

# TOWARDS FAST 3D EAR RECOGNITION FOR REAL-LIFE BIOMETRIC APPLICATIONS

G. Passalis<sup>1,2</sup>

I. A. Kakadiaris<sup>2</sup>

T. Theoharis<sup>1,2</sup>

G. Toderici<sup>2</sup>

T. Papaioannou<sup>1,2</sup>

<sup>1</sup>Department of Informatics  
University of Athens  
Athens, Greece

<sup>2</sup>Computational Biomedicine Lab  
Dept. of CS, University of Houston  
Houston, TX, USA

## Abstract

*Three-dimensional data are increasingly being used for biometric purposes as they offer resilience to problems common in two-dimensional data. They have been successfully applied to face recognition and more recently to ear recognition. However, real-life biometric applications require algorithms that are both robust and efficient so that they scale well with the size of the databases. A novel ear recognition method is presented that uses a generic annotated ear model to register and fit each ear dataset. Then a compact biometric signature is extracted that retains 3D information. The proposed method is evaluated using the largest publicly available 3D ear database appended with our own database, resulting in a database containing data from multiple 3D sensor types. Using this database it is shown that the proposed method is not only robust, accurate and sensor invariant but also extremely efficient, thus making it suitable for real-life biometric applications.*

matching.

In this paper, we present a novel method for 3D ear recognition. The proposed method uses a generic annotated ear model (AEM) to register and fit each 3D ear dataset. A compact biometric signature is extracted that retains 3D information. The metadata containing this information are stored using a regular grid of lower dimension, allowing direct comparison. By appending our own database to the largest publicly available ear database, we constructed (to the best of our knowledge) the largest ear database. Moreover, this database combines 3D data from different types of 3D sensors (laser and optical). Using this database, we show that our method is robust, accurate, sensor invariant and extremely efficient, thus making it suitable for practical biometric applications.

The rest of the paper is organized as follows: Section 2 provides an overview of ear recognition, Section 3 describes the methods that we have developed, Section 4 presents our results and the biometric databases that were used, while Section 5 summarizes our work.

## 1 Introduction

The human ear, like the human face, is considered a unique characteristic of an individual, thus making it suitable for biometric applications. Compared to the face, the ear has the added advantage that it takes no expressions. As in face recognition [1, 2], research interest in ear recognition is shifting from two-dimensional (2D) to three-dimensional (3D) methods. This is attributed to the fact that 3D data, that are becoming increasingly available, offer resilience to problems common in 2D data such as pose and illumination.

However, real-life biometric applications require algorithms that are both robust and efficient so that they scale well with the size of the databases. This requirement dictates that the biometric signature of a 3D ear must be represented in a way that it is easily comparable with other 3D ear signatures. Therefore, any computationally expensive steps should be performed during preprocessing and not during

## 2 Related Work

Bhanu and Chen have published first a method for 3D ear recognition [5]. They employed a local surface shape descriptor. On a database composed of ten individuals with two images each, they reported a 100% recognition rate. The descriptor is applied on ear meshes which were manually extracted from the 3D scan. Later, they used a two stage ICP approach for matching and a database composed of 30 subjects [6]. The results reported show that only two subjects were not recognized properly. As earlier, the meshes on which ICP was ran were manually segmented from the input data.

Yan and Bowyer [14] compare several approaches to 3D ear recognition: PCA ("eigenears") on the Z component of range data, Hausdorff distance of the edges computed from the Z component image, and the distance after running ICP

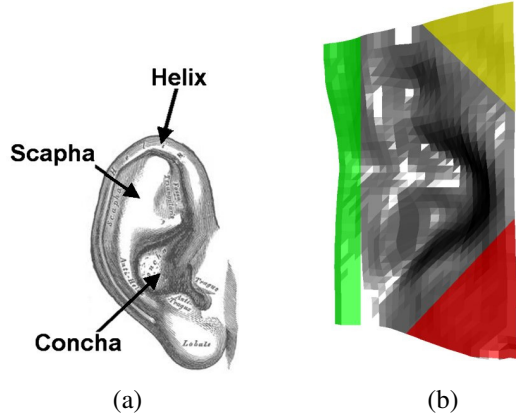


Figure 1: (a) The areas of the human ear. (b) The annotated ear model (AEM).

on the mesh representation. All methods were applied on data in which the ear was manually segmented from the profile view. On a database of 302 subjects, PCA achieved a 55.3% rank one recognition rate, while the Hausdorff approach achieved 67.5%. On the same database ICP yielded 98.7%, and on a database of 404 subjects, it provided a performance of 97.5%.

