



ing with 2D and 3D images, the system can automatically find the ear pit from the profile images by using skin detection, curvature estimation and surface segmentation and classification. After the ear pit is correctly detected, an active contour algorithm using both color and depth information is applied to expand the contour to the ear edge.

Our work has found that the snake algorithm is well suited for use in cropping out the ear from profile images. The ear pit makes an ideal starting point for the snake algorithm. The snake algorithm grows until it finds the ear edge, and is remarkably robust in its ability to exclude earrings and occluding hair.

When the snake algorithm finishes, the outlined shape is cropped from the 3D image, and the corresponding 3D data is then used as the ear image for matching. The matching algorithm achieves a rank-one recognition rate of 97.6%. Also, our approach shows a good scalability with size of dataset (see Figure 8).

## 2 Automatic Ear Extraction

Automatic ear extraction is necessary for practical ear biometrics systems. In order to locate the ear from the profile image, we need to exploit feature information about the face. This relies on 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. A refinement of the active contour (“snakes”) approach is used to segment the ear region from surrounding hair and earrings. Initial empirical studies demonstrated that the ear pit is a good and stable candidate for this purpose. 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 this case, there is likely not enough ear shape visible for matching, anyway.

### 2.1 Ear Pit detection

Ear pit detection includes four steps: preprocessing, skin detection, curvature estimation, and surface segmentation and classification. We describe each step in following sections.

#### 2.1.1 Preprocessing

In this step, we use the binary image of valid depth values to find an approximate position of the nose tip. Given the depth values of a profile image, the face contour can be easily detected. An example of the depth image is shown in Figure 1(a). A valid point has valid  $(x, y, z)$  value, and shows as white in Figure 1(b).

Then we find the X value along each row at which we first encounter a white pixel in the binary image 1(b). Using the median of the starting X values for each row, we find the approximate X value of the face contour. Within a 100 pixel range of  $X_{median}$ , the median value of the Y values for each row is an approximate Y position of the nose tip. Within a 120 pixel range of the  $Y_{median}$ , the valid point with minimum x value is the nose tip. The method avoids the possibility of locating hair or chin as the nose tip. Using the point  $P(X_{NoseTip}, Y_{NoseTip})$  as the center of a circle, we generate a sector spanning  $\pm 30$  degree from the horizontal with a radius of 20 cm. Figure 1(b) presents the steps to find the nose tip. Using the real distance for the radius helps to eliminate the effect of scale in the 2D image. One example is presented in Figure 1. With a high degree of confidence, the ear is included within the sector, along with some hair and shoulder. In all of the 830 images, none of the ears are outside the region that we look at, and the  $\pm 30$  degrees should cover people looking up or down by that much.

#### 2.1.2 Skin Region Detection

Next, a skin detection method is used to locate the skin region. This step intends to get rid of some hair and clothes. Using the 2D color image, each pixel is transformed into the YCbCr color space [10]. Together with the preprocessing step, the skin region detection step drops out irrelevant regions, such as the shoulder or hair area, helping to reduce the computation time in later steps.

#### 2.1.3 Surface Segmentation and Classification

This section describes a method that can correctly detect the ear pit from the area obtained by previous steps. A priori we know that the ear pit shows up in the 3D image as a pit in the curvature classification system [11, 12].

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. Our goal is to find the ear pit, and it is acceptable for this step to lose some other detailed curvature information. In the implementation, a Gaussian smoothing is applied on the data with a  $11 \times 11$  window size. In addition, the spikes are dropped when an angle between the optical axis and a surface normal of observed points is greater than a threshold. (Here we set threshold as 90 degrees.) Then for the  $(x, y, z)$  points within a  $21 \times 21$  window around a given point  $P$ , we establish a local X,Y,Z coordinate system defined by PCA for that point  $P$  [12]. Using this local coordinate system, a quadratic surface is fit to the points in the window. Once the coefficients of the quadratic form are obtained, their derivatives are used to estimate the Gaussian curvature and mean curvature.

