Biometric Recognition Using 3D Ear Shape

Ping Yan and Kevin W. Bowyer, *Fellow*, *IEEE*

Abstract—Previous works have shown that the ear is a promising candidate for biometric identification. However, in prior work, the preprocessing of ear images has had manual steps and algorithms have not necessarily handled problems caused by hair and earrings. We present a complete system for ear biometrics, including automated segmentation of the ear in a profile view image and 3D shape matching for recognition. We evaluated this system with the largest experimental study to date in ear biometrics, achieving a rank-one recognition rate of 97.8 percent for an identification scenario and an equal error rate of 1.2 percent for a verification scenario on a database of 415 subjects and 1,386 total probes.

Index Terms—Biometrics, ear biometrics, 3D shape, skin detection, curvature estimation, active contour, iterative closest point.

1 Introduction

Ear images can be acquired in a similar manner to face images and a number of researchers have suggested that the human ear is unique enough to each individual to allow practical use as a biometric. Several researchers have looked at using features from the ear's appearance in 2D intensity images [6], [16], [5], [27], [17], [10], [11], [23], [31], whereas a smaller number of researchers have looked at using 3D ear shape [8], [4]. Our own previous work that compared ear biometrics using 2D appearance and 3D shape concluded that 3D shape matching allowed greater performance [30]. In another previous work, we compared recognition using 2D intensity images of the ear with recognition using 2D intensity images of the face and suggested that they are comparable in recognition power [6], [27]. Also, ear biometric results can be combined with results from face biometrics. Thus, additional work on ear biometrics has the promise of leading to increased recognition flexibility and power in biometrics.

This paper builds on our previous work to present the first fully automated system for ear biometrics using 3D shape. There are two major parts of the system: automatic ear region segmentation and 3D ear shape matching. Starting with the multimodal 3D + 2D image acquired in a profile view, the system automatically finds the ear pit by using skin detection, curvature estimation, and surface segmentation and classification. After the ear pit is detected, an active contour algorithm using both color and depth information is applied to outline the visible ear region. The outlined shape is cropped from the 3D image and the corresponding 3D data is then used as the ear shape for matching. The matching algorithm achieves a rank-one recognition rate of 97.8 percent on a 415-subject data set in an identification scenario and an equal error rate (EER) of 1.2 percent in a verification scenario.

• *The authors are with the Department of Computer Science and Engineering, University of Notre Dame, 384 Fitzpatrick Hall, Notre Dame, IN 46556. E-mail: {pyan, kwb}@cse.nd.edu.*

Manuscript received 26 Dec. 2005; revised 8 Sept. 2006; accepted 11 Oct. 2006; published online 18 Jan. 2007.
Recommended for acceptance by H. Wechsler.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0735-1205.
Digital Object Identifier no. 10.1109/TPAMI.2007.1067.

This paper is organized as follows: A review of related work is given in Section 2. In Section 3, we describe the experimental method and materials used in our work. Section 4 presents details of the automatic ear segmentation system. Section 5 describes an improved iterative closest point (ICP) approach for 3D ear shape matching. In Section 6, we present the main experimental results, plus additional ear symmetry and off-angle studies. Section 7 gives the summary and conclusions.

2 Literature Review

Perhaps the best known early work on using the ear for identification is that of Iannarelli [18], who developed a manual technique. In his work, over 10,000 ears were examined and no indistinguishable ears were found. The results of this work suggest that the ear may be uniquely distinguishable based on a limited number of features or characteristics. The medical report [18] shows that variation over time is most noticeable during the period from four months to eight years old and over 70 years old. Due to the ear's uniqueness, stability, and predictable changes, ear features are potentially a promising biometric for use in human identification [5], [18], [6], [16], [5], [27], [4].

Moreno et al. [23] experiment with three neural net approaches to recognition from 2D intensity images of the ear. Their testing uses a gallery of 28 people plus another 20 people not in the gallery. They find a recognition rate of 93 percent for the best of the three approaches. They consider three methods (Borda, Bayesian, and weighted Bayesian combination) of combining results of the different approaches but do not find improved performance over the best individual method.

An "eigen-ear" approach on 2D intensity images for ear biometrics has been explored by Victor et al. [27] and Chang et al. [6]. The two studies obtained different results when compared with the performance of facial biometrics. The ear and the face showed similar performance in Chang's study, whereas ear performance is worse than the face in Victor's study. Chang suggested that the difference might be due to the differing ear image quality in the two studies.

Yuizono et al. [31] implemented a recognition system for 2D intensity images of the ear using genetic search. In their experiments, they had 660 images from 110 people with six

TABLE 1
Recent Ear Recognition Studies

Reference	Data Used	Dataset Size	Time Lapse	Number of G/P*	Method Applied	Earrings /Occlu.	Reported Perfor.
Chen & Bhanu [8]	3D	30 × 2	Same Day	1/1	ICP	No	93%
Hurley & Nixon [17]	2D	63 × 4	5 Month	1/1	PCA	No	99.2%
Moreno et al. [23]	2D	28 × 6	Different Days	1/1	Neural Net	Not mentioned	93 %
Yuizono et al. [31]	2D	110 × 6	Same Day	3/3	Genetic search	Not mentioned	100%
Victor & Chang [6]	2D	88 × 2	15 Month	1/1	PCA	No	73%
Choraś [10]	2D	N/A	Same Day	N/A	Feature-based	No	100%

*G = Gallery *and* P = Probe.

images per person. The images were selected from a video stream. The first three of these are used as gallery images and the last three are probe images. They reported that the recognition rate for the registered people was approximately 100 percent and the rejection rate for unknown people was 100 percent.

Bhanu and Chen [4] presented a 3D ear recognition method using a local surface shape descriptor. Twenty range images from 10 individuals are used in the experiments and a 100 percent recognition rate is reported. In [8], Chen and Bhanu used a two-step ICP algorithm on a data set of 30 subjects with 3D ear images. They reported that this method yielded two incorrect matches out of 30 people. In these two works, the ears are manually extracted from profile images. They also presented an ear detection method in [7]. In the offline step, they built an ear model template from each of 20 subjects using the average histogram of the shape index [21]. In the online step, first, they used step edge detection and thresholding to find the sharp edge around the ear boundary and then applied dilation on the edge image and connected-component labeling to search for ear region candidates. Each potential ear region is a rectangular box, and it grows in four directions to find the minimum distance to the model template. The region with minimum distance to the model template is the ear region. They get 91.5 percent correct detection with a 2.5 percent false alarm rate. No recognition results are reported based on this detection method.

Hurley et al. [16] developed a novel feature extraction technique using force field transformation. Each image is represented by a compact characteristic vector which is invariant to initialization, scale, rotation, and noise. The experiment displays the robustness of the technique to extract the 2D ear. Their extended research applies the force field technique to ear biometrics [17]. In the experiments, they used 252 images from 63 subjects with four images per person collected during four sessions over a five month period; any subject is excluded if the ear is covered by hair. A classification rate of 99.2 percent is claimed on this 63-person data set. The data set comes from the XM2VTS face image database [22].

