# Using Facial Symmetry to Handle Pose Variations in Real-World 3D Face Recognition

Georgios Passalis[1,2], Panagiotis Perakis[1,2], Theoharis Theoharis[1,2]
and Ioannis A. Kakadiaris[2], *Senior Member, IEEE*

**Abstract**—The uncontrolled conditions of real-world biometric applications pose a great challenge to any face recognition approach. The unconstrained acquisition of data from uncooperative subjects may result in facial scans with significant pose variations along the yaw axis. Such pose variations can cause extensive occlusions resulting in missing data. In this paper, a novel 3D face recognition method is proposed that uses facial symmetry to handle pose variation. It employs an automatic landmark detector that estimates pose and detects occluded areas for each facial scan. Subsequently, an Annotated Face Model is registered and fitted to the scan. During fitting, facial symmetry is used to overcome the challenges of missing data. The results is a pose invariant geometry image. Unlike existing methods that require frontal scans, the proposed method performs comparisons among interpose scans using a wavelet-based biometric signature. It is suitable for real-world applications as it only requires half of the face to be visible to the sensor. The proposed method was evaluated using databases from the University of Notre Dame and the University of Houston that, to the best of our knowledge, include the most challenging pose variations publicly available. In these databases the average rank-one recognition rate of the proposed method was $83.7\,\%$.

**Index Terms**—O.8.13 Biometrics, I.5.4.d Face and gesture recognition, I.3.5.i Physically based modeling.

## 1 INTRODUCTION

O Wing to an increase in the availability of 3D data, several 3D face recognition approaches have been proposed. These approaches aim to overcome the limitations of 2D face recognition by offering pose invariance. However, they mostly use frontal 3D scans and assume that the entire face is visible to the sensor. This assumption is not always valid in real-world applications, since unconstrained acquisition may lead to facial scans with extensive occlusions that result in missing data. Therefore, to take advantage of the full pose invariance potential of 3D face recognition, the problem of missing data must be addressed.

In our previous work [1], [2], we presented a 3D face recognition method (ranked first in the shape-only section of NIST's Face Recognition Vendor Test 2006). However, only frontal scans were used for recognition as the method was not designed to handle missing data. In subsequent work [3], we extended our method to work in the presense of missing data by introducing a landmark detection algorithm (based on [4]).

In this paper, we extend and integrate our previous work to present a method that offers pose invariance and high recognition rates. It allows matching among interpose facial scans, and solves the missing data problem by using facial symmetry on occluded areas (Fig. 1). The

- 1: *Computer Graphics Laboratory, Department of Informatics and Telecommunications, University of Athens, Ilisia 15784, GREECE*

- 2: *Computational Biomedicine Lab, Department of Computer Science, University of Houston, Texas 77204, USA*
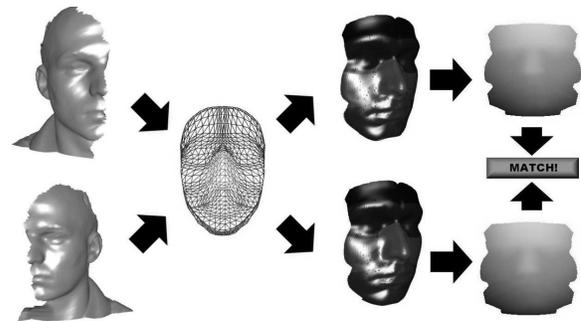
Fig. 1. Interpose matching using the proposed method (left to right): Opposite side facial scans with extensive missing data, generic AFM, deformed AFM for each scan (facial symmetry used) and extracted geometry images.

proposed method is applicable to real-world scenarios as it has the following unique combination of features:

- No user intervention is required: It is fully automatic.
- No subject cooperation is required: It can handle facial expressions and extreme pose variations.
- State-of-the-art performance: State-of-the-art performance has been experimentally demonstrated.
- Can be applied to large databases: It has a reasonable computational cost with excellent scalability.

Initial registration with the an Annotated Face Model (AFM) [2] using the detected landmarks allows the pose estimation of each facial scan. Then, the AFM is fitted to the facial scan using a subdivision-based deformable model framework that is extended to allow symmetric fitting. Symmetric fitting alleviates the missing data

problem and enables the creation of geometry images that are pose invariant. Compared to our previous work [3], we have made the following improvements:

- A far more robust automatic 3D facial landmark detector is used [5].
- Improved registration and fitting. These steps now independently handle the left and right sides of frontal scans to allow partial matching.
- A novel interpose database is introduced for additional experimental evaluation
- Significantly better results are presented indicating the proposed method's high accuracy in challenging databases.

The rest of this paper is organized as follows: in Section 2 the related work is described while in Section 3 the proposed method is presented in detail. In Section 4 the method's performance is evaluated and in Section 5 the method is summarized.

## 2 RELATED WORK

### 2.1 Face Recognition and Landmark Localization

Most face recognition methods focus only on frontal scans (see surveys of Bowyer *et al.* [6] and Chang *et al.* [7]). As a result, the performance of such methods is not evaluated with data that exhibit significant pose variations. The methods that are evaluated using data with pose variations are mentioned below. Note that none of them handles the extreme pose variations and the extensive missing data that the proposed method does.

Lu *et al.* [8], [9], [10] have presented methods to locate the positions of eye and mouth corners, and nose and chin tips, based on a fusion scheme of the shape index on range maps and the "cornerness" response on intensity maps. They also developed a heuristic method based on cross-profile analysis to locate the nose tip more robustly. Candidate landmark points were filtered out using a static (non-deformable) statistical model of landmark positions (in contrast to our approach). Although they report a 90 % rank-one recognition rate in an identification experiment, no claims where made with respect to the effects of pose variation in Face Recognition. Their proposed 3D approach [9] (evaluated using multiview scans with yaw rotations up to $45°$ from MSU) that can handle pose variations had a significant decrease in 3D landmark detection accuracy compared to their multimodal approach [8] (evaluated using near frontal scans from FRGC v1).