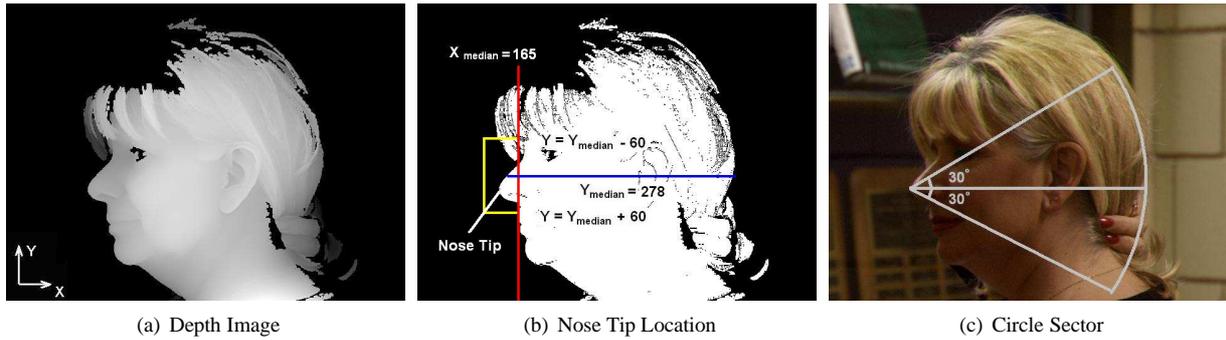


Figure 1. Using Nose Tip as Center to Generate a Sector.

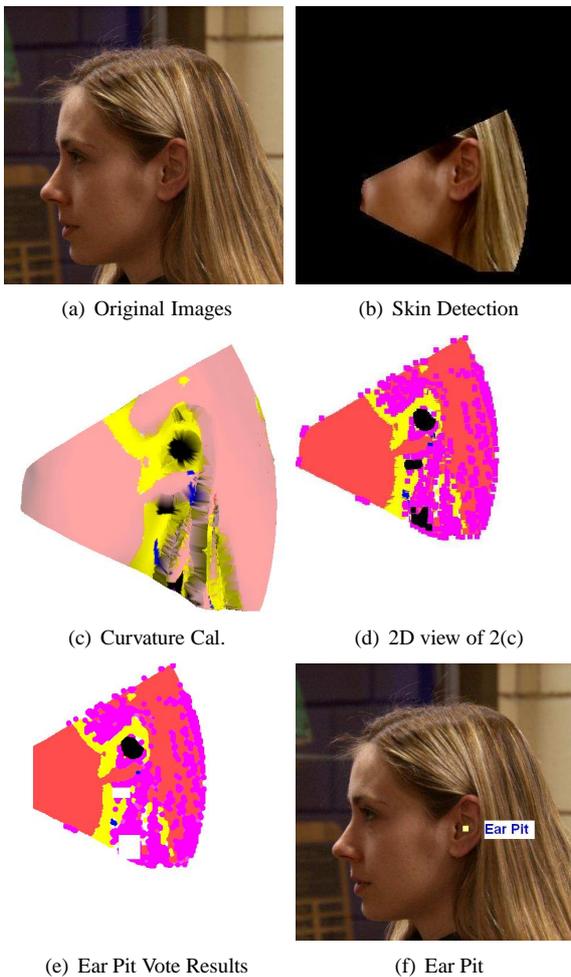


Figure 2. Starting from 2D/3D raw data: skin detection, curvature estimation, surface segmentation, region classification, ear pit detection.

### 2.1.4 Surface Curvature Estimation

The Gaussian ( $K$ ) and mean curvature ( $H$ ) from the estimation step are at a point level. We group 3D points into regions with the same curvature label.

After segmentation, we expect there is a pit region ( $K > 0$  &  $H > 0$ ) in the segmented image that corresponds to the actual ear pit. Due to numerical error and the sensitivity of curvature estimation, thresholds for Gaussian and mean curvature are required. Empirical evaluation showed that  $T_K = 0.0009$  and  $T_H = 0.00005$  provides good results. Figures 2(c) and 2(d) show an example of the face profile with curvature estimation and surface segmentation. Also, we find that the jawline close to the ear always appears as a wide valley ( $K \leq 0$  &  $H > 0$ ), and is located to the left of the ear pit region.