Choras [10], [11] introduces an ear recognition method based on geometric feature extraction from 2D images of the ear. The geometric features are computed from the edge-detected intensity image. They claim that error-free recognition is obtained on "easy" images from their database. The "easy" images are images of high quality with no earring and hair covering and without illumination changes. No detailed experimental setup is reported.

Pun and Moon [25] surveyed the literature on ear biometrics up to that point in time. They summarized elements of five approaches for which experimental results have been published [6], [16], [4], [5], [31]. In Table 1, we compare different aspects of these and other published works.

We previously looked at various methods of 2D and 3D ear recognition and found that an approach based on 3D shape matching gave the best performance. The detailed description of the comparison of different 2D and 3D methods can be found in [29]. This work found that an ICP-based approach statistically and significantly outperformed the other approaches considered for 3D ear recognition and also statistically and significantly outperformed the 2D "eigen-ear"

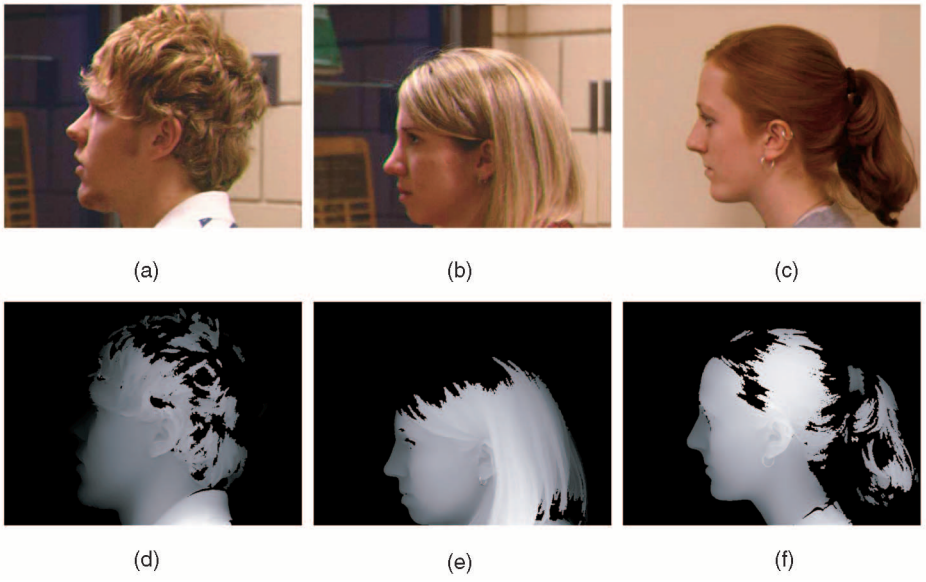


Fig. 1. Sample images used in the experiments. (a) Two-dimensional image. (b) Minor hair covering. (c) Presence of earring. (d) Three-dimensional depth image of (a). (e) Three-dimensional depth image of (b). (f) Three-dimensional depth image of (c).

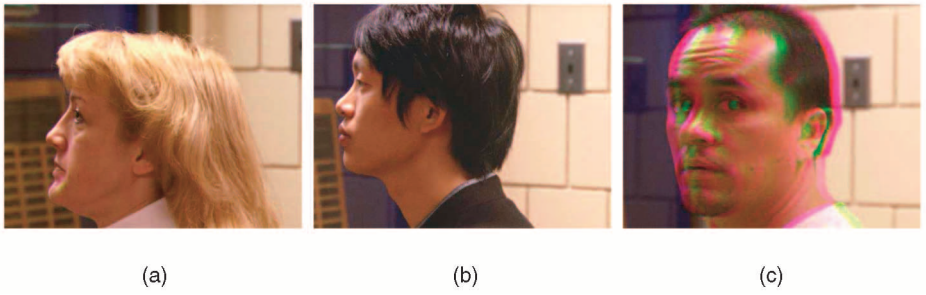


Fig. 2. Examples of images discarded for quality control reasons. (a) Hair-covered ear. (b) Hair-covered ear. (c) Subject motion.

result [6]. Approaches that rely on the 2D intensity image alone can only take into account pose change in the image plane in trying to align the probe image to the gallery image. Approaches that take the 3D shape into account can account for more general pose change. Based on our previous work, an ICP-based approach for 3D ear shape is used as the matching algorithm in this current study.

Of the publications reviewed here, only two [8], [4] deal with biometrics based on 3D ear shape. The largest data set for 2D or 3D studies, in terms of number of people, is 110 [31]. The presence or absence of earrings is not mentioned, except for [30] and [6] in which earrings are excluded.

Comparing with the publications reviewed above, the work presented in this paper is unique in several aspects. We report results for the largest ear biometrics study to date in terms of number of people, which is 415, and in terms of number of images, which is 1,801. Our work is able to deal with the presence of earrings and with a limited amount of occlusion by hair. Ours is the only work to fully automatically detect the ear from a profile view and segment the ear from the surroundings.

3 Experimental Methods and Materials

In each acquisition session, the subject sat approximately 1.5 meters away from the sensor with the sensor looking at the left side of the face. Data was acquired with a Minolta Vivid 910 range scanner. One 640×480 3D scan and one 640×480 color image were obtained in a period of several seconds. Examples of the raw data are shown in Figs. 1a and 1d. The Minolta Vivid 910 is a general-purpose 3D sensor, which is not specialized for application in face or ear biometrics.

From 497 people that participated in two or more image acquisition sessions, there were 415 who had good-quality 2D and 3D ear images in two or more sessions. Among them, there are 237 males and 178 females. There are 70 people who wore earrings at least once and 40 people who have minor hair covering around the ear. This data is not a part of the Face Recognition Grand Challenge (FRGC) data set (http://face.nist.gov/frgc/), which contains frontal face images rather than profile images.

No special instructions were given to the participants to make the ear images particularly suitable for this study and, as a result, 455 out of 2,256 images were dropped for various quality control reasons: 381 instances with hair obscuring the ear and 74 cases with artifacts due to motion during the scan. See Fig. 2 for examples of these problems. Using the Minolta scanner in the high-resolution mode that we used may make the motion artifact problem more frequent as it takes 8 seconds to complete a scan.

The earliest good image for each of the 415 people was enrolled to create the gallery for the experiments. The gallery is the set of images that a "probe" image is matched

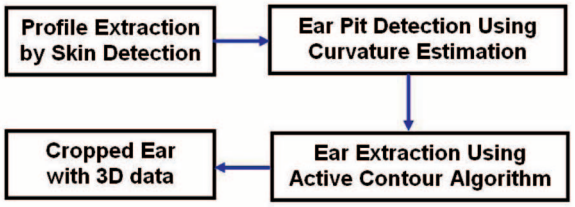


Fig. 3. Data flow of automatic ear extraction.

against for identification. The later good images of each person were used as probes. This results in an average of 17.7 weeks time lapse between the gallery and probe images used in our experiments.

4 Segmenting the Ear Region from a Profile Image

Automatically, extracting the ear region from a profile image is a key step in making a practical ear biometric system. In order to locate the ear in the profile image, we need to have a robust feature extraction algorithm which is able to handle variation in ear location in the profile images. After we find the location of the ear, segmenting the ear from the surrounding is also important. Any extra surface region around the ear could affect the recognition performance. In our system, an active contour approach [20], [13], [28] is used for segmenting the ear region.