Dibeklioglu *et al.* [11], [12] introduced a nose tip localization and segmentation method using curvature-based heuristic analysis to enable pose correction in a face recognition system that allows identification under significant pose variations. However, a limitation of the method is that it is not applicable to facial scans with yaw rotations greater than $45°$. Additionally, even though the Bosphorus database that was used consists of

3,396 facial scans, these scans were obtained from 81 subjects only. Blanz *et al.* [13], [14] presented methods on 3D face reconstruction by fitting their 3D Morphable Model on 3D facial scans. Their method is a well established approach for producing 3D synthetic faces from scanned data. However, face recognition testing is performed on FRGC database with frontal facial scans, and on FERET database with faces under pose variations which do not exceed $40°$ yaw rotation.

Bronstein *et al.* [15] presented a face recognition method that can handle missing data. Their method is based on their previous work where they used a canonical representation of the face. They report high recognition rates on a limited database of 30 subjects. In addition, the database that was used had no side scans. The scans with missing data that was used were derived synthetically by randomly removing areas from frontal scans. In Nair and Cavallaro's [16] work on partial 3D face matching, the face is divided into areas and only certain areas are used for registration and matching. The method is based on the assumption that the areas of missing data can be excluded. Using a database of 61 subjects, they show that using parts of the face rather than the whole face yielded higher recognition rates. However, both this method and their subsequent work on 3D landmark detection [17], their method is not applicable to missing data resulting from self-occlusion, particularly in the presence of holes around the nose region.

Lin *et al.* [18] introduced a coupled 2D and 3D feature extraction method to determine the positions of eye sockets using curvature analysis. The nose tip is considered as the extreme vertex along the normal direction of eye sockets. The method was used in an automatic 3D face authentication system but was only tested on data from 27 subjects with various poses and expressions. Mian *et al.* [19] introduced a heuristic method for nose tip detection and used it in a face recognition system. The method was based on a geometric analysis of the nose ridge contour projected on the $x - y$ plane. A preprocessing step to crop and pose correct the facial data was used. This method can only be applied to data with up to $90°$ roll variation and yaw and pitch variation less than $15°$, thus limiting the applicability to near frontal scans.

Methods have also been proposed that focus mainly on the detection of 3D facial landmarks. Segundo *et al.* [20] presented a face and facial feature detection method by combining a method for 2D face segmentation on depth images with surface curvature information for detecting facial features (e.g., eye corners and nose tip). The method was tested on the FRGC 2.0 data with over 99.7 % detection rate. However, this method had problems in nose and eye corner detection when the face had a significant pose variation ($> 15°$ around the yaw and roll axes). Wei *et al.* [21] proposed a nose tip and nose bridge localization method to determine facial pose. The method was based on a Surface Normal Difference

algorithm and shape index estimation, and was used as a preprocessing step in pose-variant systems to determine the pose of the face. They reported an angular error of the nose tip - nose bridge segment less than $15°$ in $98\%$ of the 2500 datasets of BU-3DFE facial database, which contains complete frontal facial scans with up to $\pm45°$ yaw variation.

Faltemier *et al.* [22] proposed a method called Rotated Profile Signatures, based on profile analysis, to robustly locate the nose tip in the presence of pose, expression and occlusion variations. Their method was tested against the NDOff2007 database which contains 7,317 facial scans (406 frontal and 6,911 in various yaw and pitch angles). They reported a $96\%$ to $100\%$ success rate with distance error threshold $10\,mm$. Although their method achieved high success rate scores, the exact localization distance error results were not presented. Xu *et al.* Finally, [23] presented a feature extraction hierarchical scheme that detected the positions of nose tip and nose ridge. They introduced the effective energy measure to describe the local distribution of neighboring points and detect the candidate nose tips. Although it was tested against various databases, no exact localization results were provided.

## 2.2 Facial Asymmetry

The human face is not perfectly symmetrical and the exact level of facial asymmetry was recently quantified in the work of Liu *et al.* [24]. In this work it was shown that facial asymmetry is statistically significant given a reasonable range of sensor noise. Additionally, facial asymmetry has been used as a biometric in several methods (e.g., Kompanets [25], Liu *et al.* [26] and Mitra *et al.* [27]). In these methods, facial asymmetry offered promising biometric results particularly in the presence of facial expressions. As pointed out by Liu *et al.* [24], facial asymmetry should not be ignored without a justification.

The proposed method, exploits facial symmetry and does not assume the human face to be perfectly symmetrical. The method is based on the assumption that the difference (caused by facial asymmetry) between the left and the right region of a subject's face is less than the difference between these regions and the regions of another subject's face. The experimental results presented in Section 4 justified that assumption for the databases that were used. The concept of using partial facial data for biometric purposes has also been investigated by Gutta *et al.* [28] in the 2D face recognition domain with promising results.

