

# Unified 3D Face and Ear Recognition using Wavelets on Geometry Images

Theoharis Theoharis<sup>a,b</sup> Georgios Passalis<sup>a,b</sup> George Toderici<sup>b</sup>

Ioannis A. Kakadiaris<sup>b</sup>

<sup>a</sup>*Computer Graphics Group*

*Department of Informatics and Telecommunications*

*University of Athens, Ilisia 15784, GREECE*

<sup>b</sup>*Computational Biomedicine Lab*

*Department of Computer Science*

*University of Houston, Texas 77204, USA*

---

## Abstract

As the accuracy of biometrics improves, it is getting increasingly hard to push the limits using a single modality. In this paper, a unified approach that fuses three-dimensional facial and ear data is presented. An annotated deformable model is fitted to the data and a geometry image is extracted. Wavelet coefficients are computed from the geometry image and used as a biometric signature. The method is evaluated using the largest publicly available database and achieves 99.7% rank-one recognition rate. The state-of-the-art accuracy of the multimodal fusion is attributed to the low correlation between the individual differentiability of the two modalities.

*Key words:* Face Recognition, Ear Recognition, Multimodal Biometrics, Wavelets, Geometry Images, Deformable Models

---

## 1 Introduction

Among the different biometric modalities, the ones that rely on three-dimensional (3D) information are constantly gaining ground. This is due to the increased availability of 3D scanners, and to the inherent advantages of 3D data which do not suffer from limitations commonly found in 2D data (e.g., pose, illumination).

Biometric recognition algorithms based on 3D face and, more recently, 3D ear data have appeared and achieved high accuracy. This is approximately 97% rank-one recognition rate on widely accepted databases. As we approach the 100% mark, progress is getting harder as the discriminatory power of the algorithms is exhausted as similar datasets from different subjects and problematic datasets exist in any single modality. We thus strongly believe that significant further progress can only result from fusing multiple modalities. To be effective such fusion must combine modalities that have low correlation in their individual differentiabilitys.

Both the human face and human ear, are considered unique characteristics of an individual, thus making them suitable for biometric applications. Each modality is widely used by many approaches and some proved to be robust and relatively accurate. However, each modality has its own limitations. For example, faces are subject to facial expressions which can affect recognition. On the other hand, the inner ear's elaborate structure cannot be fully captured

---

*Email addresses:* [theotheo@di.uoa.gr](mailto:theotheo@di.uoa.gr) (Theoharis Theoharis),  
[passalis@di.uoa.gr](mailto:passalis@di.uoa.gr) (Georgios Passalis), [gtoderici@uh.edu](mailto:gtoderici@uh.edu) (George Toderici),  
[ioannisk@uh.edu](mailto:ioannisk@uh.edu) (Ioannis A. Kakadiaris).

by modern 3D scanners due to self-occlusions.

Compared to other multimodal options the combination of face and ear offers certain advantages. The data can be captured using the same equipment and they are both represented as geometry. The latter allows the face and ear to be considered parts of the same biometric, the human head. Therefore, methods that can seamlessly handle both types of data are becoming increasingly important. In this paper, we present such a method that combines 3D face and ear data. Moreover, we show that there is a low correlation between the differentiability of 3D face and ear data. Most importantly, it boosts rank-one recognition accuracy to 99.7% on the largest publicly available multimodal database.

### *1.1 Related Work*

Hurley [10] was the first to propose a method suitable for both face and ear. He presented a force field transform that could be applied on 2D images of the face or ear. An evaluation of 2D ear and face biometrics was given by Victor [2]. According to that work, face biometrics performed significantly better than ear biometrics. On a latter work, Chang [4] contradicted the results of Victor, showing superior performance for the ear modality. Chang used an eigen-based method that allowed the combination of the two modalities presenting a multimodal biometric that performed better than each separate modality. However, in the above studies, only 2D data of the face and ear were used.

In the 3D face recognition domain, most recent works utilize the FRGC v2 database, the largest publicly available 3D face database. This database is

also used in this paper (see Sec. 3). On this database, Chang [5] examined the effects of facial expressions using two different 3D recognition algorithms. They reported a 92% rank-one recognition rate. Husken [11] presented a multimodal approach that uses hierarchical graph matching (HGM). They extended their HGM approach from 2D to 3D but the reported 3D performance is lower than the 2D equivalent. Their fusion, however, offers competitive results, 96.8% verification rate at 0.001 False Acceptance Rate, compared to 86.9% for the 3D only. Maurer [16] also presented a multimodal approach tested on the FRGC v2 database, and reported a 87% verification rate at 0.01 FAR. In our previous work on this database [12], we reported the highest scores, using the 3D face modality alone: 97% rank-one recognition and an average verification rate of 97.1% at 0.001 False Acceptance Rate.

In the 3D ear recognition domain, Chen [6] presented a method that uses a local surface patch to compute feature points. Using a subset of the UND Ear database, which is also used in this paper (see Sec. 3), they reported 96.4% rank-one recognition rate. Note that they utilized a smaller subset (302 subjects) than we utilized in this paper.

Using the same database, but using a larger subset (415 subjects), Yan and Bowyer [26,27] reported 97.6% rank-one recognition rate, for their 3D ear recognition method. They propose 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.

There has been very little work in combining the 3D face and ear modalities.

Only Woodward [25] have attempted to fuse 3D ear, face and finger data. They achieved 97% rank-one recognition rate on a small database of 85 individuals using all three modalities. To the best of our knowledge, the method proposed in this paper outperforms all previous single or multimodal approaches (3D face and ear) that presented results on similar sized databases.