Yan and Bowyer [15, 16] proposed a new ICP-based approach for ear recognition that significantly decreases their computational time, which is essential if such an approach is to be used in practice. Additionally, they propose an algorithm which uses heuristics based on some constraints of the input data, and active contours for automatic ear extraction. The results they report require no manual intervention. On a database of 415 subjects the rank one recognition rate reported is 97.8%. Chen and Bhanu [17] utilize the same database but they present lower results than Yan and Bowyer. We also employ this database for our experiments since it is the largest publicly available 3D ear database.

### 3 Methods

The main idea behind our method is to employ an annotated ear model (AEM) that is representative of human ears. This model is purely geometrical; it is used to register each ear dataset and then through a fitting process to acquire its shape. Subsequently, 3D information is extracted from the fitted model and stored as metadata. This representation is compact and directly comparable, thus making our method robust and efficient.

The AEM needs to be created only once and is based on statistical data. It is a polygonal 3D mesh containing valence-6 vertices only. Instead of building a model for the whole ear we opted for an inner ear model. This was due to the fact that the outer part of the ear is usually occluded by

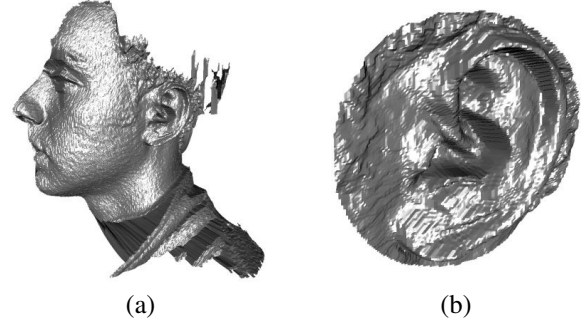


Figure 2: A 3D ear dataset (a) before segmentation, and (b) after segmentation.

hair or other accessories, thus limiting its value as a biometric. We thus based our model on the *Concha* area of the ear (Fig.1(a)). The model along with its annotation is depicted in Fig.1(b). The outline of the proposed method follows:

**Enrollment:** Raw data are converted to metadata and stored in the database.

1. *Preprocessing:* Raw data are preprocessed and segmented in order to alleviate artifacts introduced by the 3D sensor.
2. *Registration:* The raw data are registered to the AEM.
3. *Deformable Model Fitting:* The annotated model is fitted to the data.
4. *Metadata extraction:* A biometric signature is extracted using geometry information and stored in a database as metadata.

**Authentication:** Metadata retrieved from the database are directly compared using a distance metric.

Note that most methods described in the related work section use a registration step during matching, therefore registration must be performed for every dataset pair. The main feature of the proposed method is that by using the AEM, the registration step is performed only once for each dataset, during enrollment. Thus, the computationally expensive registration step is performed only once, when a new dataset is added to the database. This allows the method to be practical in real-life scenarios, since for  $n$  datasets the registration step is performed  $n$  times (during enrollment) instead of  $n^2$  (during authentication).

#### 3.1 Preprocessing

The purpose of preprocessing is twofold: to eliminate sensor-specific artifacts and to segment the ear datasets. In general, modern 3D sensors output either a range image

or 3D polygonal data. We implemented the following filters for both representations, operating on a 1-neighborhood area in both cases, thus making the method sensor invariant:

- *Median Cut*: This filter is applied to remove spikes from the data. Spikes are more common in laser range scanners, therefore stronger filtering is needed in this case.
- *Hole Filling*: Laser scanners usually produce holes in certain areas (e.g., eyes, eyebrows) so a hole filling procedure is applied.
- *Smoothing*: A smoothing filter is applied to remove white noise as most high resolution scanners produce noisy data in real-life conditions.