Initial empirical studies demonstrated that the ear pit is a good stable candidate as a starting point for an active contour algorithm. When there is so much of the ear covered by hair that the pit is not visible, the segmentation will not be able to be initialized. But, in such cases, there is not enough ear shape visible to support reliable matching anyway. From the profile image, we use skin detection, curvature estimation, and surface segmentation and classification to find the ear pit automatically. Fig. 3 presents the steps that are involved in accomplishing the automatic ear extraction.

4.1 Ear Pit Detection

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

4.1.1 Preprocessing

We start with the binary image of valid depth values to find an approximate position of the nose tip. Given the depth values of a profile image, the face contour can be easily detected. An example of the depth image is shown in Fig. 4b. A valid point has an (x, y, z) value reported by the sensor and is shown as white in the binary image in Fig. 4c.

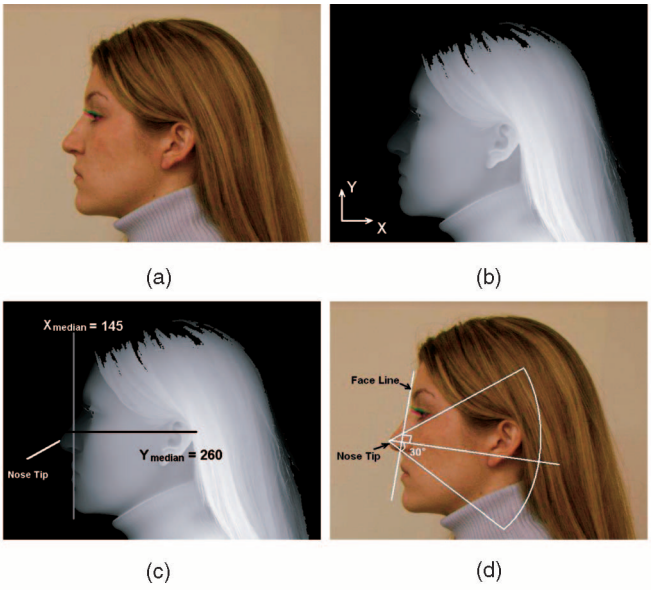


Fig. 4. Using the nose tip as the center to generate a circle sector. (a) Original 2D color image. (b) Depth image. (c) Nose tip location. (d) Circle sector.

We find the X value along each row at which we first encounter a white pixel in the binary image, as shown in Fig. 4c. Using the median of the starting X values for each row, we find the approximate X value of the face contour. Within a 5 cm range of X_{median}, the median value of the Y values for each row is at an approximate Y position of the nose tip. Within a 6 cm range of the Y_{median}, the valid point with the minimum X value is the possible nose tip.

Then, we fit a line along the face profile. Using the point $\mathrm{P}(X_{NoseTip}, Y_{NoseTip})$ as the center of a circle, we generate a sector spanning $+/-30$ degrees perpendicular to the face line with a radius of 15 cm. One example is presented in Fig. 4d. Sometimes, the possible nose tip might be located on the chin or mouth, but, in those situations, the ear still appears in the defined sector.

4.1.2 Skin Region Detection

Skin detection is computationally faster than the surface curvature computation and, so, we use skin detection to reduce the overall computational time. A skin detection method is applied to isolate the face and ear region from the hair and clothes as much as possible (Fig. 5). We do not expect

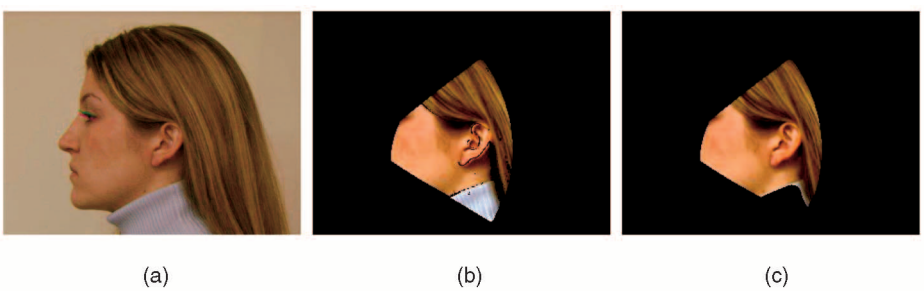


Fig. 5. Ear region with skin detection. (a) Original 2D color image. (b) After preprocessing. (c) After skin detection.

that the hair and clothes are fully removed. Our skin detection method is based on the work of Hsu et al. [15]. The major obstacle to using color to detect the skin region is that the appearance of skin-tone color can be affected by lighting. In their work, a lighting compensation technique is introduced to normalize the color appearance. In order to reduce the dependence of skin-tone color on luminance, a nonlinear transformation is applied to the luma, blue, and red chroma (YCbCr) color space. A parametric ellipse in the color space is then used as a model of skin color, as described in [15].

4.1.3 Surface Curvature Estimation

This section describes a method that can correctly detect the ear pit from the region obtained by previous steps. We know that the ear pit shows up in the 3D image as a "pit" in the surface curvature classification system [3], [14]. Flynn and Jain [14] evaluated five curvature estimation methods and classified them into analytic estimation and discrete estimation. The analytic estimation first fits a local surface around a point and then uses the parameters of the surface equation to determine the curvature value. Instead of fitting a surface, the discrete approach estimates either the curvature or the derivatives of the surface numerically. We use an analytic estimation approach with a local coordinate system determined by principal component analysis [14], [26].

In practice, the curvature estimation is sensitive to noise. For stable curvature measurement, we would like to smooth the surface without losing the ear pit feature. Since our goal at this step is only to find the ear pit, it is acceptable to smooth out other more finely detailed curvature information. Gaussian smoothing is applied on the data with an 11×11 window size. In addition, "spike" data points in 3D are dropped. A "spike" occurs when an angle between the optical axis and a surface normal of observed points is greater than a threshold. (Here, we set the threshold as 90 degrees.) Then, for the (x, y, z) points within a 21×21 window around a given point P, we establish a local X, Y, Z coordinate system defined by principal component analysis (PCA) on the points in the window [14]. Using this local coordinate system, a quadratic surface is fit to the (smoothed, despiked) points in the window. Once the coefficients of the quadratic form are obtained, their derivatives are used to estimate the Gaussian curvature, K, and mean curvature, H, for that point.

4.1.4 Surface Segmentation and Classification

The surface type at each point is labeled based on H and K. Points are grouped into regions with the same curvature label. In our experience, segmentation of the ear image by the sign of H and K is straightforward and the ear pit can always be found in the ear region if it is not covered by hair or clothes.

After segmentation, we expect that there is a pit region, defined as $K > 0$ and $H > 0$, in the segmented image that corresponds to the actual ear pit. Due to numerical error and the sensitivity of curvature estimation, thresholds are required for H and K. Empirical evaluation showed that $T_K = 0.0009$ and $T_H = 0.00005$ provide good results. Fig. 6c shows an example of the face profile with curvature estimation and surface segmentation. Also, we find that the jawline close to the ear always appears as a wide valley region ($K \leq 0$ and $H > 0$) and is located to the left of the ear pit region.