It is possible that there are multiple pit regions in the image, especially in the hair area. A systematic voting method is developed to correctly find the ear pit. Three categories contribute to the final decision: the size of the pit region, the size of the wide valley region around the pit and how close it is to the wide valley. Each category is given a score in the range of 0 to 10, and the score is simply calculated as the ratio of max area or distance at a scale of 10. The pit with the highest score is assumed to be the ear pit. The ear pit is correctly found in Figure 2(f). The experiment shows good results using only simple score calculation. Therefore it demonstrates the advantage of the technique at this scale. After the ear pit is correctly located, the next step is to segment out the unoccluded portion of the ear from the image.

## 2.2 Active Contour Algorithm

Active contours, also called snakes, are an edge-based segmentation approach. Edges are usually defined as large magnitude changes in image gradient, which also indicates locations of intensity discontinuities. These intensity discontinuities are assumed to be in the same location as the geometric discontinuities. Even if the edges are correctly

found by varying edge detection approaches, it is not clear how these edges can be connected to indicate a region of an object in the image. The classical snake function proposed by Kass, Witkin and Terzopoulos [13] is intended to solve this question. The contour  $X(s)$  starts from an explicit parametric closed curve within an image domain, and then grows under the force of both internal and external constraints, and pulls the curve toward local features (Equation 1)

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

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

$$E_{ext} = E_{image} + E_{con} \quad (3)$$

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

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

Following the description in [13, 14],  $X'(s)$  and  $X''(s)$  denote the first and second derivative of the curve  $X(s)$ .  $\alpha$  and  $\beta$  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.  $E_{con}$  represents the external constraints force that usually is a function of  $X(s)$ . Here we used pressure force proposed by Cohen [15], where  $\vec{n}(s_i)(x, y) = \frac{s_{i-1}(x, y) - s_{i+1}(x, y)}{Distance(s_{i-1}, s_{i+1})}$ ,  $s_i$  is the point  $i$  on curve  $s$ . Figure 3 shows how the snake algorithm grows toward the image edge step by step.

### 2.3 Ear Extraction

Several factors contribute to the complexity of segmenting the ear out of the image. First, ear size and shape vary widely for different persons. Second, the ear is always close to the hair, potentially obscuring the ear in an image. Third, if earrings are present, they overlap or touch the ear and should not be treated as part of the ear. These characteristics make it hard to use a fixed template to crop the ear. A bigger template will include too much hair, while a smaller template will lose shape features. Also it is hard to distinguish the ear from hair or earrings, especially when hair and earrings have similar color to skin or are very close to the ear. Starting with the ear pit determined in the previous step, we apply the active contour algorithm on both the 2D and 3D images. The initial contour is an ellipse with ear pit as center, major axis as 20 pixels and minor axis as 30 pixels. For the 2D color images, three color spaces, RGB, HSV and YCbCr were considered. It turns out that the YCbCr's Cr channel yields the best segmentation results. For the 3D images, we only use Z depth images. Figure 4 shows results when only color or depth information is used for the active contour algorithm. In the top two images, ear contours are correctly found. The active contour algorithm works well when there is color change (Figure 4(a)) or depth change (Figure 4(b)) around the ear contour in the image. But when there is no clear color or depth change, it is hard for the algorithm to stop expanding, as shown in Figures 4(c) and