## *1.2 Overview*

In this paper, we propose a combined face and ear approach that uses 3D data. We extend our previous work on intra-class 3D object retrieval [18] to handle human ears. We then incorporate improvements that we successfully deployed in the face recognition domain [12]. The result is a novel unified approach that can seamlessly handle both faces and ears.

An annotated deformable model is constructed for each object class, face and ear. Each model is fitted to the corresponding 3D datasets using a subdivision-based deformable framework. Subsequently, the geometry image of the deformed model is computed, and wavelet coefficients are extracted. These coefficients form a multimodal biometric signature that achieves state-of-the-art performance. The method is automatic, robust and efficient and it requires no training as it does not use statistical data. It is shown that each modality confutes the shortcomings of the other, thus making 3D faces and ears a very accurate multimodal biometric.

The rest of the paper is organized as follows: Sec. 2 describes the methods we have developed, Sec. 3 describes the biometric databases, Sec. 4 presents our state-of-the-art performance, while Sec. 5 summarizes our work.

## 2 Methods

The proposed method processes each face and ear dataset through a common pipeline of algorithms. The only difference between the processing of faces and ears is that each uses its own annotated model. This model is representative of the respective classes (face and ear) and is purely geometrical. The model is used for registering each dataset and then, through a fitting process, acquires its shape. A regularly sampled representation called the geometry image is extracted and a wavelet transform is applied to extract the biometric signature. The steps that each face or ear dataset goes through are:

- (1) *Preprocessing*: Raw data are preprocessed and segmented.
- (2) *Registration*: The raw data are registered to the annotated model using a two-phase algorithm.
- (3) *Deformable Model Fitting*: The annotated model is fitted to the data. The fitted model is then converted to a geometry image.
- (4) *Wavelet Analysis*: A wavelet transform is applied on the geometry image (and derived normal image) and the wavelet coefficients are exported and stored.

### 2.1 Annotated Models

We have constructed two annotated models, the Annotated Face Model (AFM) and the Annotated Ear Model (AEM), both depicted in Fig. 1. These models need to be created only once, and are used to process any number of datasets. Both models share some basic characteristics. They are both average shapes of their class, created from statistical data. They are represented by a 3D

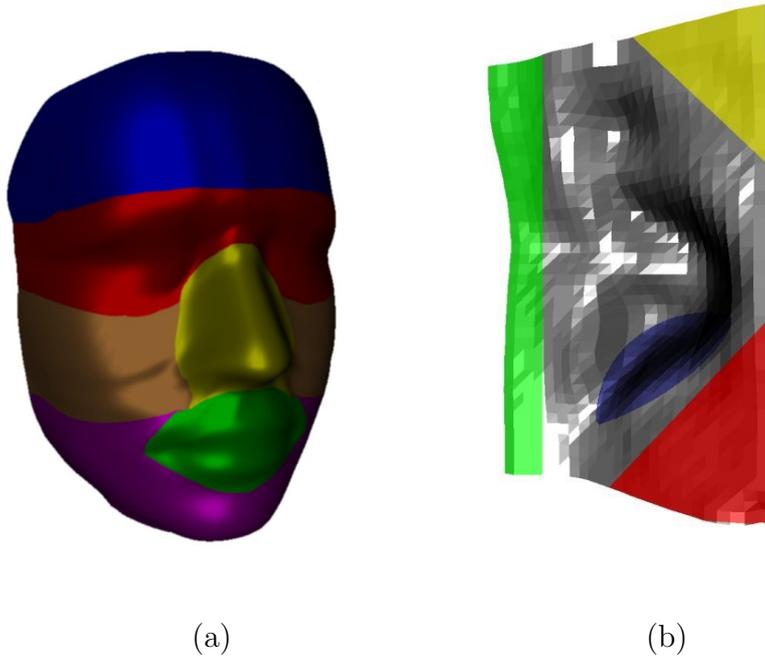


Fig. 1. (a) Annotated Face Model; (b) Annotated Ear Model. Annotation areas are marked with different color.

polygonal mesh that mainly has valence-6 vertices. This is due to the fact that these vertices define the control points of a subdivision surface as explained in Sec. 2.4. The AEM utilizes only the inner area of the ear (called *concha*) because the outer part of the ear is usually occluded by hair or other accessories, thus limiting its value as a biometric characteristic.

Certain areas are annotated on each model (see Fig. 1). The annotation was done empirically and is based on the physiology of the face and ear. These areas have different properties associated with them which are utilized by our method (e.g., rigidity, feature importance, resilience to noise). For example the mouth area, is considered less rigid than the nose area in the fitting step and is also considered less important in the comparison step. In the AEM, the bottom of the ear pit (marked blue in Fig. 1 (b)), is expected to have spikes due to self-occlusion, and is assigned a lower weight during comparison.