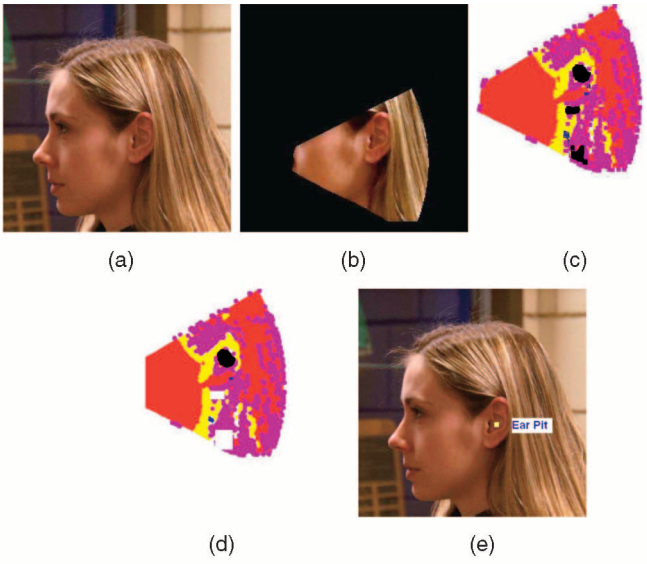


Fig. 6. Steps of finding the ear pit: (a) 2D or 3D raw data, (b) skin detection, (c) curvature estimation, (d) surface curvature segmentation, and (e) region classification, ear pit detection. In (c) and (d), black represents pit region, yellow represents wide valley, magenta represents peak, and red represents ridge, wide peak, and saddle ridge.

It is possible that there are multiple pit regions in the image, especially in the hair around the ear. A systematic voting method is developed to select the pit region that corresponds to the ear pit. Three types of information contribute to the final decision: the size of the pit region, the size of the wide valley region around the pit, and how close the ear pit region is to the wide valley. Each category is given a score in the range of 0 to 10, calculated as the fraction of max area or distance at a scale of 10. For example, the largest pit region P_1 in the image has a score of 10 and the score of any other pit region P_2 is calculated as $Area(P_2)/Area(P_1) \times 10$. The pit with the highest average score is assumed to be the ear pit. In order to validate our automatic ear extraction system, we compare the results ($X_{Auto\_Ear\_Pit}$, $Y_{Auto\_Ear\_Pit}$) with the manually marked ear pit ($X_{Manual\_Ear\_Pit}$, $Y_{Manual\_Ear\_Pit}$) for the 1,801 images used in this study. The maximum distance difference between ($X_{Auto\_Ear\_Pit}$, $Y_{Auto\_Ear\_Pit}$) and ($X_{Manual\_Ear\_Pit}$,$Y_{Manual\_Ear\_Pit}$) is 29 pixels. There are slightly different results from the active contour algorithm when using automatic ear pit finding versus manual ear pit marking. But, the difference does not cause problems for the active contour algorithm finding the ear region, at least on any of the 1,801 images considered here. Using a manual marking of the center of the ear pit rather than the automatically found center of the ear pit results in a minimal difference in rank-one recognition rate, 97.9 to 97.8 percent. Fig. 7 illustrates that, as long as the starting point is near the ear pit, the active contour algorithm can find a reasonable segmentation of the ear region, which is useful for recognition.

Our experiments used several parameters obtained from empirical results. Ear pit finding can be more complicated when great pose variation is involved. Therefore, further study combining ear features should result in more robust results.

(a) (b)

Fig. 7. Varying ear pit location versus segmentation results. (a) Ear pit (automatically found). (b) Ear pit (manually found).

4.2 Ear Segmentation Using Active Contour Algorithm

The 3D shape matching of the ear relies upon correct and accurate segmentation of the ear. Several factors contribute to the complexity of segmenting the ear out of the image. First, ear size and shape vary widely between different people. Second, there is often hair touching or partially obscuring the ear. Third, if earrings are present, they overlap or touch the ear but should not be treated as a part of the ear shape. These characteristics make it hard to use a fixed template to crop the ear shape from the image (as in, for example, [6]). A bigger template will include too much hair, whereas a smaller template may lose shape information. Also, it is hard to distinguish the ear from hair or earrings, especially when hair and earrings have a similar color to the skin or are very close to the ear.

Edges are usually defined as large magnitude changes in image gradient. We wish to find edges that indicate the boundary of the visible ear region. The classical active contour function proposed by Kass et al. [20] is used to grow from the ear pit to the outline of the visible ear region. Thus, we have

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

$$E_{int} = \frac{1}{2}\left[\alpha|X'(s)|^2 + \beta|X''(s)|^2\right], \tag{2}$$

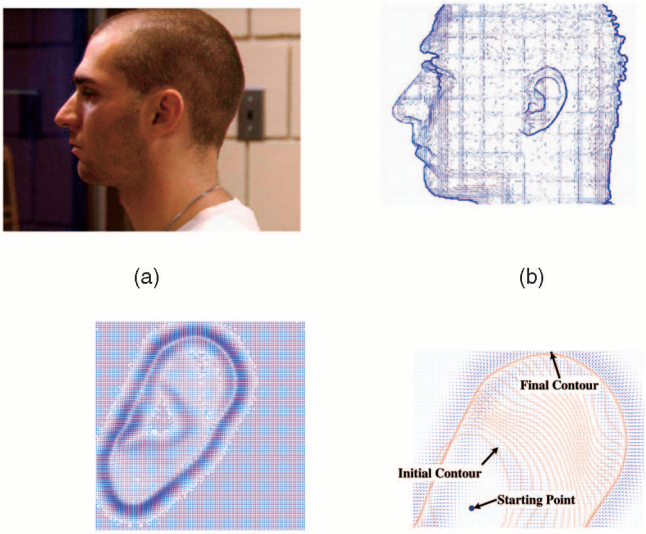
$$E_{ext} = E_{image} + E_{con}, \tag{3}$$

$$E_{image} = \nabla Image(x, y), \tag{4}$$

$$E_{con} = -w_{con}\ \vec{n}\ (s). \tag{5}$$

The contour X(s) starts from a closed curve within the region and then grows under internal and external constraints to move the curve toward local features (1). Following the description in [20], $X'(s)$ and $X''(s)$ denote the first and second derivative of the curve X(s). α and β are weighting parameters for measuring the contour tension and rigidity, respectively. The internal function E_{int} restrains the curve from stretching or bending. The external function E_{ext} is derived from the image so that it can drive the curve to areas with high image gradient and lock on to close edges. It includes E_{image} and E_{con}. E_{image} is image energy, which is used to drive the curve to salient image features such as lines, edges, and terminations. In our case, we use edge feature as E_{image}.

The traditional active contour algorithm suffers from instability due to image force. When the initial curve is far away from image features, the curve is not attracted by the E_{image} and would shrink into a point or a line, depending on the initial curve shape. Cohen [12] proposed a "balloon" model to give more stable results. The "pressure force" E_{con} (5) is introduced and it pushes the curve outward so that it does not shrink to a point or a line. Here, $\vec{n}(s_i)(x,y) = \frac{s_{i-1}(x,y) - s_{i+1}(x,y)}{Distance(s_{i-1}, s_{i+1})}$, s_i is the point i on curve s. Fig. 8 shows how the active contour algorithm grows toward the outline of the ear region.