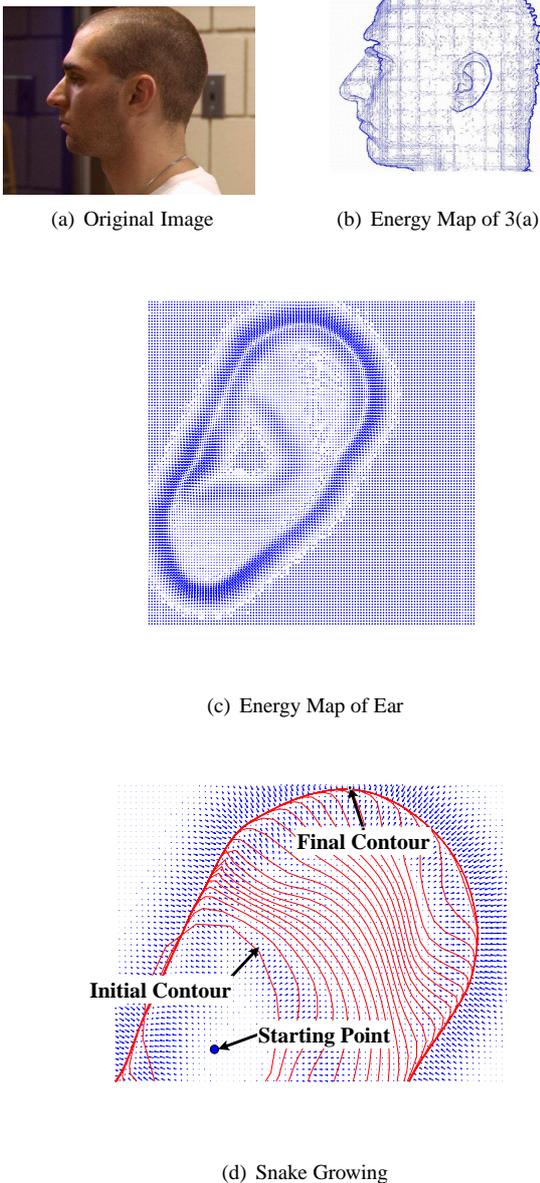
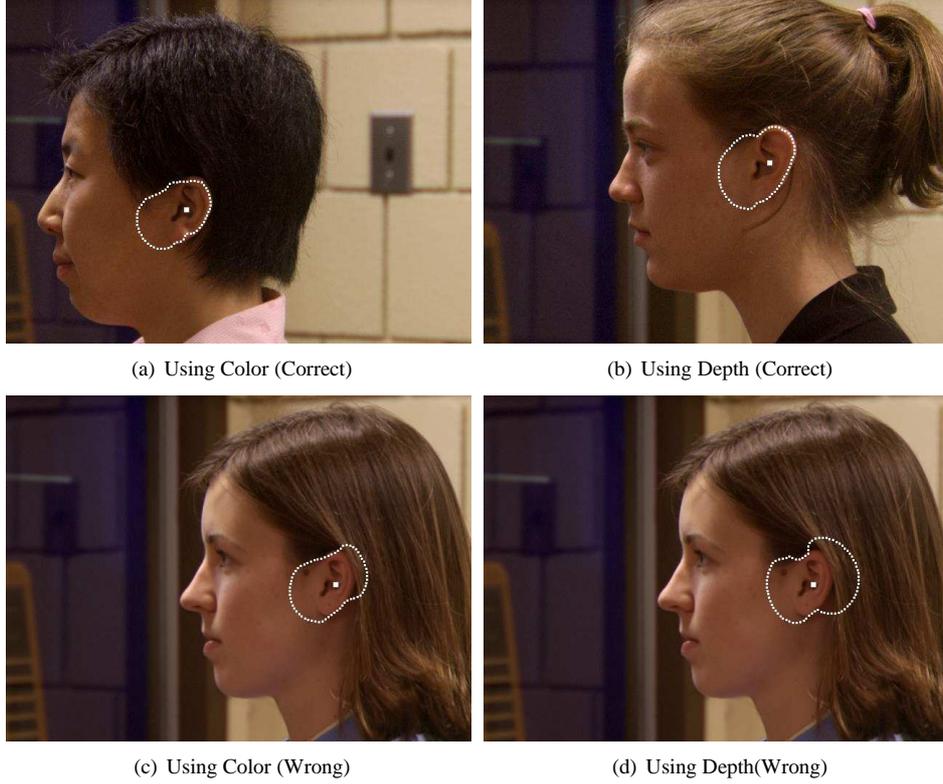


Figure 3. Snake Growing on Ear Image.



**Figure 4. Active Contour Results using Only Color or Depth.**

4(d).