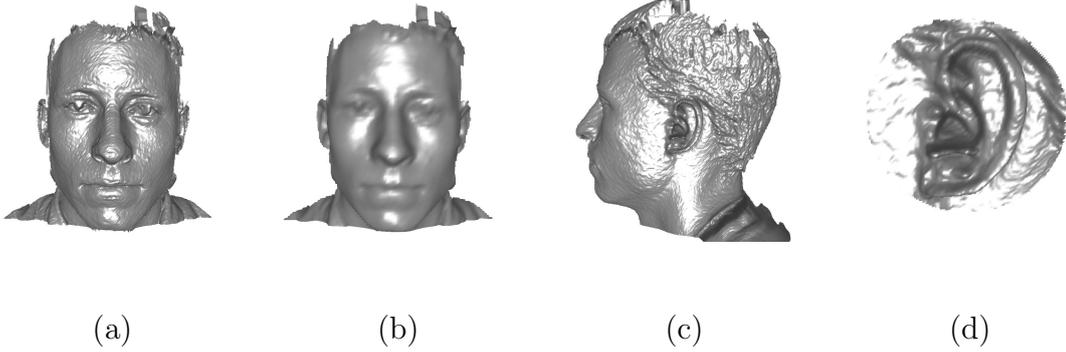


Fig. 2. Raw (a,c) and processed (b,d) datasets of the same subject.

Finally, a continuous global UV parametrization is applied to the annotated models. The specific parametrization is essentially a mapping from  $R^3$  to  $R^2$  and has two significant properties: it is area-preserving and injective. The area-preserving property minimizes the stretching imposed by the mapping. The injective property allows us to convert the annotated models from a polygonal representation to an equivalent geometry image representation. Geometry images [9,20] are regularly sampled 2D images that have three channels, encoding geometric information ( $x$ ,  $y$  and  $z$  components of a vertex in  $R^3$ ). In this work, the number of channels in the geometry image is greater than three, as apart from geometric information we also encode normal and annotation information.

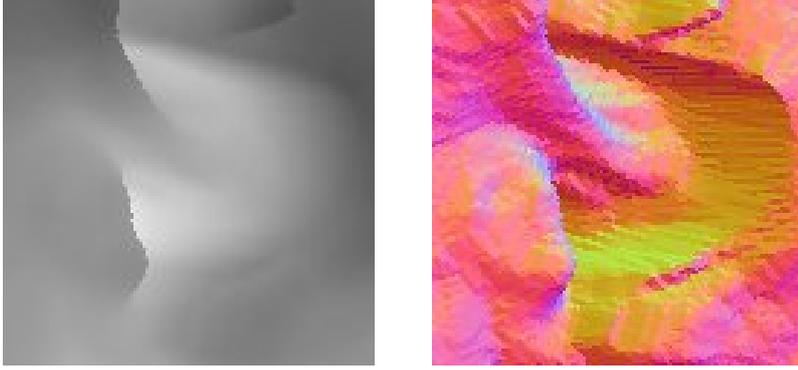
## 2.2 Preprocessing

The purpose of preprocessing is mainly to eliminate sensor-specific problems and for the case of the ear datasets, to segment them from the rest of the head (see Fig. 2). In general, modern 3D sensors output either a range image or 3D polygonal data, but in our experiments we used only range images from laser scanners. Therefore, the following preprocessing algorithms operate directly on the range data, before the conversion to polygonal data.

- *Segmentation*: Only ear datasets need segmentation; the face datasets omit this step. We keep the 3D geometry that resides within a sphere of a certain radius that is centered roughly on the ear pit. Using a custom 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 and we are currently in the process of automating it.
- *Median Cut*: This filter is applied to remove spikes from the data. Spikes are common issues with laser range scanners, especially in certain areas such as the eyes or the ear pit. A  $3 \times 3$  kernel is used.
- *Hole Filling*: Laser scanners usually produce holes in certain areas where hair are present, 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. A  $3 \times 3$  kernel is used.
- *Subsampling*: Only face datasets need subsampling, the ear datasets omit this step. Using data with resolutions higher than the AFM is a waste of computational time, as the AFM effectively resamples the data during fitting.

### 2.3 Registration

Data acquired from 3D scanners have arbitrary orientation. Before the annotated model (AFM or AEM) is fitted to these data, the two must be coregistered. To this end, we employ a two-phase *rigid* registration step. In both phases the dataset is registered with the annotated model, with the results of



(a)

(b)

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

the first phase being used as input for the second.

The first phase utilizes is the Iterative Closest Point (ICP) algorithm [3]. ICP determines correspondences between vertices and minimizes the sum of the square of their distances. We employ an improvement suggested by Turk et al.[24] to reject vertex pairs containing points on surface boundaries. Even though ICP always gives a plausible registration, our experiments showed that it is not optimal.

To this end, as a second phase, we utilize the registration algorithm presented by Papaioannou [17] that uses a global optimization technique (Simulated Annealing [13,21]) applied to depth images. We modified the algorithm to compute normal images as well (Fig. 3). In an OpenGL environment, the depth images are derived directly from the z-buffer and the normal images from the color buffer. The orientation of the raw data is controlled by the parameter vector  $\omega = [t_x, t_y, t_z, \phi, \theta, \psi]$  where  $(t_x, t_y, t_z)$  is the translation vector and  $\phi, \theta, \psi$  are the Euclidean rotation angles. This parameter vector controls the

Simulated Annealing process which minimizes the following objective function:

$$E_{total} = E_d + w \cdot E_c$$

where  $E_d$  is the depth buffer error,  $E_c$  is the color buffer error, and  $w$  is an empirically selected normalization weight. The depth and color buffer errors are 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.

Registration is the most critical step of our method as errors introduced here cannot be alleviated later, and generally result in failures. 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 datasets. However, normal images are less sensitive to these problems but do not provide information about translation along the Z axis. Their combination allows Simulated Annealing to find an excellent solution.