(a) (b) (c) (d)

Fig. 8. Active contour growing on ear image. (a) Original image. (b) Energy map of (a). (c) Energy map of ear. (d) Active contours growing.

Starting with the ear pit determined in the previous step, the active contour grows until it finds the ear edge. Usually, there is either depth or color change, or both, along the ear edge. These attract the active contour to grow toward and stop at the ear boundary.

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

Fig. 9 shows examples when only color or depth information is used for the active contour algorithm. When there is no clear color or depth change along the ear edge, it is hard for the algorithm to stop expanding. As shown in Figs. 9a and 9b, by using 2D alone or 3D alone, the active contour can easily keep growing after it reaches the boundary of the ear. We ran the active contour algorithm using color or depth alone on the 415 gallery images. Using only color information, 88 out of 415 (21 percent) images are incorrectly segmented. Using only depth information, 60 out of 415 (15 percent) images are incorrectly segmented. All of the incorrectly segmented images in these two situations can be correctly segmented by using the combination of color and depth information. These examples in Fig. 9 imply that, in order to improve the robustness of the algorithm, we need to combine both the color and 3D information in the active contour algorithm. To

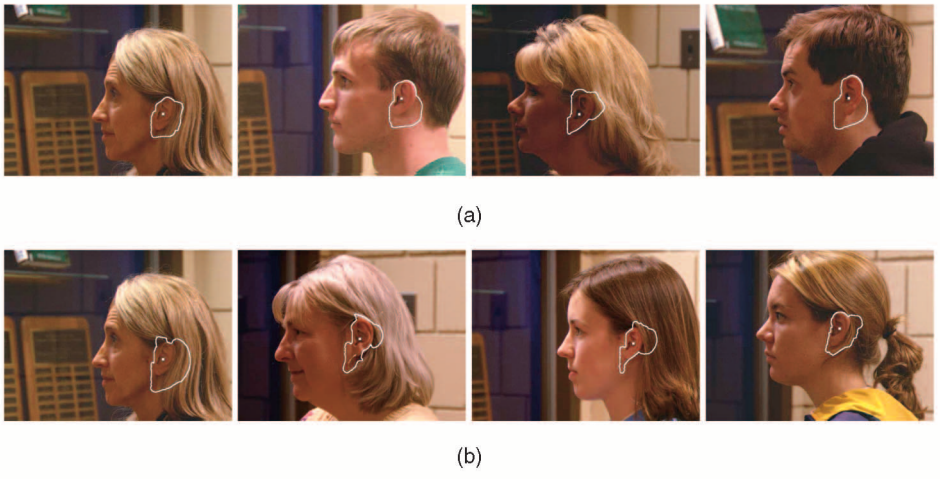


Fig. 9. Active contour results using only color or depth information. (a) Only using color (incorrect segmentation). (b) Only using depth (incorrect segmentation).

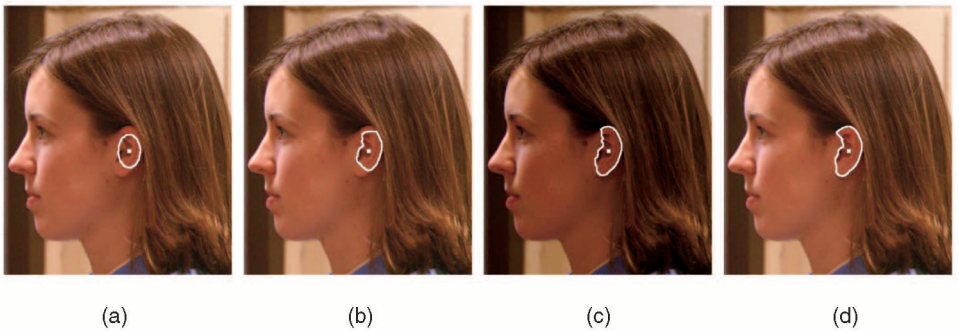


Fig. 10. Active contour growing on a real image. (a) $\mathrm{Iteration} = 0$. (b) $\mathrm{Iteration} = 25$. (c) $\mathrm{Iteration} = 75$. (d) $\mathrm{Iteration} = 150$.

do this, the E_{image} in (3) is replaced by (6). Consequently, the final energy E is represented by (7):

$$E_{Image} = w_{depth}\nabla Image_{depth}(x,y) + w_{Cr}\nabla Image_{Cr}(x,y), \quad (6)$$

$$\begin{aligned} E = \int_0^1 \frac{1}{2}\left[\alpha|X'(s)|^2 + \beta|X''(s)|^2\right] \\ + w_{depth}\nabla Image_{depth}(x,y) + w_{Cr}\nabla Image_{Cr}(x,y) \\ - w_{con}\ \vec{n}\ (s). \end{aligned} \quad (7)$$

In order to prevent the active contour from continuing to grow toward the face, we modify the internal energy of points to limit the expansion when there is no depth jump within a 3×5 window around the given point. The threshold for the maximum gradient within the window is set as 0.01. With these improvements, the active contour algorithm works effectively in separating the ear from the hair and earrings and the active contour stops at the jawline close to the ear.

The initial contour is an ellipse with the ear pit as center. Approximately, the major axis is 15 mm and the minor axis is 10 mm and the major axis is vertical. Fig. 10 illustrates the steps of active contour growing for a real image. Fig. 11 shows examples in which the active contour deals with hair and earrings. The 3D shape within the final contour is cropped out of the image for use in the matching algorithm.

5 Matching 3D Ear Shape for Recognition

We have previously compared using an ICP approach on a point-cloud representation of the 3D data and a PCA-style approach on a range-image representation of the 3D data [29] and found better performance using an ICP approach on the point-could representation. The problem with using a range image representation of the 3D data is that landmark points must be selected ahead of time to use for normalizing the pose and creating the range image. Errors or noise in this process can lead to recognition errors in the PCA or other algorithms that use the range image. Our experience is that the ICP style approach using the point cloud representation can better adapt to inexactness in the initial registration, though, of course, at the cost of some increase in the computation time for the matching step.