Figure 4 implies that either color information or depth information alone is not adequate for contour growing. Therefore, we must combine both color and depth data. In order to combine them, the gradients of the depth images and the Cr channel from YCbCr together form  $E_{image}$ . The final energy  $E$  is represented by Equation (7).

$$\begin{aligned}
 E_{Image} &= w_{depth} \nabla Image_{depth}(x, y) \\
 &+ w_{Cr} \nabla Image_{Cr}(x, y) \quad (6) \\
 E &= \int_0^1 \frac{1}{2} [\alpha |X'(s)|^2 + \beta |X''(s)|^2] \\
 &+ w_{depth} \nabla Image_{depth}(x, y) \\
 &+ w_{Cr} \nabla Image_{Cr}(x, y) \\
 &- w_{con} \vec{n}(s) \quad (7)
 \end{aligned}$$

In Figure 4, the snake grows towards the side of the face more than expected. In order to correct this, we modify the internal energy of points to limit the expansion when there is no depth jump within a  $3 \times 5$  window around the given point. With these improvements, the new active contour algorithm works effectively in separating the ear from hair and earrings and the snake stops at the jawline close to the ear. Figure 5 illustrates the steps of snake growing for a real image. And Figure 6 shows examples when snakes deal with hair and earrings. As the active contour algorithm

finishes, the ear contour is obtained. We assume the area inside the contour is the ear region and it is cropped out of the image for use in the matching algorithm.

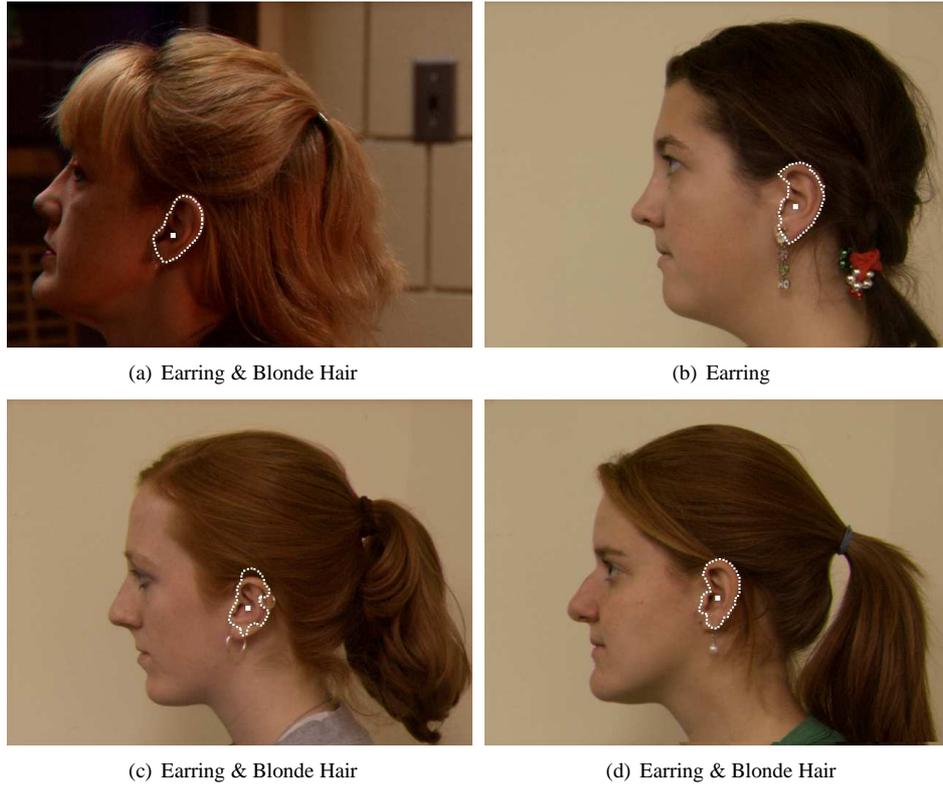
We adapted the Matlab code from [14] but modified the algorithm to use both color and intensity gradient as external energy. We used  $\alpha = 0.05$ ,  $\beta = 0$ ,  $w_{depth} = 15$ ,  $w_{Cr} = 15 * 1.5$ , and  $w_{con} = 0.3$ . The number of iterations is 150. The system runs on a dual-processor Pentium Xeon 2.8GHz machines with 2GB RAM, and it takes approximately 10-20 seconds to find the ear boundary for one image.

### 3 Ear Recognition

ICP is a well known algorithm for 3D shape matching [16, 17]. In our 3D ear recognition we use a refined ICP algorithm similar to [18], using the kd-tree data structure, automated outlier removal and point-to-triangle refinement. The starting point of the ICP algorithm is the ear pit of each image which is found in the previous step.