## 3 UR3D-S: A SYMMETRIC FACE RECOGNITION METHOD

The proposed method, UR3D-S, processes each facial scan and derives a wavelet-based biometric signature. This representation is a biometric signature that can
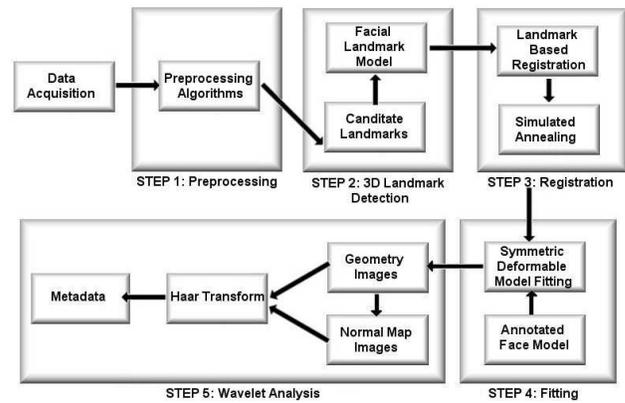


Fig. 2. Depiction of the UR3D-S flow chart.

be directly compared with other signatures using an $L_1$ distance metric, allowing efficient matching in both identification and verification scenarios. The novelty of UR3D-S is that the signature is independent of the initial pose and of missing data caused by occlusions (as long as half of the face with respect to the yaw axis is visible in the scan). Specifically, in order to perform interpose matching we require that the following five landmarks are visible on the same side of the face: inner and outer eye corner, nose tip, mouth corner and chin tip. This allows seamless comparisons among frontal, left and right side scans, making UR3D-S suitable for real-life biometric applications. The processing pipeline of each facial scan consists of the following fully automated steps (Fig. 2):

Step 1 *Preprocessing:* Standard preprocessing techniques are used to filter the raw data.

Step 2 *3D Landmark Detection:* A robust landmark detector is used for pose estimation (determining if it is a frontal, left or right scan).

Step 3 *Registration:* The raw data are registered to the AFM using a two–stage approach.

Step 4 *Symmetric Deformable Model Fitting:* The AFM is fitted to the data using facial symmetry. The fitted model is then converted to a geometry image and a normal image.

Step 5 *Wavelet Analysis:* A wavelet transform is applied on the geometry and normal images and the wavelet coefficients are stored as a biometric signature.

### 3.1 Preprocessing

UR3D-S can use as input both polygonal and range data obtained from optical or laser scanners. A preprocessing step is applied to the raw data to convert them to a unified representation and to eliminate any sensor-specific problems [2]. Particularly for the range data, the following preprocessing algorithms were applied before converting the data to a polygonal representation:

- *Median Cut:* To remove spikes (that are common in range images from laser scanners) a median cut
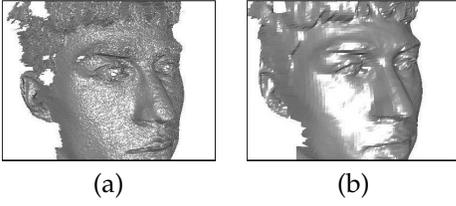
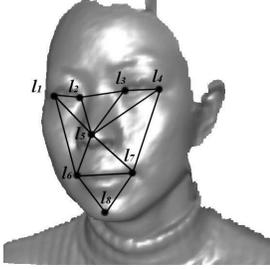Fig. 3. Example of a facial scan (a) before and (b) after preprocessing.



Fig. 4. Depiction of the landmark model (FLM8) overlaid on a frontal facial scan.

filter with a $3 \times 3$ window was applied.

- *Hole Filling:* Laser scanners usually produce holes in certain areas (e.g., eyes, eyebrows). A hole filling procedure that uses bilinear interpolation was applied.
- *Smoothing:* A smoothing filter with a $3 \times 3$ window was applied to remove white noise.
- *Subsampling:* The range data were subsampled at a $1 : 4$ ratio.

An example of range data suffering with noise and holes is shown in Fig. 3, before (a) and after (b) preprocessing.

### 3.2 3D Landmark Detection

Our method for 3D landmark detection and pose estimation (described in detail in [5]) uses 3D information to extract candidate interest points which are identified and labeled as landmarks by matching them with a Facial Landmark Model (FLM) [4], [3]. Once anatomical landmarks are localized, then the corresponding rigid transformation is computed in order to register the facial scans.

We use a set of eight anatomical landmarks: right eye outer corner ($l_1$), right eye inner corner ($l_2$), left eye inner corner ($l_3$), left eye outer corner ($l_4$), nose tip ($l_5$), mouth right corner ($l_6$), mouth left corner ($l_7$) and chin tip ($l_8$) (Fig. 4). Notice that five of these points are visible on profile and semi-profile facial scans. Thus, the complete set of eight landmarks can be used for frontal and almost-frontal faces and two subsets of five landmarks each can be used for semi-profile and profile faces. The right side landmark set contains the landmarks $l_1$, $l_2$, $l_5$, $l_6$, and $l_8$, and the left side the landmarks $l_3$, $l_4$, $l_5$, $l_7$ and $l_8$.

Each of these sets of landmarks constitute a corresponding Facial Landmark Model (FLM). For the re-

maining of this paper, the model of the complete set of eight landmarks will be referred to as FLM8 and the two reduced sets of five landmarks (left and right) as FLM5L and FLM5R, respectively. The FLMs are created in the following manner:

- A statistical mean shape for each landmark set (FLM8, FLM5L and FLM5R) is computed from a training set of 150 manually annotated facial scans with neutral expressions, randomly chosen from the FRGC v2 database.
- Variations of each FLM are computed using Principal Component Analysis (PCA).

#### 3.2.1 The Facial Landmark Models

The mathematical representation of an $n$-point shape in $d$ dimensions can be defined by concatenating all landmark coordinates into a $k = nd$ vector and establishing a *shape space* [29], [30], [31]. Since shape is invariant to Euclidean similarity transformations, translational, scale and rotational effects need to be filtered out. This is accomplished by establishing a common coordinate reference to which all landmark shapes are aligned; a procedure commonly known as *Procrustes Analysis*. Since for the purposes of this paper the size of the shape is of great importance, it is not filtered out by scaling shapes to unit size. Alignment is performed by minimizing the *Procrustes distance* $D_P^2 = |\mathbf{x}_i - \mathbf{x}_m|^2$ of each shape $\mathbf{x}_i$ to the mean shape $\mathbf{x}_m$. The alignment procedure is used to compute the mean shape $\mathbf{x}_m$ of landmarks, which is the *Procrustes mean* $\mathbf{x}_m = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i$ of all $N$ example shapes $\mathbf{x}_i$ .