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

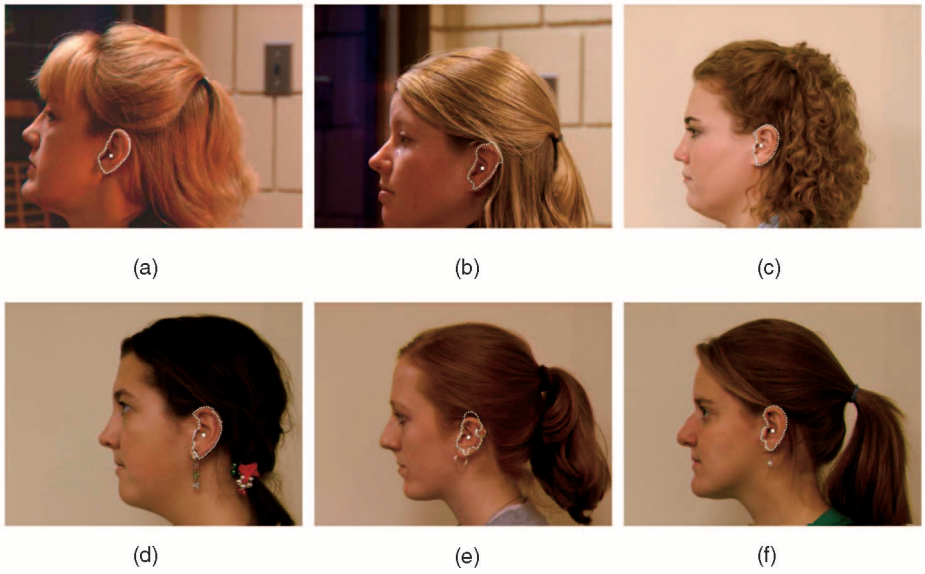


Fig. 11. Active contour algorithm dealing with earring and blonde hair. (a) Earring and blonde hair. (b) Blonde hair. (c) Earring and blonde hair. (d) Earring. (e) Earring and blonde hair. (f) Earring and blonde hair.

5.1 Computation Time Reduction

It is well known that the basic ICP algorithm can be time consuming. In order to make it more practical for use in biometric recognition, we use a k-d tree data structure in the search for closest points, limit the maximum number of iterations to 40, and stop if the improvement in mean square difference between iterations drops below 0.001. This allows a probe shape to be matched against a gallery of 415 ear shapes in 10 minutes or better than 40 shape matches per minute. This is with an average of 6,000 points in a gallery image and 1,400 in a probe image. The ICP algorithm is implemented in C++ based on the VTK 4.4 library [1] and run on a dual-processor 2.8-GHz Pentium Xeon system. The current computation speed is obviously more than sufficient for a verification scenario in which a probe is matched against a claimed identity. It is also sufficient for an identification scenario involving a few tens of subjects.

5.2 Recognition Performance Improvement

Ideally, if two scans come from the same ear with the same pose, the error distance should be close to zero. However, with pose variation and scanning error, the registration results can be greatly affected by data quality. Our approach to improve performance focuses on reducing the effect of noise and using a point-to-surface error metric for sparse range data.

5.2.1 Outlier Elimination

The general ICP algorithm requires no extracted features or curvature computation [2]. The only preprocessing of the range data is to remove "spike" outlier points. In a 3D face image, the eyes and mouth are common places for holes and spikes to occur. Three-dimensional ear images do exhibit some spikes and holes due to oily skin or sensor error, but these occur less frequently than in 3D face images.

An "outlier" match occurs when there is a poor match between a point on the probe and a point on the gallery. To improve performance, outlier match elimination is accomplished in two stages. During the calculation of the transformation matrix, the approach is based on the assumption that, for a given noise point p on the probe surface, the distance from p to the associated closest point g_p on the gallery surface will be much larger than the average distance [32], [19]. For each point p on the probe surface, we find the closest point g_p on the gallery surface. Let $D = d(p, g_p)$ represent the distance between the two points. Only those pairs of points whose D is less than a threshold are used to calculate the transformation matrix. Here, the threshold is set as $\text{mean distance} + R * 2$, where R is the resolution of the probe surface.

The second stage occurs outside the transformation matrix calculation loop. After the first step, a transformation matrix is generated to minimize the error metric. We apply this transformation matrix on the source surface S and obtain a new surface S'. Each point on the surface S' will have a distance to the closest point on the target surface. We sort all of the distance values and use only the lower 90 percent to calculate the final mean distance. Other thresholds (99, 95, 85, 80, and 70 percent) were tested and 90 percent gives the best performance, which is consistent with the experiments of other researchers [24].

5.2.2 Point-to-Point versus Point-to-Surface Approach

Two approaches are considered for matching points from the probe to points on the gallery: point-to-point [2] and point-to-surface [9]. In the point-to-point approach, we try to find the closest point on the target surface. In the point-to-surface approach, we use the output from the point-to-point algorithm first. Then, from the closest point obtained earlier on the target surface, all of the triangles around this point are extracted. Then, the real closest point is the point

TABLE 2
ICP Performance by Using Point-to-Surface, Point-to-Point, and Revised Version, and Time Is for One Probe Matched to One Gallery Shape

G = Gallery	Point-To-Surface		Point-To-Point		Mixed: P-To-P and P-To-S	
P = Probe	Performance	Run Time (s)	Performance	Run Time (s)	Performance	Run Time (s)
G1P1	97.3%	15-18 sec	96.6%	5-6 sec	97.6%	5-8 sec
G1P2	96.9%	2-3 sec	96.6%	1-2 sec	97.1%	1-2 sec (*)
G1P4	96.1%	1-2 sec	95.7%	0.3-0.7 sec	95.9%	0.5-0.7 sec
G2P2	96.1%	2-4 sec	95.7%	0.7-1.3 sec	96.4%	0.7-1.3 sec
G2P4	96.4%	0.7-1.3 sec	95.2%	0.2-0.3 sec	96.1%	0.2-0.3 sec
G4P4	93.3%	0.6-1.2 sec	72.8%	0.1-0.2sec	88.4%	0.1-0.2 sec

**Recognition rates and execution times quoted elsewhere in the paper are for the G1, P2 instance of the algorithm using our "mixed" ICP.*

on any of these triangles with the minimum distance to the source point. In general, point-to-surface is slower, but also more accurate in some situations.

As shown in Table 2, the point-to-point approach is fast and accurate when all of the points on the source surface can find a good closest point on the target surface. But, if the gallery is subsampled, the point-to-point approach loses accuracy. Since the probe and gallery ear images are taken on different days, they vary in orientation. When both gallery and probe images are subsampled, it is difficult to match points on the probe surface to corresponding points on the gallery surface. This generally increases the overall mean distance value. But, this approach is much faster than point-to-surface.

On the other hand, the greatest advantage of the point-to-surface approach is that it is accurate through all of the different subsample combinations. Even when the gallery is subsampled by every four rows and columns, the performance is still acceptable.

Our final algorithm attempts to exploit the trade-off between performance and speed. The point-to-point approach is used during the iterations to compute the transformation matrix. One more point-to-surface iteration is done after obtaining the transformation matrix to compute the error distance. This revised algorithm works well due to the good quality of the gallery images, which makes it possible for the probe images to find the corresponding points. As a biometrics application and especially in a verification scenario, we can assume that the gallery image is always of good quality and the ear orientation exposes the most part of ear region. The final results reflecting the revised algorithm are shown in Table 2.

Table 2 leads to two conclusions: The first is that, when the gallery and probe surfaces have similar resolution, the mixed algorithm is always more accurate than pure point-to-point matching and has similar computation time. The second is that, when the gallery surface is more densely sampled than the probe surface, the mixed algorithm is both faster and more accurate than point-to-surface ICP.