### 4 Experimental Results

Data was acquired with a Minolta Vivid 910 range scanner. One 640x480 3D scan and one 640x480 color image are obtained nearly simultaneously. In each acquisition session, the subject sat approximately 1.5 meters away from the sensor, with the left side



**Figure 6. Active Contour Algorithm Dealing with Earring and Blonde Hair.**

of the face facing the camera. If the subject is still during acquisition, the 3D and 2D scans are registered as obtained. In some cases with slight subject motion in between the 3D and 2D scans, the registration could be manually corrected by marking the same point in each scan and translating the 2D in X and Y. In cases of large subject motion, or motion during the several seconds of the 3D scanning time, the data had to be discarded. A total of 415 people had good quality 2D and 3D ear images in two or more sessions.

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

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  $(X_{Manual\_Ear\_Pit}, Y_{Manual\_Ear\_Pit})$ . The maximum distance difference between  $(X_{Auto\_Ear\_Pit}, Y_{Auto\_Ear\_Pit})$  and  $(X_{Manual\_Ear\_Pit}, Y_{Manual\_Ear\_Pit})$  is 20 pixels. There are slightly different results from the active contour algorithm when using automatic ear pit finding and manual ear pit marking. But as far as we can tell, the differences do not cause problems for the active contour algorithm, which means as long as the starting point is near the ear pit, the snake can find the ear boundary.

The ICP-based approach shows good performance on 3D ear recognition. The rank-one recognition rate achieves 97.6% with

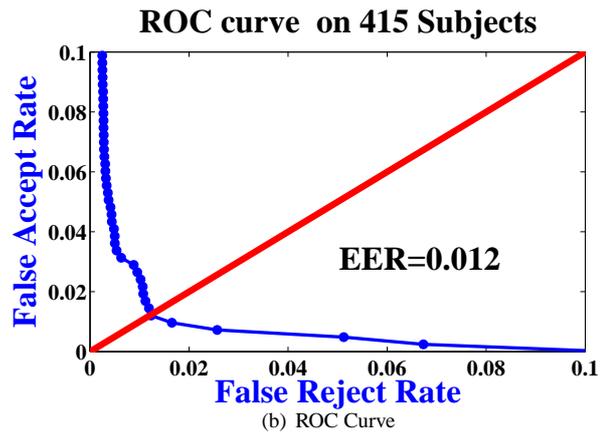
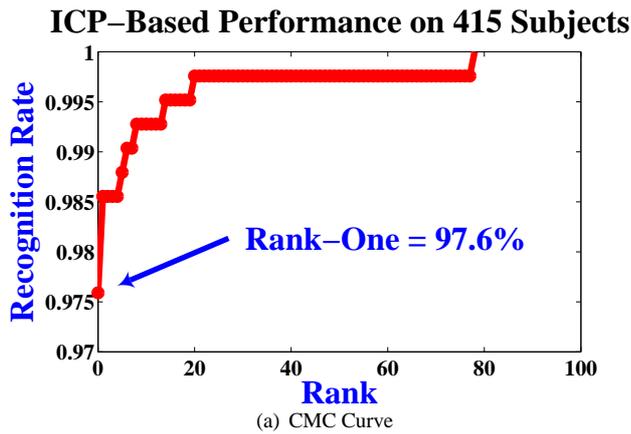
our 415-subject dataset. 380 were not wearing earrings in either image and 35 were wearing earrings in one or both images. The Cumulative Match Characteristic (CMC) curve is shown in Figure 7(a) and Receiver Operating Characteristic (ROC) curve in Figure 7(b). The equal error rate (ERR) is 0.012, which indicates a high accuracy of the biometric system. Also, the rank-one recognition is 94.2% (33 out of 35) for the 35 cases that involve earrings. Thus the presence of earrings in the image causes only minimal loss in accuracy.

The recognition system runs on a dual-processor Pentium Xeon 2.8GHz machines with 2GB RAM, and the implementation is written in C++. It takes approximately 5-8 minutes to match one probe ear against 415 gallery images when there is no subsample for gallery images, and 2 x 2 subsample for probe images.