#### 2.4 Deformable Model Fitting

In order to fit the annotated model (AFM or AEM) to the raw data, a deformable model framework [23] is utilized. The main idea is that the deformation of the AFM is controlled by internal and external forces. The internal forces correspond to the elastic properties of the model's surface (e.g., strain energy, material stiffness) and resist the deformation. The external forces de-

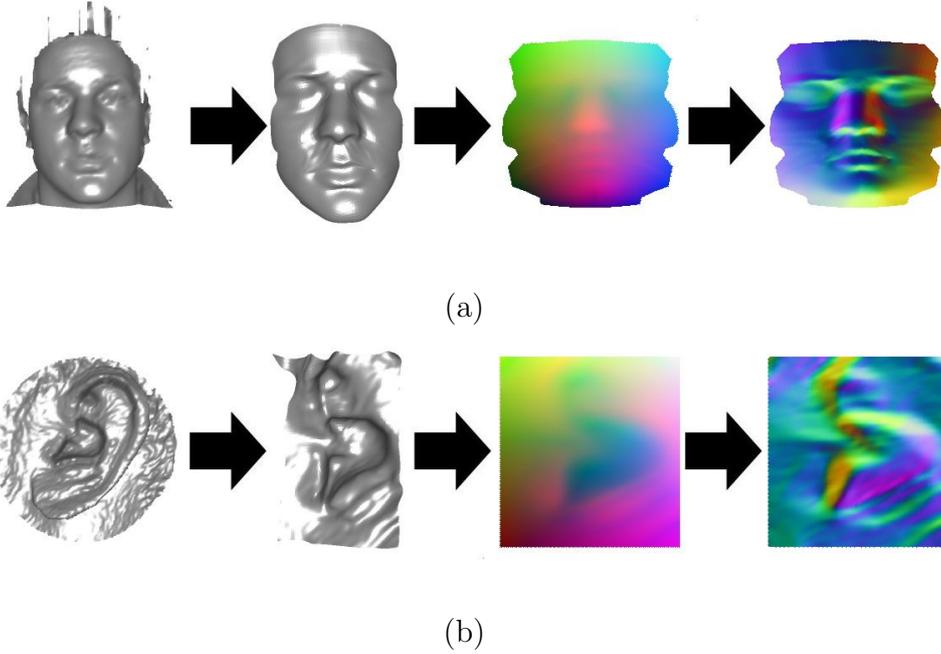


Fig. 4. From left to right, for face (a) and ear (b) datasets of the same subject: Raw data  $\rightarrow$  Fitted AFM  $\rightarrow$  Extracted geometry image  $\rightarrow$  Computed normal image. form the model so that it gradually acquires the shape of the raw data. The analytical equations are solved using an iterative Finite Element Method approximation.

**Analytical Formulation:** The basic equation of the deformable model framework is given by:

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

where  $\vec{q}$  is the control points vector,  $\mathbf{M}_q$  is the mass matrix,  $\mathbf{D}_q$  is the damping matrix,  $\mathbf{K}_q$  is the stiffness matrix, and  $f_q$  are the external forces. For data fitting purposes we used  $\mathbf{M} = \emptyset$  and  $\mathbf{D} = \emptyset$ . The stiffness matrix is the most important component as it resists the external forces and determines elastic properties of the model. It can be decomposed into three matrices  $\mathbf{K} = \mathbf{K}_{fo} + \mathbf{K}_{so} + \mathbf{K}_{sp}$ . The matrix  $\mathbf{K}_{fo}$  is related to the first order strain energy,  $\mathbf{K}_{so}$  to the second order strain energy and  $\mathbf{K}_{sp}$  is related to the spring forces energy:

$$E_{fo} = \frac{1}{2}\kappa_{fo}\vec{q}^T\mathbf{K}_{fo}\vec{q},$$

$$E_{so} = \frac{1}{2}\kappa_{so}\vec{q}^T\mathbf{K}_{so}\vec{q},$$

$$E_{sp} = \frac{1}{2}\kappa_{sp}\vec{q}^T\mathbf{K}_{sp}\vec{q}$$

where  $\kappa_{fo}$ ,  $\kappa_{so}$ ,  $\kappa_{sp}$  are the individual weights.

**Finite Element Method:** In our implementation, we employed the subdivision-based Finite Element Method approximation proposed by Mandal [15]. This approximation solves the above equations in an iterative way. We build a Loop subdivision surface using the AFM as the control mesh. We selected to use the Loop subdivision scheme [14] for two reasons: it produces a limit surface with  $C^2$  continuity and only 1-neighborhood area information is needed for each vertex. In the above equations the vector  $\vec{q}$  corresponds to the control mesh. However, the equation is evaluated using a vector  $\vec{Q}$  that corresponds to the limit surface. As shown by Mandal [15], by using the inverse subdivision matrices, the computed forces on  $\vec{Q}$  can be transferred to  $\vec{q}$ . The external forces  $f_q$  are also computed on the limit surface and then transferred to the control mesh. For each vertex of  $\vec{Q}$  the nearest neighbor on the raw data surface is found and this creates a deformation force proportional to their distance. Therefore, the raw data surface acts as an attractor to the annotated model and drives the deformation. Note that in order to reduce the computational cost of the nearest neighbor search we employed a space partitioning technique [8].

When the deformation concludes, the annotated model acquires the shape of the raw data. Since the deformation has not violated the properties of the original model, the deformed model can be converted to a geometry image.