6 Experimental Results

In an identification scenario, our algorithm achieves a rank-one recognition rate of 97.8 percent on our 415-subject data set with 1,386 probes. The cumulative match characteristic (CMC) curve is shown in Fig. 12a. In a verification scenario, our algorithm achieves an EER of 1.2 percent. The receiver operating characteristic (ROC) curve is shown in Fig. 12b. This is an excellent performance in comparison to previous

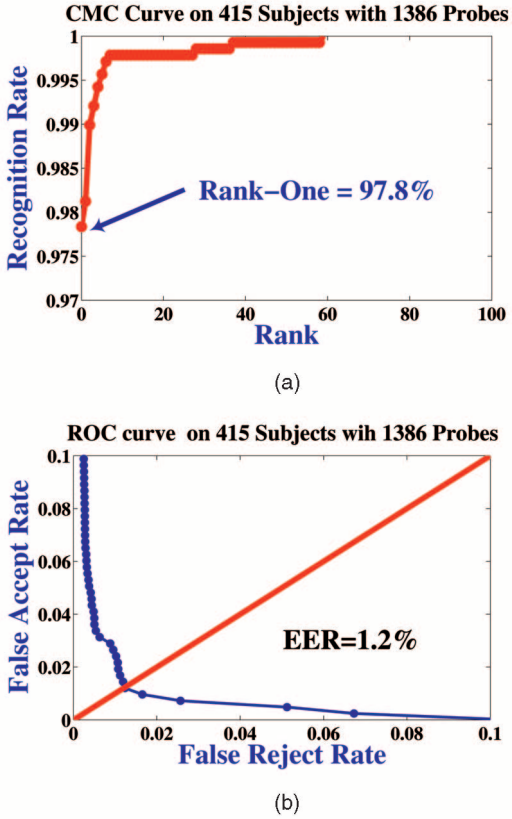


Fig. 12. The performance of ear recognition. (a) CMC curve. (b) ROC curve.

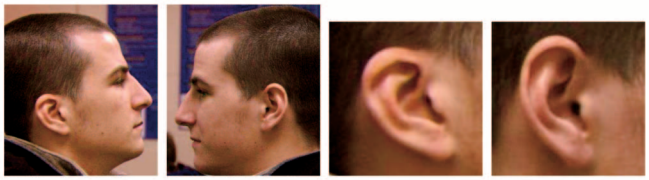


Fig. 13. Examples of asymmetric ears. (a) Right ear. (b) Left ear. (c) Right ear. (d) Mirrored left ear.

work in ear biometrics; where higher performance values were reported, they were for much smaller data sets.

Also, the rank-one recognition is 95.7 percent (67 out of 70) for the 70 cases that involve earrings. This is a difference of just one of the 70 earring probes from the rank-one recognition rate for probes without earrings. Thus, the presence of earrings in the image causes only a minimal loss in accuracy.

Chang et al. [6] obtained a 73 percent rank-one recognition rate for an "eigen-ear" approach on 2D intensity images with 88 people in the gallery and a single time-lapse probe image per person. Our rank-one recognition rate for PCA-based ear recognition using 2D intensity images for the first 88 people in our 415 person data set is 76.1 percent, which is similar to the result obtained by Chang et al., even though we used a completely different image data set acquired by a different sensor and used different landmark points. For the same 88 people, our ICP-based ear recognition gave a 98.9 percent rank-one recognition rate.

6.1 Ear Symmetry Experiment

The ear data used in our experiments in previous sections are gallery and probe images that are approximately straight-on views of the same ear which were acquired on different days. One interesting question to explore is the use of bilateral symmetry; for example, matching a mirrored left ear to a right ear. This means that, for one subject, we enroll his right ear and try to recognize using his mirrored left ear. One example is shown in Figs. 13a and 13b. For our initial experiment to investigate this possibility, both ear images were taken on the same day. The rank-one recognition rates from matching a mirrored image of an ear are around 90 percent on a 119 subject data set [30]. By analyzing the results, we found that most people's left and right ears are approximately bilaterally symmetric. But, some people's left and right ears have recognizably different shapes. Fig. 13 shows an example of this. Thus, it seems that symmetry-based ear recognition cannot be expected to be as accurate, in general, as matching two images of the same ear.

TABLE 3
Results of Off-Angle Experiments with a 24-Subject Data Set

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

6.2 Off-Angle Experiment

Another dimension of variability is the degree of pose change between the enrolled gallery ear and the probe ear. To explore this, we enroll a right ear that was viewed straight on and try to recognize a right ear viewed at some amount of angle. In this experiment, there are four different angles of view for each ear: straight-on, 15 degrees off center, 30 degrees off center, and 45 degrees off center, as shown in Fig. 14. The 45 degree images were taken on the first week. The 30 degree images were taken the second week. Finally, the 15 degree and straight-on images were both taken on the third week. For each angle of ear image, we match it against all images in the different angle data sets.

Twenty-four subjects participated in this set of image acquisitions. Two observations are drawn from Table 3. The first is that 15 and 30 degrees off center have better overall performance than the straight-on and 45 degrees off center. This observation makes sense since there is more ear area exposed to the camera when the face is 15 and 30 degrees off center. Also, matching is generally good for 15 degrees difference, but gets worse for more than 15 degrees. This is an initial experiment and additional work with a larger data set is still needed.

7 Summary and Discussion

We have presented a fully automatic ear biometric system using 2D and 3D information. The automatic ear extraction algorithm can crop the ear region from the profile image,

Fig. 14. Example images acquired for off-angle experiments. (a) Straight-on. (b) Fifteen degrees off. (c) Thirty degrees off. (d) Forty-five degrees off.

separating the ear from hair and earring. The recognition subsystem uses an ICP-based approach for 3D shape matching. The experimental results demonstrate the power of our automatic ear extraction algorithm and 3D shape matching applied to biometric identification. The system has a 97.8 percent rank-one recognition rate and a 1.2 percent EER on a time-lapse data set of 415 persons with 1,386 probe images.

The system as outlined in this paper is a significant and important step beyond existing work in ear biometrics. It is fully automatic, handling preprocessing, cropping, and matching. The system addresses issues that plagued earlier attempts to use 3D ear images for recognition, specifically partial occlusion of the ear by hair and earrings.

There are several directions for future work. We presented techniques for extracting the ear image from hair and earrings, but there is currently no information on whether the system is robust when subjects wear eyeglasses. We intend to examine whether eyeglasses can cause a shape variation in the ear and whether this will affect the algorithm. Additionally, we are interested in further quantifying the effect of pose on ICP matching results. Further study should result in guidelines that provide best practices for the use of 3D images for biometric identification in production systems. Also, speed and recognition accuracy remain important issues. We have proposed several enhancements to improve the speed of the algorithm, but the algorithm might benefit from adding feature classifiers. We have both 2D and 3D data and they are registered with each other, which should make it straightforward to test multimodal algorithms.

The 2D and 3D image data sets used in this work are available to other research groups. See the Web page at www.nd.edu/~cvrl for the release agreement and details.

ACKNOWLEDGMENTS

Biometrics research at the University of Notre Dame is supported by the US National Science Foundation under Grant CNS01-30839, by the Central Intelligence Agency, by the US Department of Justice/National Institute for Justice under Grants 2005-DD-CX-K078 and 2006-IJ-CX-K041, by the National Geo-Spatial Intelligence Agency, and by UNISYS Corp. The authors would like to thank Patrick Flynn and Jonathon Phillips for useful discussions about this work. The authors would also like to thank the anonymous reviewers for providing useful feedback. These comments were important in improving the clarity and presentation of the research.