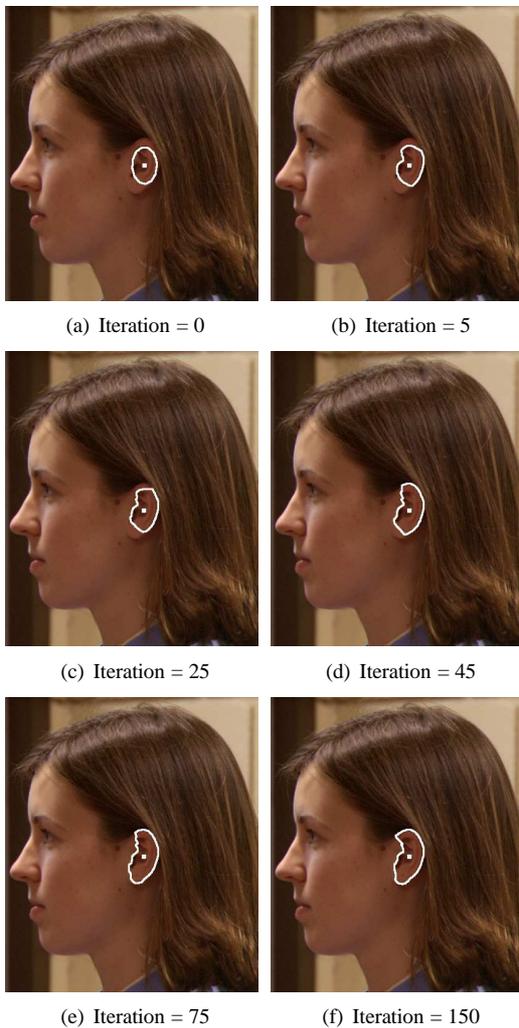
#### 4.1 Scaling with Dataset Size

Scaling of performance with dataset size is a critical issue in biometrics. The decrease in observed recognition performance with increased dataset size was observed in FRVT 2002 for 2D face recognition. “For every doubling of database size, performance decreases by two to three overall percentage points. In mathematical terms, identification performance decreases linearly with respect to the logarithm of the database size.” [19]

When the gallery size becomes bigger, the possibility of a false match increases. Some techniques scale better to larger datasets than others. Figure 8 shows the scalability of our ear recognition

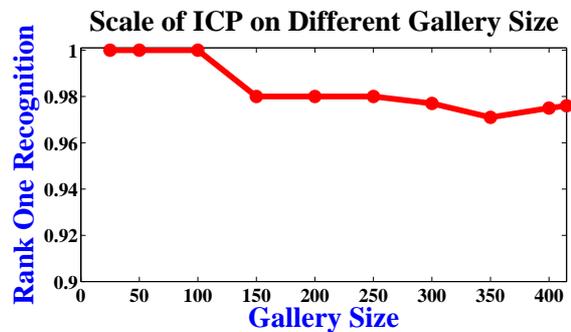


**Figure 7. The Performance of Ear Recognition**



**Figure 5. Snake Growing on A Real Image.**

algorithm. For a given data set size, the images are randomly selected from the 415 persons.



**Figure 8. Scalability of ICP-Based Ear Recognition**

#### 4.2 Comparison with 2D Face

To relate the performance level achieved here to the commercial biometric state of the art, we experimented with FaceIT version 6.1. We used frontal face images which were taken on the same date for the same person as the ear images used in the experiments. Only 411 out of the original 415 subjects are used due to data quality problems with some frontal face images. None of the 4 people who are not included in the face experiment has ear biometric recognition errors. The images are taken by Canon Power Shot G2 and Nikon D70, under controlled studio lighting and with no lighting condition change between gallery and probe. Each image has resolution 1704x2272. With normal expression both in gallery and probe, the rank one recognition rate is 98.7%. Subjectively, our impression is that FaceIT version 6 has improved on version 5 in a noticeable way. This also suggests that an ICP-based 3D ear recognition algorithm provides competitive recognition power to the commercial state of the art in face recognition.