Before each dataset is used in the the next steps of our method, ear segmentation must be performed, as shown in Figs. 2(a,b). We keep only the 3D geometry that resides within a sphere of a certain radius that is centered roughly on the ear pit. Using a custom specialized tool, a human operator places the center of this sphere, guided by information such as the center of mass and the average normal. This segmentation is the only part of our method that is not fully automatic. Even though the process itself is not time-consuming and does not require an experienced operator, human input can be eliminated by employing the methods described by Yan *et al.* [15, 16]. In their approach, they show that the performance of the automatic ear segmentation can match that of the human operator.

### 3.2 Registration

Each dataset is registered with the AEM before it is fitted. Registration is the most critical step of our method since errors introduced here cannot be alleviated later. In order to achieve the best possible accuracy we employ two different registration algorithms. In both cases each dataset is registered with the AEM, with the results of the first algorithm being used as input for the second.

The first algorithm we employ is the Iterative Closest Point (ICP) [4]. ICP determines correspondences between vertices and minimizes the sum of the square of their distances. We employ an improvement suggested by Turk *et al.* [13] to reject vertex pairs containing points on surface boundaries. Our experiments showed that even though ICP is relatively robust and gives a good approximation, it is not accurate enough.

To this end, we employ an additional fine-tuning registration algorithm based on the work of Papaioannou *et al.* [11] that uses a global optimization technique (Simulated Annealing [7, 12]) applied to depth and normal images. As in most registration algorithms, we optimize a total of six parameters (three for translation and three for rotation). Both

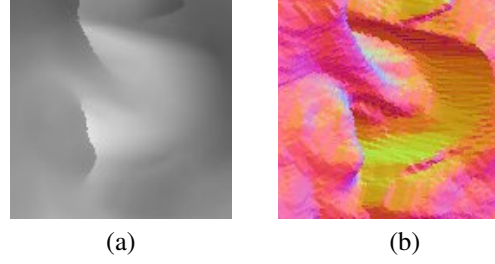


Figure 3: Images from an ear dataset as used in Simulated Annealing: (a) depth image, and (b) normal image.

the dataset and the AEM are rendered using OpenGL in order to extract the depth and normal images, depicted in Figs. 3(a, b), respectively. The depth image is derived directly from the z-buffer while the normal image from the color buffer. In order to derive the normal image from the color buffer the vertex normals must be used as vertex colors. The discrete sum of the differences of these buffers is computed as follows:

$$E_d = \sum_{i=1}^R \sum_{j=1}^R |D_{model}(i, j) - D_{data}(i, j)|$$

and

$$E_c = \sum_{i=1}^R \sum_{j=1}^R (|C_{model}^R(i, j) - C_{data}^R(i, j)| + |C_{model}^G(i, j) - C_{data}^G(i, j)| + |C_{model}^B(i, j) - C_{data}^B(i, j)|)$$

where  $R$  is the spatial resolution of the buffers,  $D$  is the z-buffer, and  $C^R$ ,  $C^G$  and  $C^B$  are the three chromatic components of the color buffer. The errors from depth and color buffers are summed up and the total error is used in the optimization function:  $E = E_d + w \cdot E_c$  where  $w$  is a normalization weight that was selected empirically so that both error components contribute equally.

We used both depth and normal images because each one has its own advantages and disadvantages. Depth images are extremely sensitive to spikes and other quality problems commonly found in the ear database. On the contrary, normal images are less sensitive to these problems but do not provide information about the translation along the Z axis. The combination allows Simulated Annealing to find the optimal solution. An empirically selected weight was used to balance the contribution of each image type to the final error metric.

### 3.3 Deformable Model Fitting

The annotated ear model is fitted to each individual dataset in order to capture the geometric characteristics of the subject's ear. This is achieved using a deformable model approach and the derived deformed model offers an alternative representation of the ear's geometry. It utilizes the

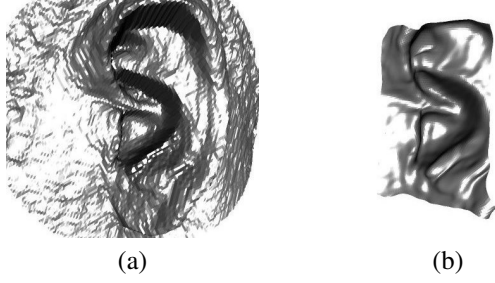


Figure 4: Fitting of AEM: (a) raw polygonal data, and (b) fitted model.

deformable model framework [10]. Based on the work of Mandal *et al.* [9] the framework is combined with Loop subdivision surfaces [8].