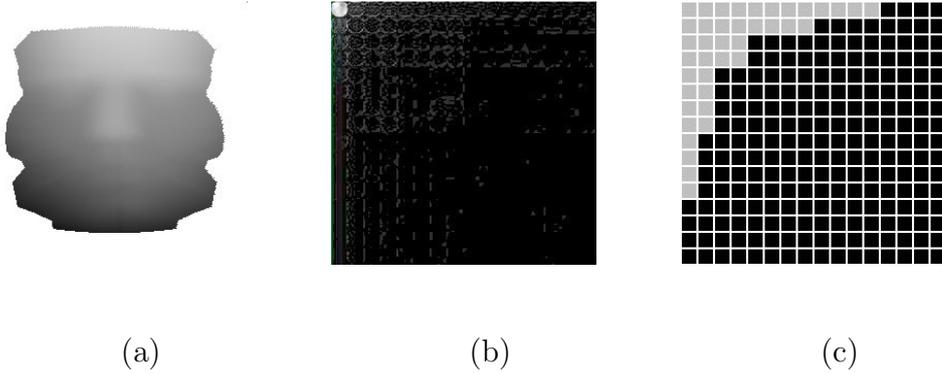


Fig. 5. Wavelet analysis of a facial geometry image. For visualization purposes, the coefficient’s magnitude is mapped as greyscale intensity: (a) Original image ; (b) Four level Walsh transform; (c) Mask that selects 15% of wavelet bands.

We also compute the normal image. This process is depicted in Fig. 4 for face and ear datasets of the same subject.

## 2.5 Wavelet Analysis

We apply a wavelet transform on the derived geometry and normal images in order to extract a descriptive and compact biometric signature. To this end, we employ the Walsh Transform [22]. Each channel of the geometry and normal image is treated as a separate image for the wavelet analysis. The Walsh wavelet transform for images is a decimated wavelet decomposition using tensor products of the full Walsh wavelet packet system. The 1D Walsh wavelet packet system is constructed by repeated application of the Haar filter bank, a two-channel multirate filter bank based on the Haar conjugate mirror filter. Both channels output the result of convolving a 1D discrete signal with a Haar filter and then downsampling by a factor of two. The low-pass and high-pass Haar filters are  $g$  and  $h$ , respectively:  $g = \frac{1}{\sqrt{2}}[1 \ 1]$  and  $h = \frac{1}{\sqrt{2}}[1 \ -1]$ . For images, we use tensor products of these 1D filters. This means that the

filter bank operations are applied separately to the rows and columns of the image, resulting in a four channel filter bank with channels LL, LH, HL, and HH (corresponding to the filters  $g^t * g$ ,  $g^t * h$ ,  $h^t * g$  and  $h^t * h$  respectively). In other words, channel LL (low-pass) captures the local averages of the image, and channels LH, HL and HH (high-pass) capture horizontal, vertical and diagonal edges, respectively. We recursively apply this decomposition to each of the four output channels to construct the full Walsh wavelet packet tree decomposition (see Fig. 5 (b)).

Conjugate mirror filter banks achieve perfect reconstruction, so the Walsh transform preserves all information originally present in the signal. However, for efficiency purposes we keep only certain coefficients that retain most of the energy of the image. We thus favor the low-pass bands of the packet tree decomposition (see Fig. 5 (c)). The selected coefficients are directly compared during identification without reconstructing the original image, using a weighted  $L^1$  distance metric. The weights are empirically selected and depend on the annotation of the face or ear model. For each modality, the biometric signature is the concatenation of the wavelet coefficients that were extracted from the geometry and normal images. For the multimodal fusion, the coefficients of each modality are concatenated using an empirical global normalization weight.

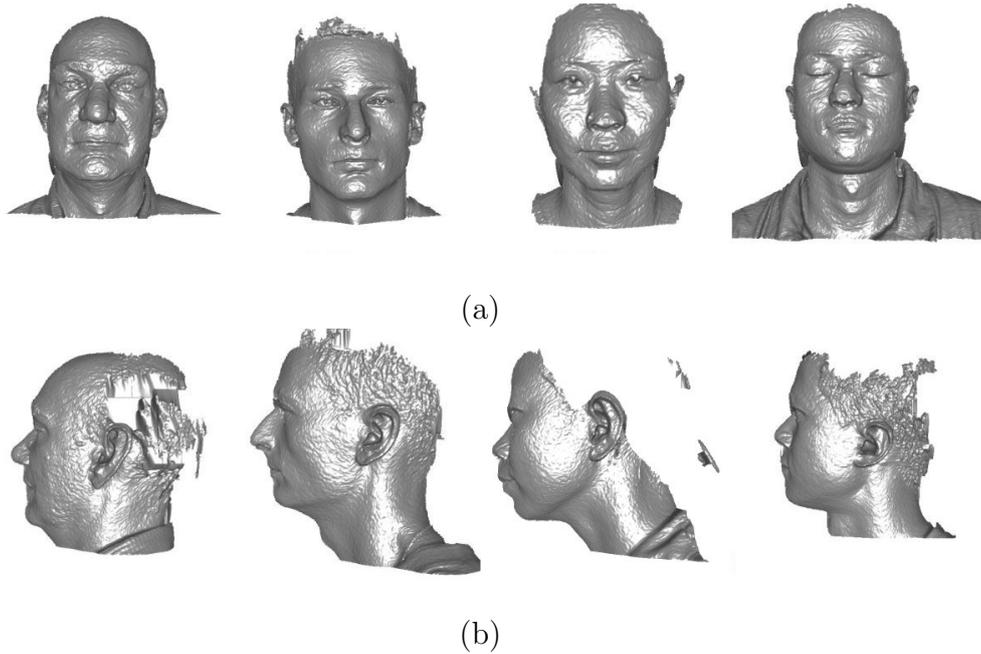


Fig. 6. Unprocessed multimodal data pairs from four subjects: (a) Facial data from FRGC v2; (b) Ear data from UND Ear.