## 5 Summary And Discussion

We present experimental results of a fully automatic ear recognition system using 2D and 3D information. The automatic ear extraction algorithm can crop the ear region from the profile image and separate the ear from hair and earring. The ear recognition uses an ICP-based approach for 3D shape matching. The rank-one recognition rate is 97.6% for the 415 people, and 94.2% on the subset of people wearing earrings. The experimental results demonstrate the power of our automatic ear extraction algorithm and of the ear as a potential biometric for human identification. However, the active contour algorithm may fail if there are no gradient changes in either color or depth image. An improvement might focus on using shape and texture constraints to help the segmentation. It might be possible to build in some preferred shape, like an ellipse, or to penalize small irregular parts to the outline.

Interestingly, we find that the ICP-based approach to 3D ear recognition scales quite well with increasing size of dataset. This result is encouraging in that it suggests the uniqueness of the human ear and its potential applicability as a biometric. However, the ICP-based approach is still computationally expensive in comparison to other approaches considered. This is a topic to be explored further in future work.

### Acknowledgements

Biometrics research at the University of Notre Dame is supported by the National Science Foundation under grant CNS-0130839, by the Central Intelligence Agency, and by the Department of Justice under grant 2004-DD-BX-1224. The authors would like to thank Dr. Patrick Flynn and Dr. Jonathon Phillips for useful discussions about this area.

### References

- [1] B. Moreno, A. Sanchez, and J. Velez, "On the use of outer ear images for personal identification in security applications," in *IEEE International Carnahan Conference on Security Technology*, 1999, pp. 469–476.
- [2] T. Yuizono, Y. Wang, K. Satoh, and S. Nakayama, "Study on individual recognition for ear images by using genetic local search," in *Proceedings of the 2002 Congress on Evolutionary Computation*, 2002, pp. 237–242.
- [3] Bir Bhanu and Hui Chen, "Human ear recognition in 3D," in *Workshop on Multimodal User Authentication*, 2003, pp. 91–98.
- [4] Hui Chen and Bir Bhanu, "Contour matching for 3D ear recognition," in *Seventh IEEE Workshops on Application of Computer Vision*, 2005, pp. 123–128.
- [5] Hui Chen and Bir Bhanu, "Human ear detection from side face range images," in *International Conference on Pattern Recognition*, 2004, pp. 574–577.
- [6] D. Hurley, M. Nixon, and J. Carter, "Force field energy functionals for image feature extraction," *Image and Vision Computing Journal*, vol. 20, pp. 429–432, 2002.
- [7] D. Hurley, M. Nixon, and J. Carter, "Force field energy functionals for ear biometrics," *Computer Vision and Image Understanding*, vol. 98, no. 3, pp. 491–512, 2005.
- [8] Ping Yan and Kevin W. Bowyer, "Empirical evaluation of advanced ear biometrics," in *IEEE Computer Society Workshop on Empirical Evaluation Methods in Computer Vision*, 2005.
- [9] Ping Yan and Kevin W. Bowyer, "Ear biometrics using 2D and 3D images," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, 2005, p. 121.
- [10] Jin Chang, *New multi-biometric approaches for improved person identification*, Ph.D. thesis, Department of Computer Science and Engineering, University of Notre Dame, USA, 2004.
- [11] Paul J. Besl and Ramesh C. Jain, "Invariant surface characteristics for 3D, object recognition in range images," *Computer Vision Graphics Image Processing*, vol. 33, pp. 30–80, 1986.
- [12] P. Flynn and A. Jain, "Surface classification: Hypothesis testing and parameter estimation," in *Computer Vision Pattern Recognition*, 1988, pp. 261–267.
- [13] Michael Kass, Andrew Witkin, and Demetri Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1987.
- [14] Chenyang Xu and Jerry L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 359–369, 1998.
- [15] 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.
- [16] P. Besl and N. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 2, pp. 239–256, 1992.
- [17] Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," *IJIVC*, vol. 10, no. 3, pp. 145–155, 1992.
- [18] Ping Yan and Kevin W. Bowyer, "ICP-based approaches for 3D ear recognition," in *Biometric Technology for Human Identification II, Proceedings of SPIE*, 2005, vol. 5779, pp. 282–291.
- [19] P.J. Phillips, P. Grother, R.J. Micheals, D.M. Blackburn, E. Tabassi, and J.M. Bone, "Frvt 2002: Overview and summary," 2003.