The analytical formulation that was used is given by:

$$\mathbf{M}_q \frac{d^2 q}{dt^2} + \mathbf{D}_q \frac{dq}{dt} + \mathbf{K}_q q = \mathbf{f}_q$$

where  $\mathbf{M}_q$  is the mass matrix,  $\mathbf{D}_q$  is the damping matrix,  $\mathbf{K}_q$  is the stiffness matrix, and  $\mathbf{f}_q$  are the external forces. Note that  $q$  represents the degrees of freedom of the model. The external forces drive the deformation, the stiffness matrix resists the deformation, and the mass and damping matrices control the velocity and acceleration of the vertices, respectively. This equation is solved in an iterative way using a Finite Element Method (FEM) approximation. When the deformation is completed the AEM has acquired the shape of the individual dataset (Fig.4). The deformed AEM effectively resamples the polygonal dataset. It is therefore a representation independent of the dataset’s original resolution and free of artifacts (e.g., missing areas, spikes).

### 3.4 Metadata Extraction and Comparison

The deformed AEM is used to produce the metadata. As was the case with registration, a depth and a normal image are created. The biometric signature coefficients are the concatenation of the pixel values of these two images. Note that since the metadata are in fact a two dimensional structure, image transforms (e.g., Fourier, Wavelet) can be employed to reduce their storing requirements. For comparison purposes we use an  $L^1$  metric on the coefficients. The total difference between two coefficient sets is the sum of the differences among all the individual coefficient pairs. Therefore, the comparison is a very efficient process that does not include any computational expensive step.

## 4 Results

In order to evaluate our method we employ the largest database for ear recognition. We use this database to

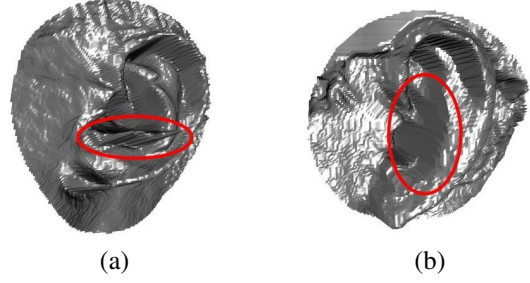


Figure 5: Sensor-related quality problems; problematic areas circled: (a) spikes, and (b) missing information.

measure our method’s efficiency and performance. The database is divided into gallery and probe sets, with each subject having exactly one dataset in the gallery set. The performance is measured using a Cumulative Match Characteristic (CMC) curve and the rank-one recognition rate is reported.

### 4.1 Database

The utilized database is the combination of two separate components. Both components of the database include left ears only, acquired with the 3D sensor pointing directly to the ear (straight ear profile). In order to process right ears we would have to mirror the AEM with the rest of the method remaining unchanged.

The first component is the Ear Database from the University of Notre Dame (UND) [3], a publicly available database. It contains 830 datasets from 415 subjects and the data were acquired from fall 2003 to fall 2004. A Minolta 910 laser scanner was used, which produces range images with a resolution of 640x480. The second component is a database we have acquired using a 3dMD™ optical multi-pod system during Fall 2005. This database contains 201 3D polygonal datasets from 110 subjects.

The combined multi-sensor database contains a total of 1031 datasets from 525 subjects. Several of the datasets have quality problems (depicted in Fig.5) but were included to increase the difficulty of the experiments. To the best of our knowledge this is the most extensive and most challenging database reported in 3D ear recognition.

### 4.2 Efficiency Evaluation

Biometric experiments on databases usually require the creation of a similarity matrix that has row count equal to the number of probe datasets and column count equal to the number of gallery datasets. Therefore, the number of comparisons required are  $O(n^2)$ . These comparisons are the major bottleneck for most methods when applied to large databases, such as the one used in this paper.



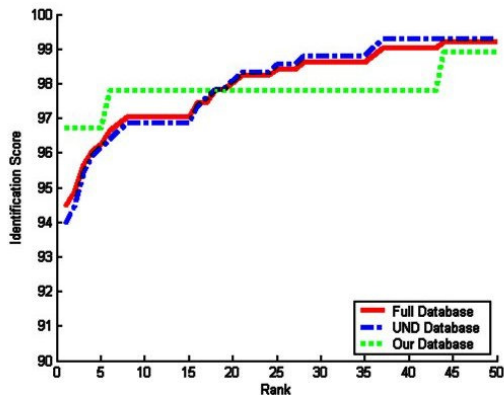


Figure 6: CMC curves for the proposed method reported on the full database and on the two components.