### 3 Databases

**Face Database:** For facial data, we use the FRGC v2 database [19], the largest publicly available 3D face database. It contains a total of 4007 range images (e.g., Fig. 6(a)), acquired between 2003 and 2004. The hardware used to acquire these range data was a Minolta Vivid 900 laser range scanner, with a resolution of 640x480. These data were obtained from 466 subjects and contain various facial expressions (e.g., happiness, surprise). The subjects are 57% male and 43% female, and the age distribution is 65% 18-22 years old, 18% 23-27 and 17% 28 or over. The datasets are divided into gallery and probe sets, with one gallery per subject, resulting in 466 gallery and 3541 probes.

**Ear Database:** For ear data, we use the Ear Database from the University of Notre Dame (UND) [1], collections F and G. We used a subset defined by the makers of the database that contains 830 range images (e.g., Fig. 6(b)) from

415 subjects and the data were acquired from fall 2003 to fall 2004. Again, the Minolta laser scanner was used, with a resolution of 640x480. Note that compared to the facial datasets, the ear datasets are more prone to missing data and noise due to the elaborate structure of the human ear, which includes many concavities (see Fig. 6). The datasets are divided into gallery and probe sets, with one gallery per subject, resulting in 415 gallery and 415 probes.

**Multimodal Database:** In order to create a multimodal face and ear database we combined the above databases. Even though these databases were created separately, they share a lot of common subjects. Using the subject ID it is possible to identify facial and ear data that belong to the same subject. An example is given in Fig. 6, where four pairs of datasets from four subjects are shown. To perform multimodal experiments, we need to select subjects that have at least one probe and one gallery in each database. There are a total of 324 such subjects, resulting in a subset of FRGC v2 with 324 gallery and 3259 probe face datasets and a subset of UND Ear with 324 gallery and 324 probe ear datasets. We combined these datasets in two ways:

- We created 324 gallery and 324 probe multimodal datasets. Each dataset is a unique (face, ear) pair from the same individual. This multimodal database will be referred to as *MULTI1*.
- We created 324 gallery and 3259 probe multimodal datasets. In this case each ear probe was paired with multiple face probes of the same individual, since there were fewer ear probes than face probes. This multimodal database will be referred to as *MULTI2*.

Table 1

Rank-one recognition rates for two databases.

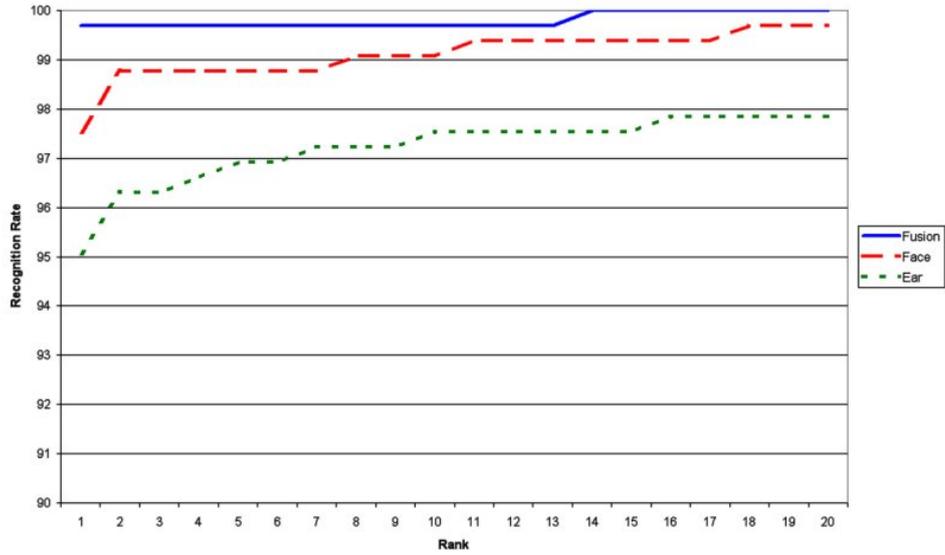
	<i>MULTI1</i>	<i>MULTI2</i>
Fusion	99.7%	99.4%
Face	97.5%	97.2%
Ear	95.0%	95.0%