Aligned shape vectors form a distribution in the shape space, thus, we can model this distribution by estimating a vector $\mathbf{b}$ of parameters that describes shape's deformations [32], [31], [33], [30]. This shape decomposition is performed by applying PCA to the aligned shapes.

If $\mathbf{A}$ contains (in columns) the $p$ eigenvectors $\mathbf{A}_i$ corresponding to the $p$ largest eigenvalues $\lambda_i$ of the covariance matrix $\mathbf{C}$ of the aligned shape vectors, then $\mathbf{b}$ is a $p$-dimensional vector given by $\mathbf{b} = \mathbf{A}^T \cdot (\mathbf{x} - \mathbf{x}_m)$. The vector $\mathbf{b}$ is the projection of $\mathbf{x}$ onto the subspace spanned by the $p$ most significant eigenvectors of the eigenspace (principal components). By applying limits to each $b_i$ (e.g., $|b_i| \leq 3\sqrt{\lambda_i}$) we can create marginal mean shape deformations. The number $p$ of most significant eigenvectors and eigenvalues to retain (modes of variation) can be chosen so that the model represents a given proportion of the total variance of the data. The result is the FLM [4], [3].

We incorporated 15 eigenvalues (out of the total 24) in FLM8, which represent 99 % of total shape variations of the complete landmark shapes. We also incorporated seven eigenvalues (out of the total 15) in FLM5L and FLM5R, which represent 99 % of total shape variations of the FLM5L and FLM5R.

General-purpose feature detection methods do not identify and label the detected candidate landmarks. It

is clear that some topological properties of facial landmarks must be taken into consideration. Thus, candidate landmarks, irrespective of the way they are produced, must be consistent with the corresponding FLM. This is done by fitting a candidate landmark set $\mathbf{y}$ to the FLM $\mathbf{x}_m$ by minimizing the Procrustes distance $|\mathbf{y} - \mathbf{x}_m|$ in a simple iterative approach. By projecting $\mathbf{y}$ to the shape eigenspace, the deformation parameters $\mathbf{b}$ are determined, and the landmark shape is considered as plausible, if it is consistent with marginal shape deformations [31], [33].

### 3.2.2 Landmark Detection & Selection

To detect landmark points, we have used two 3D local shape descriptors that exploit the 3D geometry-based information of facial scans: *shape index* and *spin images*.

The shape index is extensively used for 3D landmark detection [34], [35], [8], [9], [10]. It is a continuous mapping of principal curvature values ($k_{max}$, $k_{min}$) of a 3D object point $\mathbf{p}$ into the interval [0,1], according to the formula:

$$SI(\mathbf{p}) = \frac{1}{2} - \frac{1}{\pi} tan^{-1} \frac{k_{max}(\mathbf{p}) + k_{min}(\mathbf{p})}{k_{max}(\mathbf{p}) - k_{min}(\mathbf{p})} .$$

The SI values represent the type of local curvatures of shapes (Cup = 0.0, Rut = 0.25, Saddle = 0.5, Ridge = 0.75, and Cap = 1.0).

Local maxima (Caps) are candidate landmarks for nose tip and chin tip and local minima (Cups) are candidate landmarks for eye corners and mouth corners. Local maxima and minima are located on a facial dataset and sorted in descending order of shape index values. The most significant subset of points for each group is retained (a maximum of 512 Caps and 512 Cups). In Fig. 5(a), black boxes represent Caps, and white boxes Cups.

However, our experiments indicated that the shape index value alone is not sufficiently robust for detecting anatomical landmarks on facial scans in a variety of poses. Therefore, candidate landmarks estimated from shape index values are further classified and filtered out according to their relevance with corresponding spin image templates.

A spin image encodes the coordinates of points on the surface of a 3D object with respect to a so-called *oriented point* $(\mathbf{p}, \mathbf{n})$, where $\mathbf{n}$ is the normal vector at a point $\mathbf{p}$ of a 3D object's surface [36]. A spin image at an oriented point $(\mathbf{p}, \mathbf{n})$ is a 2D grid accumulator of 3D points, as the grid is rotated around $\mathbf{n}$ by $360°$. Thus, a spin image is a descriptor of the global or local shape of the object, invariant under rigid transformations. For our purposes of representing facial features on facial scans, a $16 \times 16$ spin image grid with $2\ mm$ bin size was used. This represents the local shape spanned by a cylinder of $3.2\ cm$ height and $3.2\ cm$ radius.

In order to identify interest points on facial scans, we create spin image templates that represent the classes of the landmarks used. Notice that due to the symmetry of
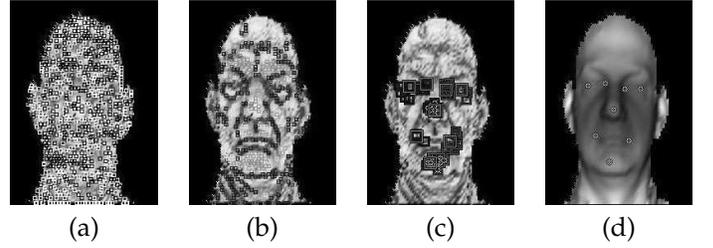


Fig. 5. Depiction of landmark detection procedure for face registration: (a) shape index extrema; (b) spin image classification; (c) consistent landmark sets and (d) best landmark set.

the face, only five classes (which represent the eye outer corner, eye inner corner, nose tip, mouth corner and chin tip landmarks) need to be created. Spin image templates are generated using spin image grids of landmark points from 975 available manually annotated frontal facial scans from FRGC v2 database, and represent the mean spin image grid associated with each of the five classes of the landmarks used. These spin image templates are generic and can be used on other test databases with no additional training.

