range Images," *Proc. Int'l Conf. Pattern Recognition*, vol. 3, pp. 574-577, 2004.
- [30] H. Chen and B. Bhanu, "Shape Model-Based 3D Ear Detection from Side Face Range Images," *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshop Advanced 3D Imaging for Safety and Security*, 2005.
- [31] P. Besl and N.D. Mckay, "A Method of Registration of 3D Shapes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239-256, Feb. 1992.
- [32] Y. Chen and G. Medioni, "Object Modelling by Registration of Multiple Range Images," *Image and Vision Computing*, vol. 10, no. 3, pp. 145-155, 1992.
- [33] Z. Zhang, "Iterative Point Matching for Registration of Free-Form Curves and Surfaces," *Int'l J. Computer Vision*, vol. 13, no. 2, pp. 119-152, 1994.
- [34] S. Rusinkiewicz and M. Levoy, "Efficient Variants of the ICP Algorithm," *Proc. Int'l Conf. 3-D Digital Imaging and Modeling*, pp. 145-152, 2001.
- [35] H. Chui and A. Rangaraian, "A New Algorithm for Non-Rigid Point Matching," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 44-51, 2000.
- [36] S. Wang, T. Kubota, and T. Richardson, "Shape Correspondence through Landmark Sliding," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 143-150, 2004.
- [37] J. Lim and M.H. Yang, "A Direct Method for Modeling NonRigid Motion with Thin Plate Spline," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 196-202, 2005.
- [38] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *Int'l J. Computer Vision*, vol. 1, no. 14, pp. 321-331, 1988.
- [39] B. Horn, "Close-Form Solution of Absolute Orientation Using Unit Quaternions," *J. Optical Soc. Am.*, vol. 4, no. 4, pp. 629-642, 1987.
- [40] R.J. Campbell and P.J. Flynn, "A Survey of Free-Form Object Representation and Recognition Techniques," *Computer Vision and Image Understanding*, vol. 81, pp. 166-210, 2001.
- [41] P. Flynn and A. Jain, "On Reliable Curvature Estimation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 110-116, 1989.
- [42] C. Dorai and A. Jain, "COSMOS-A Representation Scheme for 3D Free-Form Objects," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 10, pp. 1115-1130, Oct. 1997.
- [43] J.J. Koenderink and A.V. Doorn, "Surface Shape and Curvature Scales," *Image Vision Computing*, vol. 10, no. 8, pp. 557-565, 1992.
- [44] A. Johnson and M. Hebert, "Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 433-449, May 1999.

- [45] B. Schiele and J. Crowley, "Recognition without Correspondence Using Multidimensional Receptive Field Histograms," *Int'l J. Computer Vision*, vol. 36, no. 1, pp. 31-50, 2000.
- [46] B. Bhanu, R. Wang, and X. Tan, "Predicting Fingerprint Recognition Performance from a Small Gallery," *Pattern Recognition Letters*, vol. 28, pp. 40-48, 2007.
- [47] H. Chen, B. Bhanu, and R. Wang, "Performance Evaluation and Prediction for 3D Ear Recognition," *Proc. Conf. Audio and Video-Based Biometric Person Authentication*, pp. 748-757, 2005.
- [48] C. Kotropoulos, A. Tefas, and I. Pitas, "Frontal Face Authentication Using Discriminating Grids with Morphological Feature Vectors," *IEEE Trans. Multimedia*, vol. 2, no. 1, pp. 14-26, Jan.-Mar. 2000.
- [49] W. Shen, M. Surette, and R. Khanna, "Evaluation of Automated Biometric-Based Identification and Verification Systems," *Proc. IEEE*, vol. 85, no. 9, pp. 1464-1478, 1997.
- [50] P. Yan, "Ear Biometrics in Human Identification," PhD dissertation, Dept. of Computer Science and Eng., Univ. of Notre Dame, 2006.



**Hui Chen** received the MS and BS degree in control theory and applications from the University of Science and Technology, Beijing, China, in 1999 and 1996, respectively, and the PhD degree in electrical engineering from the University of California at Riverside in 2006. He worked as a software engineer at the Chinese Academy of Sciences, Beijing, China, during 1999-2001. He also worked as an intern at Siemens Medical Solutions, Malvern, Pennsylvania, during the summer of 2005. His research interests are 3D object detection, tracking, and recognition, medical image processing, 3D biometrics, shape registration, and range image analysis. He is a student member of the IEEE.



**Bir Bhanu** received the SM and EE degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, the PhD degree in electrical engineering from the Image Processing Institute, University of Southern California, Los Angeles, and the MBA degree from the University of California, Irvine. He was the founding professor of electrical engineering and served as its first chair at the University of California at Riverside (UCR). He has been the Cooperative Professor of Computer Science and Engineering and the director of the Visualization and Intelligent Systems Laboratory (VISLab) since 1991. Currently, he also serves as the founding director of an interdisciplinary Center for Research in Intelligent Systems (CRIS) at UCR. Previously, he was a Senior Honeywell Fellow at Honeywell Inc., Minneapolis, Minnesota. He has been on the faculty of the Department of Computer Science at the University of Utah, Salt Lake City, and has worked at Ford Aerospace and Communications Corporation, California, INRIA-France, and IBM San Jose Research Laboratory, California. He has been the principal investigator of various programs for the NSF, DARPA, NASA, AFOSR, ARO, and other agencies and industries in the areas of video networks, video understanding, learning and vision, image understanding, pattern recognition, target recognition, biometrics, navigation, image databases, and machine vision applications. He is the coauthor of the books *Computational Learning for Adaptive Computer Vision* (Springer-Verlag, 2007), *Evolutionary Synthesis of Pattern Recognition Systems* (Springer-Verlag, 2005), *Computational Algorithms for Fingerprint Recognition* (Kluwer, 2004), *Genetic Learning for Adaptive Image Segmentation* (Kluwer, 1994), and *Qualitative Motion Understanding* (Kluwer, 1992), and the coeditor of the book *Computer Vision Beyond the Visible Spectrum* (Springer-Verlag, 2004). He holds 11 US and international patents and more than 250 reviewed technical publications in the areas of his interest. Dr. Bhanu has received two outstanding paper awards from the Pattern Recognition Society and has received industrial and university awards for research excellence, outstanding contributions, and team efforts. He has been on the editorial board of various journals and has edited special issues of several IEEE transactions (*TPAMI*, *IP*, *SMC-B*, *R&A*, *IFS*) and other journals. He was general chair for the IEEE Conference on Computer Vision and Pattern Recognition, IEEE Workshops on Applications of Computer Vision, IEEE Workshops on Learning in Computer Vision and Pattern Recognition; chair for the DARPA Image Understanding Workshop, and program chair for the IEEE Workshops on Computer Vision Beyond the Visible Spectrum and Multi-Modal Biometrics. He is a fellow of the IEEE, the IEEE Computer Society, American Association for the Advancement of Science (AAAS), International Association of Pattern Recognition (IAPR), and the International Society for Optical Engineering (SPIE).

► For more information on this or any other computing topic, please visit our Digital Library at [www.computer.org/publications/dlib](http://www.computer.org/publications/dlib).