## 4 Results

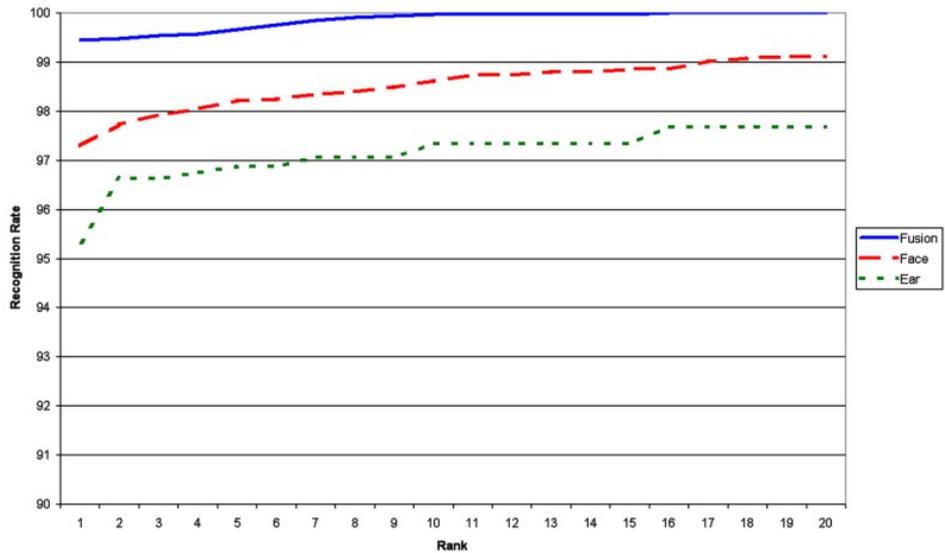
### 4.1 Performance

Using the gallery/probe division of our databases, we performed an identification experiment. The performance is measured using a Cumulative Match Characteristic (CMC) curve and the rank-one recognition rate is reported. For comparison purposes we also report the results for each modality separately.

The fusion of the face and ear performs significantly better than each modality as seen in Table 1. Also, the face modality performs better than the ear modality, despite the challenging nature of the FRGC v2 database. The importance of the multimodal fusion is more evident in Fig. 7. The fusion curve can reach 100% recognition rate before rank 15. On the contrary, neither single modality can reach 100% before rank 20. This indicates that the cases where the method completely fails to identify a subject are uncorrelated, therefore one modality can confute the shortcoming of the other.



(a)



(b)

Fig. 7. Recognition rates for fusion, face and ear modalities on two databases: (a) *MULTI1* (b) *MULTI2*.

#### 4.2 Correlation Analysis

In order to measure the correlation between the face and ear modalities we used the Pearson Correlation Coefficient [7]. This coefficient indicates the

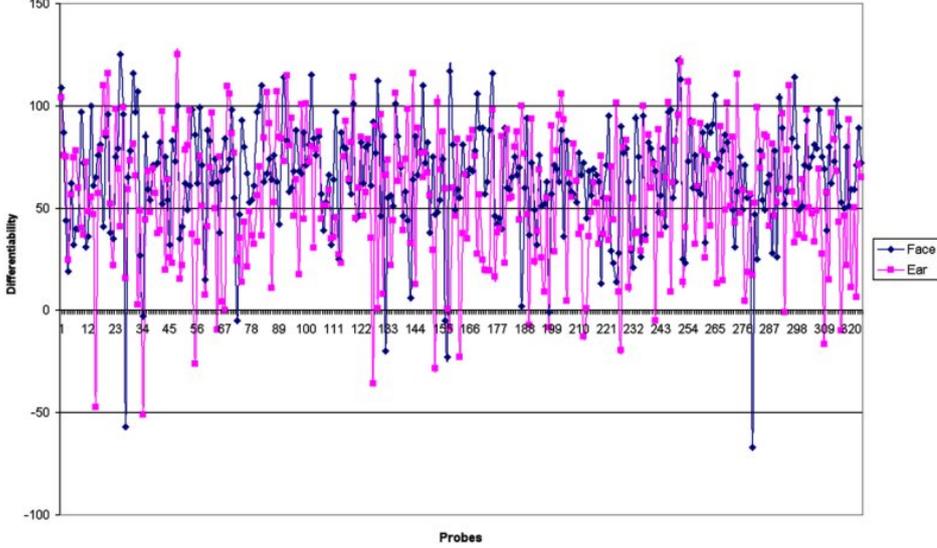


Fig. 8. Differentiability for the 324 probe datasets of *MULTII*. For each probe, negative values indicate that the closest gallery dataset does not belong to the same subject.

strength of a linear relationship between two random variables, and is given by

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x}_i)(y_i - \bar{y}_i)}{(n - 1)s_x s_y}$$

where  $x$  and  $y$  are two variables with  $n$  measurements,  $\bar{x}$  and  $\bar{y}$  are their mean values, and  $s_x$  and  $s_y$  their standard deviations. For positive correlations,  $r_{xy}$  can range from 0 to 1; higher values indicate stronger correlation.

For each of the face and ear modalities, we first computed the differentiability of the probes using database *MULTII*. This is a 1-dimensional array with size equal to the number of probes of the database. For each probe, the differentiability is defined as its distance from the nearest gallery dataset that belongs to a different subject minus the distance from the gallery dataset that belongs to the same subject. The higher this value is, the more differentiability we have, since the distance of the correct match to the nearest outlier increases. If the value is negative, this indicates a mismatch, as the outlier was closer than the

correct match. The differentiability graph for the face and ear modalities is depicted in Fig. 8. From this graph, it is shown that the negative spikes are not common for the face and ear (with only 2 exceptions), indicating small correlation.

This is confirmed by the Pearson Correlation Coefficient, as the correlation between the face and ear was 0.161. From a statistical point of view this correlation is considered small, which shows that these modalities are ideal for combination using our method<sup>1</sup>. For comparison purposes, we also computed the Pearson Correlation Coefficient between the face and the multimodal fusion, and between ear and the multimodal fusion. These correlations were 0.449 and 0.915 respectively. These values indicate strong correlation which was to be expected since the multimodal fusion is a linear combination of the two modalities.

## 5 Conclusions

We have presented a unified multimodal approach that seamlessly handles 3D face and ear data. Geometry images are obtained after a fitting process of an Annotated Face Model and an Annotated Ear Model. Wavelet coefficients are then extracted which provide a descriptive and compact biometric signature.