Instead of searching all points of a facial dataset to determine the correspondence with the spin image templates, we use the candidate landmark points derived from the shape index. Thus, local maxima and minima of the shape index map (Caps and Cups) are further classified into five classes (eye outer corner, eye inner corner, nose tip, mouth corner and chin tip) according to the similarity $S(P, Q(T))$ of their spin images $P$ with the spin image templates $Q$ that represent each landmark class:

$$S(P, Q) = \frac{N \sum p_i q_i - \sum p_i \sum q_i}{\sqrt{[N \sum p_i^2 - (\sum p_i)^2][N \sum q_i^2 - (\sum q_i)^2]}} ,$$

where $p_i$, $q_i$ denotes each of the $N$ elements of spin images $P$ and $Q$, respectively [36].

The landmarks of the five spin image classes are sorted in descending order of significance according to their similarity measure with their corresponding spin image template. The most significant subset for each landmark class is retained (a maximum of 128 eye outer corners, 64 eye inner corners, 64 nose tips, 256 mouth corners and 128 chin tips). In Fig. 5(b), boxes of various shades of grey represent the landmarks of the five spin image classes. Notice that some of the classified landmark boxes overlap due to similarity with different templates.

For each facial scan, the procedure for landmark detection, labeling and selection has the following steps (Fig. 5):

1. Extract candidate landmarks from the Shape Index map.
2. Classify candidate landmarks by matching them with the corresponding Spin Image templates.
3. Create feasible combinations of five landmarks from the candidate landmark points.
4. Compute the rigid transformation that best aligns

the combinations of five candidate landmarks with the FLM5R and FLM5L.

5. Filter out those combinations that are not consistent with FLM5L or FLM5R by applying the fitting procedure as previously described.

6. Sort consistent right (FLM5R) and left (FLM5L) landmark sets in descending order according to a distance metric from the corresponding FLM.

7. Fuse accepted combinations of five landmarks (left and right) in complete landmark sets of eight landmarks.

8. Compute the rigid transformation that best aligns the combinations of eight landmarks with the FLM8.

9. Discard combinations of landmarks that are not consistent with the FLM8 by applying the fitting procedure as previously described.

10. Sort consistent complete landmark sets in descending order according to a distance metric from the FLM8.

11. Select the best combination of landmarks (consistent with FLM5R, FLM5R or FLM8) based on the distance metric to the corresponding FLM.

12. Obtain the corresponding rigid transformation for registration.

In Fig. 5(c), the landmark sets consistent with the FLM5R, FLM5L and FLM8 are depicted as boxes. Notice that some of the consistent landmarks overlap. Also note that the FLM8 consistent landmark set is not always the best solution; FLM5L and FLM5R are usually better solutions for side facial scans.

In order to find the best solution, and compare the consistent solutions we need a distance measure from the corresponding FLM. Since FLM5R, FLM5L, FLM8 have different dimensions in shape space, we must use a distance measure that is independent of shape space dimension. For this purpose, we define an intuitive combined *normalized Procrustes* and *spin similarity* distance:

$$D_{NPSS} = D_{NP} \cdot D_{SS} = \frac{D_P}{k^2} \cdot \frac{1}{2} \left[ 1 - \frac{\sum_{i=1}^{n} S(P_i, Q_i)}{n} \right] ,$$

where $D_{NP}$ is the normalized Procrustes distance, $D_{SS}$ is the mean spin image similarity distance (normalized to [0,1]), $D_P$ is the Procrustes distance, $S(P_i, Q_i)$ is the similarity measure between the landmark spin image $P_i$ and the corresponding template $Q_i$, $k$ is the shape space dimension, and $n$ is the number of landmarks. The division by $k^2$ instead of $k$ is preferred to give a bias to a complete solution.

Finally, using the selected best solution, the registration transformation is computed, the yaw-angle is estimated, and the facial scan is classified according to pose as frontal, left side or right side.

Note that the use of landmark sets of five landmarks serves two purposes: (i) it is the potential solution for semi-profile and profile faces and (ii) it reduces the combinatorial search space for creating the complete landmark sets in a divide-and-conquer manner. Instead
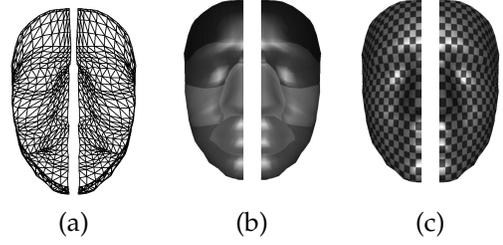


Fig. 6. Left & right sides of the AFM: (a) control mesh, (b) facial area anotation and (c) UV parametrization.

of creating 8-tuples of landmarks out of N candidates, which generates $N^8$ combinations to be checked for consistency with the FLMs, we create 5-tuples of landmarks, and check $N^5 + N^5 = 2N^5$ combinations for consistency with FLM5L and FLM5R. We retain 256 landmark sets consistent with FLM5L and 256 landmark sets consistent with FLM5R. By fusing them and checking consistency with FLM8 we have an extra of $256 \times 256$ combinations to be checked. Thus, by this approach $2N^5 + 256^2 \ll N^8$ combinations are checked, with $O(N^5) \ll O(N^8)$. For $N = 128$ we have approximately $69 \times 10^9$ instead of $72 \times 10^{15}$ combinations to be checked. Also, since an exhaustive search of all possible combinations of the candidate five landmarks is not feasible, simple length constraints from the corresponding FLM and its deformations are used to reduce the search space (pruning).

## 3.3 Annotated Face Model

In all the subsequent steps of the proposed method (registration, fitting and wavelet analysis), the *Annotated Face Model* (AFM) [2] is used. It is an anthropometrically correct 3D model of the human face [37]. The AFM was constructed only once and consists of a polygonal representation (later used as the control mesh of the subdivision surface), a facial area annotation and a UV parametrization (Fig. 6).