The computational cost and storage requirements of the proposed method are summarized on the following table:

Enrollment	30 sec per dataset
Authentication	less than 1 msec per comparison
Storage	28KB per metadata

Enrollment is performed for every dataset and includes registration, fitting and metadata extraction. Authentication is performed on every pair in the similarity matrix. These measurements were carried out on a typical modern PC with a Pentium 4 at 3Ghz and 1GB of RAM.

In their latest work Yan and Bowyer [15, 16] also use the UND database. They report that it takes 30-50 minutes to compute a row of the similarity matrix (a single probe compared to 415 gallery datasets). Since there are 415 rows for this database, this translates to a total time of 12450 to 20750 minutes, or 276 hours on average. For the same database the proposed method takes approximately 7 hours for enrollment and a few minutes for authentication (computing the full similarity matrix).

Therefore, compared to a typical ICP-based approach the proposed method takes less than 3% of the time to process the exact same database. Moreover, if both methods are applied to even larger databases the difference will increase further. Even though a number of implementation details can affect these results, it is evident that avoiding costly registration steps during matching offers a definite efficiency advantage. The total computational cost is decreased by an order of magnitude ( $O(n)$  instead of  $O(n^2)$ ).

### 4.3 Performance Evaluation

We performed an identification experiment where each of the 506 probe datasets of the database are compared with each of the 525 gallery datasets. The CMC curve of this experiment is depicted in Fig.6. The rank-one recognition rate

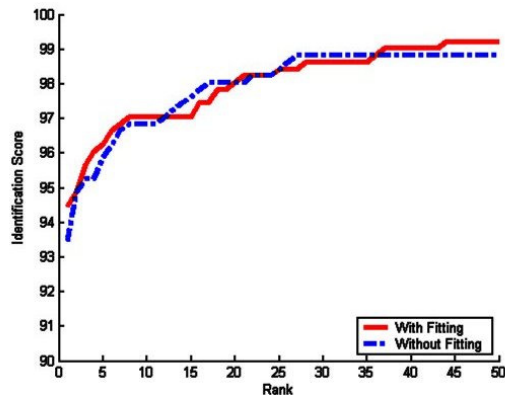


Figure 7: CMC curves for the proposed method reported on the full database with and without the fitting step.

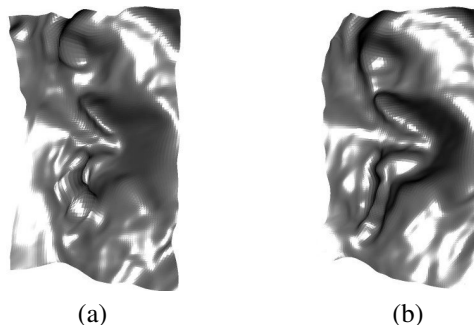


Figure 8: Failure case: (a) gallery, and (b) probe from the same individual.

for the full database is 94.4%. In this figure separate CMC curves are provided for each component of the database: for the UND database the rank-one rate is 93.9% while for our database it is 96.7%.

The proposed method has an option to completely omit the deformable model fitting step. If this step is omitted, the metadata are extracted from the registered raw polygonal data. The advantage is that the computational cost is decreased; the time needed to extract the metadata from an individual dataset is reduced, from 30 seconds to 15 seconds. The performance penalty is approximately 1%, with the rank-one recognition rate dropping from 94.4% to 93.4% (Fig.7). Depending on the timing requirements of the biometric application, this trade-off could be desirable.

### 4.4 Failure Analysis

The rank-one recognition rate of 94.4% for the full database corresponds to 28 failure cases (from 506 probes). A detailed visual inspection of these cases allowed us to group the failures into three categories:

1. *Poor quality of data:* Certain datasets of the database have quality problems (e.g., spikes, missing information) that cause inconsistencies between the gallery and probe datasets of the same individual (Fig.8). This is the most common source of failures.
2. *Erroneous registration:* The registration algorithms always try to find the global optimal solution, but it is possible that they get trapped in a local minimum.
3. *Incorrect fitting:* In cases of ears with intricate geometric structure it is possible that the AEM fails to deform correctly. This failure category is sparsely observed.