Using the largest publicly available database we presented state-of-the-art performance that reaches 99.7% rank-one recognition rate. Moreover, we show that there is a low correlation between the differentiability of 3D face and ear

---

<sup>1</sup> The computed correlation depends on both the datasets and the method used to compute the distances.

data. Future work, will focus on utilizing the whole human head as a single 3D modality that will include facial and ear structures.

## References

- [1] University of Notre Dame biometrics database, <http://www.nd.edu/>
- [2] K. B. B. Victor, S. Sarkar, An evaluation of face and ear biometrics, in: Proc. of International Conf. on Pattern Recognition (ICPR), 2002.
- [3] P. Besl, N. McKay, A method for registration of 3-D shapes, *IEEE Trans. on Pattern Analysis and Machine Intelligence* 14 (2) (1992) 239–256.
- [4] K. Chang, K. Bowyer, S. Sarkar, B. Victor, Comparison and combination of ear and face images in appearance-based biometrics, *IEEE Trans. on Pattern Analysis and Machine Intelligence* 25 (9) (2003) 1160–1165.
- [5] K. I. Chang, K. W. Bowyer, P. J. Flynn, Adaptive rigid multi-region selection for handling expression variation in 3D face recognition, in: Proc. of IEEE Workshop on Face Recognition Grand Challenge Experiments, San Diego, CA, 2005.
- [6] H. Chen, B. Bhanu, Human ear recognition in 3d, *IEEE Trans. on Pattern Analysis and Machine Intelligence* 29 (4) (2007) 718–737.
- [7] J. Cohen, *Statistical power analysis for the behavioral sciences*, 2nd ed., Hillsdale, NJ: Lawrence Erlbaum Associates, 1998.
- [8] A. Finkel, J. Bentley, Quad trees: A data structure for retrieval of composite keys, *Acta Informatica* 4 (1) (1974) 1–9.
- [9] X. Gu, S. Gortler, H. Hoppe, Geometry images, in: ACM SIGGRAPH '02, 2002.

- [10] D. Hurley, M. Nixon, J. Carter, A new force field transform for ear and face recognition, in: Proc. of IEEE International Conf. on Image Processing (ICIP), 2000.
- [11] M. Husken, M. Brauckmann, S. Gehlen, C. von der Malsburg, Strategies and benefits of fusion of 2D and 3D face recognition, in: Proc. of IEEE Workshop on Face Recognition Grand Challenge Experiments, San Diego, CA, 2005.
- [12] I. Kakadiaris, G. Passalis, G. Toderici, N. Murtuza, Y. Lu, N. Karampatziakis, T. Theoharis, 3D face recognition in the presence of facial expressions: An annotated deformable model approach, IEEE Trans. on Pattern Analysis and Machine Intelligence 29 (4) (2007) 640–649.
- [13] S. Kirkpatrick, C. Gelatt, M. Vecchi, Optimization by simulated annealing, Science 22 (4598) (1983) 671–680.
- [14] C. Loop, Smooth subdivision surfaces based on triangles, Master’s thesis, Department of Mathematics, University of Utah (1987).
- [15] C. Mandal, A dynamic framework for subdivision surfaces, Ph.D. thesis, University of Florida (1998).
- [16] T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West, G. Medioni, Performance of Geometrix ActiveIDTM 3D face recognition engine on the frgc data, in: Proc. of IEEE Workshop on Face Recognition Grand Challenge Experiments, San Diego, CA, 2005.
- [17] G. Papaioannou, E. Karabassi, T. Theoharis, Reconstruction of three-dimensional objects through matching of their parts, IEEE Trans. on Pattern Analysis and Machine Intelligence 24 (1) (2002) 114–124.
- [18] G. Passalis, I. Kakadiaris, T. Theoharis, Intra-class retrieval of non-rigid 3D objects: Application to face recognition, IEEE Trans. on Pattern Analysis and Machine Intelligence 29 (2) (2007) 218–229.

- [19] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the Face Recognition Grand Challenge (Jun. 2005).
- [20] E. Praun, H. Hoppe, Spherical parametrization and re-meshing, in: ACM SIGGRAPH '03, 2003.
- [21] P. Siarry, G. Berthiau, F. Durbin, J. Haussy, Enhanced simulated annealing for globally minimizing functions of many-continuous variables, *ACM Trans. on Mathematical Software* 23 (2) (1997) 209–228.
- [22] E. Stollnitz, T. DeRose, D. Salesin, *Wavelets for Computer Graphics: Theory and Applications*, Morgan Kaufmann Publishers, Inc, 1996.
- [23] D. Terzopoulos, J. Platt, A. Barr, K. Fleischer, Elastically deformable models, in: *ACM SIGGRAPH '87*, vol. 21(4), 1987.
- [24] G. Turk, M. Levoy, Zippered polygon meshes from range images, in: *ACM SIGGRAPH '94*, 1994.
- [25] D. Woodard, T. Faltemier, P. Yan, P. Flynn, K. Bowyer, Comparison of 3d biometric modalities, in: *Proc. of Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, 2006.
- [26] P. Yan, K. Bowyer, An automatic 3d ear recognition system, in: *Third International Symposium on 3D Data Processing, Visualization and Transmission*, University of North Carolina, Chapel Hill, 2006.
- [27] P. Yan, K. Bowyer, Biometric recognition using three-dimensional ear shape, *IEEE Trans. on Pattern Analysis and Machine Intelligence*.