The UV parametrization allows the conversion of the AFM into an equivalent representation called *geometry image*. A geometry image is the result of mapping all the vertices of a 3D object ($x$, $y$ and $z$ coordinates) to a 2D grid representation ($u$, $v$ coordinates) [38]. Thus, a geometry image is the regular sampling of a 3D model represented as a 2D image, with each $u$, $v$ pixel corresponding to the original $x$, $y$, $z$ coordinates. Two dimensional geometry images have at least three channels assigned to each pair of $u$, $v$ coordinates encoding geometric information ($x$, $y$, $z$ coordinates and/or normals). Note that the UV parametrization of the AFM offers an injective mapping from a sphere in $R^3$ to a plane in $R^2$. This property is not violated even if its vertices are deformed, thus allowing the creation of a geometry image from a deformed AFM. Since the AFM is a topologically open model, a simple cylindrical mapping technique was used to create the UV parametrization. For a topologically closed and genus zero model (suitable for the full human
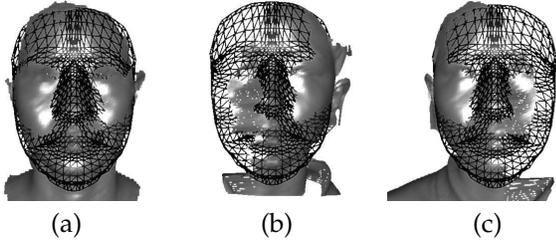
Fig. 7. AFM (black wireframe) and facial scans (grey polygons) superposed after registration: (a) frontal scan; (b) $45°$ left side scan; and (c) $45°$ right side scan.

head) Praun and Hoppe's octahedron-based parameterization [39] would be more appropriate.

## 3.4 Registration

To fit the AFM to each facial scan, they must both be defined in the same coordinate system (Fig. 7). To this end, we register the facial scans with the AFM using a two–stage approach. First, an initial registration is obtained from the landmarks detected in Section 3.2. Second, an algorithm based on Simulated Annealing is used to fine tune the registration.

The initial registration of each facial scan to the AFM can be accomplished by minimizing the Procrustes distance between a set of landmark points $\mathbf{x}$ on the scan and the corresponding landmark points $\mathbf{x}_0$ on the AFM in an iterative approach. If $\mathbf{T}$ translates $\mathbf{x}$ so that its centroid is at the origin (0,0,0), $\mathbf{T}_0$ translates $\mathbf{x}_0$ so that its centroid is at the origin (0,0,0), and $\mathbf{R}$ is an optimal rotation that minimizes the Procrustes distance of $\mathbf{x}$ to the reference shape $\mathbf{x}_0$, then, the final transformation to register a facial scan with vertices $\mathbf{v}_i$ to the AFM is:

$$\mathbf{v}'_i = \mathbf{T}_0^{-1} \cdot \mathbf{R} \cdot \mathbf{T} \cdot \mathbf{v}_i \ .$$

The landmark set detected on a facial scan (frontal, right or left) determines the model (FLM8, FLM5R or FLM5L) that will be used to aid registration with the AFM. However, in practice, when a frontal scan is detected, instead of using the FLM8, the frontal scan is considered as a pair of side scans and two independent registrations using FLM5R and FLM5L are computed. In this case the remaining steps of the method (Sections 3.5 and 3.6) are repeated twice, and two independent signatures are finally derived.

To improve the registration we use the algorithm presented by Papaioannou *et al.* [40] that uses a global optimization technique (Simulated Annealing [41], [42]) applied to depth images. The Simulated Annealing process minimizes the following objective function:

$$O = \sum_{i=1}^{r} \sum_{j=1}^{r} |Z_m(i,j) - Z_d(i,j)| \ ,$$

where $Z_m$ and $Z_d$ are the Z-buffers of the model and data respectively (normalized to $[0,1]$) and $r$ is the spatial resolution of the buffers. For side scans, only one half of

the model's z-buffer is used in the objective function. The other half is excluded as it would have been registered with areas that may have missing data. Since we assume that the initial registration provides a valid approximation, the Simulated Annealing algorithm is only allowed to produce limited translations and rotations. It is only used to fine-tune the registration; it does not alleviate errors caused by erroneous landmark detection.

## 3.5 Symmetric Deformable Model Fitting

We use the subdivision-based deformable model framework (presented in [2]) to fit the AFM to each facial scan (already registered by the previous step). During fitting, the AFM deforms in order to capture the shape of the facial scan. The forces that drive this deformation are called *external forces*. The forces that resist this deformation are called *internal forces* and correspond to the elastic properties of the model's surface (e.g., strain energy and material stiffness).

We have modified the framework in [2] by incorporating the notion of *symmetric fitting* to handle missing data. The fitting step can now handle the left and right sides of the AFM independently (Fig. 6 (a)). The idea is that facial symmetry can be used to avoid the computation of the external forces on areas of possible missing data. The internal forces are not affected and remain unmodified thereby ensuring the continuity of the fitted surface. As a result, when fitting the AFM to facial scans classified as left side, the external forces are computed on the left side of the AFM and mirrored to the right side (and vice versa for right side scans). This can be applied also to frontal scans since they are handled as a pair of independent left and right side scans (as stated in Section 3.4). Therefore, for each frontal scan, two fitted AFMs are computed: one that has the left side mirrored to the right and another that has the right side mirrored to the left.