## 5 Conclusion

A novel ear recognition method was presented that uses 3D geometry information of the ear. An annotated ear model is used for registration and fitting purposes and a compact biometric signature is extracted. The main feature of the proposed method is that the AEM allows the registration step to be performed only once per dataset and not during matching. Using an extensive database, the proposed method's efficiency and performance were evaluated. It offers competitive performance, top efficiency and sensor invariance. Due to the extremely efficient matching step the computational complexity is proportional to the size of the database ( $O(n)$  instead of  $O(n^2)$ ). Combined with an automated ear segmentation method, the proposed method is suitable for real-life biometric applications that require large databases. Future work will be directed towards applying the proposed method to even larger databases with the goal to establish ear recognition as a reliable biometric.

### Acknowledgment

The authors would like to acknowledge financial support from the GSRT under project 05NON-EU-91, the NSF instrumentation award CNS-0521527 and 3dMD<sup>TM</sup> for providing the ear sensor on loan.

## References

- [1] I.A. Kakadiaris, G. Passalis, G. Toderici, M.N. Murtuza, Y. Lu, N. Karampatziakis and T. Theoharis. Three-Dimensional Face Recognition in the Presence of Facial Expressions: An Annotated Deformable Model Approach. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29(4):640–649, 2007
- [2] P.J. Phillips, W.T. Scruggs, A.J. O'Toole, P.J. Flynn, K.W. Bowyer, C.L. Schott, and M. Sharpe. FRVT 2006 and ICE 2006 Large-Scale Results. *National Institute of Standards and Technology, NISTIR 7408*, <http://face.nist.gov>, 2007
- [3] University of Notre Dame biometrics database. <http://www.nd.edu/%7Ecivl/UNDBiometricsDatabase.html>
- [4] P. Besl and N. McKay. A method for registration of 3-D shapes. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(2):239–256, Feb. 1992.
- [5] B. Bhanu and H. Chen. Human ear recognition in 3D. In *Proc. of Workshop on Multimodal User Authentication*, pages 91–98, 2003.
- [6] H. Chen and B. Bhanu. Contour matching for 3D ear recognition. In *Proc. of WACV-MOTION*, volume 1, pages 123–128, Breckenridge, CO, 5-7 January 2005.
- [7] S. Kirkpatrick, C. Gelatt, and M. Vecchi. Optimization by simulated annealing. *Science*, 22(4598):671–680, 1983.
- [8] C. Loop. Smooth subdivision surfaces based on triangles. Master's thesis, Department of Mathematics, University of Utah, 1987.
- [9] C. Mandal, H. Qin, and B. Vemuri. A novel FEM-based dynamic framework for subdivision surfaces. In *Proc. of fifth ACM symposium on Solid modeling and applications*, pages 191–202, New York, NY, USA, 1999.
- [10] D. Metaxas and I. Kakadiaris. Elastically adaptive deformable models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(10):1310–1321, 2002.
- [11] G. Papaioannou, E. Karabassi, and T. Theoharis. Reconstruction of three-dimensional objects through matching of their parts. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(1):114–124, January 2002.
- [12] P. Siarry, G. Berthiau, F. Durbin, and J. Haussay. Enhanced simulated annealing for globally minimizing functions of many-continuous variables. *ACM Trans. on Mathematical Software*, 23(2):209–228, 1997.
- [13] G. Turk and M. Levoy. Zippered polygon meshes from range images. In *Proc. of SIGGRAPH*, pages 311–318, Orlando, FL, July 1994.
- [14] P. Yan and K. Bowyer. Ear biometrics using 2D and 3D images. In *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, page 41, San Diego, CA, 20-26 June 2005.
- [15] P. Yan and K. Bowyer. An automatic 3D ear recognition system. In *Proc. of Third International Symposium on 3D Data Processing, Visualization and Transmission*, University of North Carolina, Chapel Hill, June 14-16 2006.
- [16] P. Yan and K. Bowyer. Biometric recognition using three-dimensional ear shape. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, In press, 2007.
- [17] H. Chen, B. Bhanu. Human ear recognition in 3D, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29(4):718–737. 2007.