REFERENCES

[1] http://www.vtk.org, year?
[2] P. Besl and N. McKay, "A Method for Registration of 3-D Shapes," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 14, pp. 239-256, 1992.
[3] P.J. Besl and R.C. Jain, "Invariant Surface Characteristics for 3D Object Recognition in Range Images," *Computer Vision Graphics Image Processing,* vol. 33, pp. 30-80, 1986.
[4] B. Bhanu and H. Chen, "Human Ear Recognition in 3D," *Proc. Workshop Multimodal User Authentication,* pp. 91-98, 2003.
[5] M. Burge and W. Burger, "Ear Biometrics in Computer Vision," *Proc. 15th Int'l Conf. Pattern Recognition,* vol. 2, pp. 822-826, 2000.
[6] K. Chang, K. Bowyer, and V. Barnabas, "Comparison and Combination of Ear and Face Images in Appearance-Based Biometrics," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, pp. 1160-1165, 2003.
[7] H. Chen and B. Bhanu, "Human Ear Detection from Side Face Range Images," *Proc. Int'l Conf. Image Processing,* pp. 574-577, 2004.
[8] H. Chen and B. Bhanu, "Contour Matching for 3D Ear Recognition," *Proc. Seventh IEEE Workshop Application of Computer Vision,* pp. 123-128, 2005.
[9] Y. Chen and G. Medioni, "Object Modeling by Registration of Multiple Range Images," *Image and Vision Computing,* vol. 10, pp. 145-155, 1992.
[10] M. Choras, "Ear Biometrics Based on Geometrical Feature Extraction," *Electronic Letters on Computer Vision and Image Analysis,* vol. 5, pp. 84-95, 2005.
[11] M. Choras, "Further Developments in Geometrical Algorithms for Ear Biometrics," *Proc. Fourth Int'l Conf. Articulated Motion and Deformable Objects,* pp. 58-67, 2006.
[12] L.D. Cohen, "On Active Contour Models and Balloons," *Computer Vision, Graphics, and Image Processing. Image Understanding,* vol. 53, no. 2, pp. 211-218, 1991.
[13] D. Cremers, "Statistical Shape Knowledge in Variational Image Segmentation," PhD dissertation, Dept. of Math. and Computer Science, Univ. of Mannheim, Germany, July 2002.
[14] P. Flynn and A. Jain, "Surface Classification: Hypothesis Testing and Parameter Estimation," *Proc. IEEE Conf. Computer Vision Pattern Recognition,* pp. 261-267, 1988.
[15] R.-L. Hsu, M. Abdel-Mottaleb, and A. Jain, "Face Detection in Color Images," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, pp. 696-706, 2002.
[16] D. Hurley, M. Nixon, and J. Carter, "Force Field Energy Functionals for Image Feature Extraction," *Image and Vision Computing J.,* vol. 20, pp. 429-432, 2002.
[17] D. Hurley, M. Nixon, and J. Carter, "Force Field Energy Functionals for Ear Biometrics," *Computer Vision and Image Understanding,* vol. 98, pp. 491-512, 2005.
[18] A. Iannarelli, *Ear Identification.* Paramont Publishing, 1989.
[19] A.E. Johnson http://www-2.cs.cmu.edu/vmr/software/meshtoolbox, year?
[20] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *Int'l J. Computer Vision,* vol. 1, pp. 321-331, 1987.
[21] J. Koenderink and A. van Doorn, "Surface Shape and Curvature Scales," *Image and Vision Computing,* vol. 10, pp. 557-565, 1992.
[22] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: The Extended M2VTS Database," *Audio and Video-Based Biometric Person Authentication,* pp. 72-77, 1999.
[23] B. Moreno, A. Sanchez, and J. Velez, "On the Use of Outer Ear Images for Personal Identification in Security Applications," *Proc. IEEE Int'l Carnaham Conf. Security Technology,* pp. 469-476, 1999.
[24] K. Pulli, "Multiview Registration for Large Data Sets," *Proc. Second Int'l Conf. 3-D Imaging and Modeling,* pp. 160-168, Oct. 1999.
[25] K. Pun and Y. Moon, "Recent Advances in Ear Biometrics," *Proc. Sixth Int'l Conf. Automatic Face and Gesture Recognition,* pp. 164-169, May 2004.
[26] H.-Y. Shum, M. Hebert, K. Ikeuchi, and R. Reddy, "An Integral Approach to Free-Form Object Modeling," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, pp. 1366-1370, 1997.
[27] B. Victor, K. Bowyer, and S. Sarkar, "An Evaluation of Face and Ear Biometrics," *Proc. 16th Int'l Conf. Pattern Recognition,* pp. 429-432, 2002.
[28] C. Xu and J. Prince, "Snakes, Shapes, and Gradient Vector Flow," *IEEE Trans. Image Processing,* vol. 7, pp. 359-369, 1998.
[29] P. Yan and K.W. Bowyer, "Ear Biometrics Using 2D and 3D Images," *Proc. 2005 IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR '05)—Workshops,* p. 121, 2005.
[30] P. Yan and K.W. Bowyer, "Empirical Evaluation of Advanced Ear Biometrics," *Proc. 2005 IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR '05)—Workshops,* p. 41, 2005.
[31] T. Yuizono, Y. Wang, K. Satoh, and S. Nakayama, "Study on Individual Recognition for Ear Images by Using Genetic Local Search," *Proc. 2002 Congress Evolutionary Computation,* pp. 237-242, 2002.
[32] Z. Zhang, "Iterative Point Matching for Registration of Freeform Curves and Surfaces," *Int'l J. Computer Vision,* vol. 13, pp. 119-152, 1994.

Ping Yan received the BS (1994) and MS (1999) degrees in computer science from Nanjing University and the PhD degree in computer science and engineering from the University of Notre Dame in 2006. Her research interests include computer vision, image processing, evaluation, and implementation of 2D/3D biometrics and pattern recognition. She is currently a postdoctoral researcher at the University of Notre Dame.

Kevin W. Bowyer currently serves as the chair of the Department of Computer Science and Engineering, University of Notre Dame. His research efforts have concentrated on data mining and biometrics. The Notre Dame Biometrics Research Group has been active as part of the support team for the US government's Face Recognition Grand Challenge program and Iris Challenge Evaluation program. His paper "Face Recognition Technology: Security Versus Privacy," published in *IEEE Technology and Society*, was recognized with a 2005 Award of Excellence from the Society for Technical Communication, Philadelphia Chapter. He is a fellow of the IEEE and a golden core member of the IEEE Computer Society. He has served as editor-in-chief of the *IEEE Transactions on Pattern Analysis and Machine Intelligence* and on the editorial boards of *Computer Vision and Image Understanding*, *Image and Vision Computing Journal*, *Machine Vision and Applications*, *International Journal Pattern Recognition and Artificial Intelligence*, *Pattern Recognition*, *Electronic Letters in Computer Vision and Image Analysis*, and *Journal of Privacy Technology*. He received an Outstanding Undergraduate Teaching Award from the University of South Florida College of Engineering in 1991 and the Teaching Incentive Program Awards in 1994 and 1997.