The basic equation of the deformable model framework is given by Newton's second law:

$$M_q \frac{d^2\mathbf{q}}{dt^2} + D_q \frac{d\mathbf{q}}{dt} + K_q\mathbf{q} = f_\mathbf{q} \ .$$

The term $\mathbf{q}$ is the control points vector that determines the degrees of freedom of the AFM (each point has three degrees of freedom). The term $M_q$ is the mass matrix and is multiplied with the acceleration vector in order to control the kinetic energy. The damping matrix $D_q$ is multiplied with the velocity vector in order to control the energy dissipation. For data fitting purposes we set $M_q \frac{d^2\mathbf{q}}{dt^2} = 0$ and $D_q \frac{d\mathbf{q}}{dt} = 0$, as they represent the translational effects of the external forces. The term $f_\mathbf{q}$ represents the external forces vector; during fitting it consists of forces that pull the control points vector towards the surface of the facial scan. The term $K_q$ is the stiffness matrix and determines the elastic properties of the AFM that resist the external forces. It can be decomposed into three matrices $K_q = K_\alpha + K_\beta + K_\gamma$. $K_\alpha$ is related to the first order strain energy, $K_\beta$ to the second

order strain energy and $K_\gamma$ is related to the spring forces energy:

$$E_\alpha = \tfrac{1}{2}\kappa_\alpha \mathbf{q}^T K_\alpha \mathbf{q} \ ,$$
$$E_\beta = \tfrac{1}{2}\kappa_\beta \mathbf{q}^T K_\beta \mathbf{q} \ ,$$
$$E_\gamma = \tfrac{1}{2}\kappa_\gamma \mathbf{q}^T K_\gamma \mathbf{q} \ ,$$

where $\kappa_\alpha$ ,$\kappa_\beta$ and $\kappa_\gamma$ are the individual weights.

The analytical equations are solved using an iterative Finite Element Method (FEM) approximation. In our implementation, we employed the subdivision-based FEM approximation proposed by Mandal [43]. This approximation solves the above equations in an iterative manner. The AFM is used as the control mesh of a subdivision surface. At each step, the internal and external forces are computed on the limit surface and, by using the inverse subdivision matrices, are transferred to the control mesh (**q**). Details of this implementation have been presented in [2]. The key aspects of the implementation are outlined below:

- The resolution of the control mesh determines the degrees of freedom but does not affect the accuracy of the approximation (which is determined by the resolution of the limit surface).
- The Loop subdivision scheme [44] is used because it produces a limit surface with $C^2$ continuity, and only 1-neighborhood area information is needed for each vertex.
- Computing the external forces requires multiple nearest neighbor searches between the AFM and the surface of the facial scan. To decrease the computational cost a space partitioning technique (Octrees [45], [46]) is used.

When the deformation stops, the annotated model acquires the shape of the raw data. Since the deformation has not violated the injective nature of the AFM's UV parametrization, the deformed AFM can be converted to a geometry image. The normal image which is the first spatial derivative of the geometry image is also derived.

### 3.6 Wavelet Analysis

We apply a wavelet transform on the derived geometry and normal images in order to extract a descriptive and compact biometric signature. As explained in Section 3.5, the derived geometry and normal images represent the full face even if half of the facial data are missing from the scan. When facial symmetry is used (for side scans) the information in the geometry and normal images is redundant, because half of it is the mirror of the other half. However, we keep both sides in order to have a common representation that is independent of the initial pose.

Each channel of the geometry and normal image is treated as a separate image for wavelet analysis (resulting in six channels, three for each image type). The Walsh wavelet transform [47] for images is a decimated wavelet decomposition using tensor products of the full Walsh wavelet packet system. The 1D Walsh wavelet
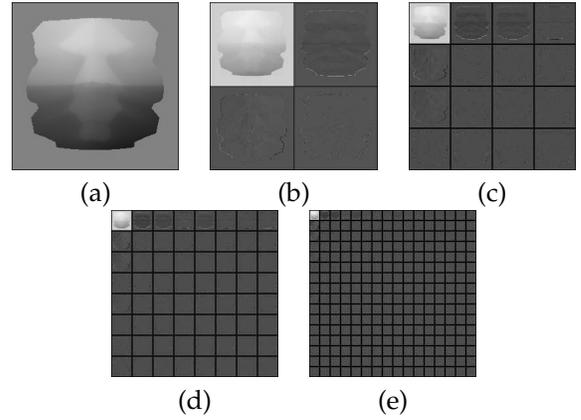


Fig. 8. Wavelet analysis of a frontal facial geometry image (the intensity of the coefficients was adjusted for visualization purposes): (a) original image, (b-e) depiction of the coefficients from $1^{st}$, $2^{nd}$, $3^{rd}$ and $4^{th}$ level Walsh transform.

packet system is constructed by repeated application of the Haar filterbank, a two-channel multirate filterbank based on the Haar conjugate mirror filter. The choice of Haar filters was based on their excellent localization properties. The application of the Haar filterbank is conceptually simple and computationally efficient. It is performed by applying a low-pass filter and a high-pass filter on a one-dimensional input, and repeating the same process on the two resulting outputs. The low–pass and high–pass Haar filters are $g$ and $h$, respectively: $g = \frac{1}{\sqrt{2}}[1 \quad 1]$ and $h = \frac{1}{\sqrt{2}}[1 \quad -1]$. For images, there will be four outputs for each level of the Haar wavelet: $g^t * g, g^t * h, h^t * g$ and $h^t * h$ (corresponding to Low-Low, Low-High, High-High, High-Low respectively). We compute a level 4 decomposition (i.e., we apply the filters four times), which yields $16 \times 16$ wavelet packets (Fig. 8). Since the geometry and normal images are of resolution $256 \times 256$, each wavelet packet has a resolution of $16 \times 16$.

The level 4 decomposition produces 256 ($16 \times 16$) wavelet packets. Since not all packets encode the same amount of information, most of the packets can be ignored without losing significant information from the original image. In our approach, 40 wavelet packets (roughly 15 %) are retained to create an efficient and compact biometric signature. We favor the wavelet packets with the minimum variation among the scans of the same subject as well as the maximum variation among the scans of different subjects. The selection of the 40 wavelet packets was optimized using a validation database with frontal facial scans.

The coefficients contained within the signature can be directly compared, without having to reconstruct the original image, using a weighted $L_1$ metric. The difference between a probe and a gallery geometry image is measured as:
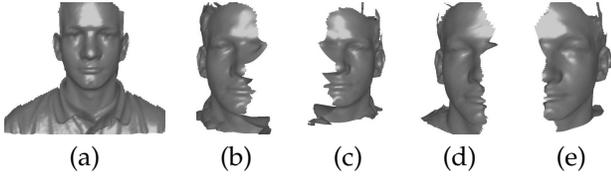
$$\text{DG} = \sum_{i=1}^{n} \mathbf{w}_i |G_P[i] - G_G[i]| \ ,$$

Fig. 9. Depiction of scans from the combined UND database from a single subject: (a) frontal, (b) $45^o$ right, (c) $45^o$ left, (d) $60^o$ right, and (e) $60^o$ left. Note the extensive missing data in (b-e).
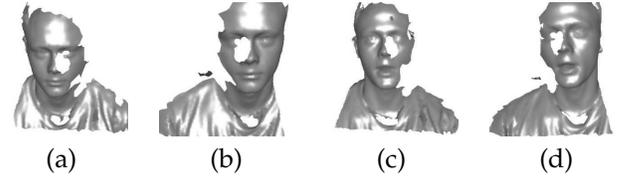


Fig. 10. Depiction of scans from the UH database from a single subject. (a,b) Right and left scans with neutral expression were acquired simultaneously. (c,d) Right and left scans with open mouth were acquired simultaneously.

where $G_P$ and $G_G$ are the wavelet coefficients derived from the geometry images of the probe and gallery scans respectively, $n$ is the number of coefficients, and $\mathbf{w}_i$ is a weight mapping function. The localization properties of the Walsh transform allow per area mapping, therefore for $k$ annotated facial areas, the $\mathbf{w}_i$ function will have $k$ distinct values. For the experiments, these values were selected empirically based on the biometric importance of each area.

The difference between a probe and a gallery normal image, DN, is computed in a similar manner. The final difference between a probe and a gallery scan is given by

$$D = DG + \mathbf{w}_n DN \ ,$$

where $w_n$ is a normalization weight. Since the ratio of the average $L_1$ difference between two geometry images over the average $L_1$ difference between two normal images is approximately $1 : 8$, we set $\mathbf{w}_n$ equal to 8. Note that the normal images, being the first spatial derivative of the geometry images, are less sensitive to positional (but not rotational) errors introduced during registration. As a result, the interpose matching favors the normal images over the geometry images.

# 4 EXPERIMENTAL RESULTS

## 4.1 Databases

### 4.1.1 Combined UND Databases

To evaluate the performance of UR3D-S a combination of the largest publicly available 3D face and ear databases was used. For frontal facial scans, we use the FRGC v2 database [48], [49]. It contains a total of 4007 range images, 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 $640 \times 480$. These data were obtained from 466 subjects and contain various facial expressions (e.g., happiness, surprise). For side facial scans, the Ear Database from the University of Notre Dame (UND) [50], collections F and G were used. This database (which was created for ear recognition purposes) contains side scans with yaw rotations of $45^o$, $60^o$ and $90^o$. In the $90^o$ side scans, both sides of the face are occluded from the sensor, and do not contain any useful information for face recognition purposes. Hence, only the $45^o$ side scans (118 subjects, 118 left and 118 right) and the $60^o$ side scans (87 subjects, 87 left and 87 right) were used. Even though these side scans are

denoted as $45^o$ and $60^o$, the measured average angle of rotation is $65^o$ and $80^o$ respectively (Fig. 9). Nevertheless, the database notation ($45^o$ and $60^o$) will be used. Since not all subjects exist in both databases, the number of common subjects between the frontal scans and the $45^o$ side scans is 39 and between the frontal scans and the $60^o$ side scans is 32.

For our experiments, we define the following collections:

- UND45LR: Contains $45^o$ side scans from 118 subjects. For each subject, the left scan is considered gallery and the right is considered probe. *Total: 236 scans.*
- UND60LR: Contains $60^o$ side scans from 87 subjects. For each subject, the left scan is considered gallery and the right is considered probe. *Total: 174 scans.*
- UND00LR: Gallery set has one frontal scan for each of the 466 subjects. Probe set has two $45^o$ side scans (left and right) for each of the 39 subjects and two $60^o$ side scans (left and right) for each of the 32 subjects. *Total: 608 scans.*

In all cases there is only one gallery scan per subject. Also, all subjects present in a probe set are also present in the gallery set (the opposite is not always true).

### 4.1.2 UH Databases

In addition to the UND databases we used a database with data collected at the University of Houston [51]. The database contains 1075 left and 1075 right scans of 281 subjects. A unique feature of this database is that each pair of left and right side scans was acquired simultaneously (Fig. 10). The scans were acquired using a 3dMD$^{\text{TM}}$system. This system consists of one left and one right optical scanner that acquire data simultaneously but are independent of each other. We consider these side scans comparable to the UND $45^o$ scans. During the acquisition of the data, each subject was asked to remove accessories on the face (e.g., glasses). An initial scan was acquired while the subject assumed a neutral facial expression. Subsequently, several scans were acquired while the subject was reading loudly a predefined text (thus assuming arbitrary facial expressions). All scans of a particular subject were acquired on the same day.

For our experiments, we define the following collections:

- UHDB7L: Contains left side scans from 281 subjects. For each subject, one scan is considered gallery and the rest are considered probe. The minimum and maximum left scans per subject are 1 and 6, respectively. *Total: 1075 scans.*
- UHDB7R: Contains right side scans from 281 subjects. For each subject, one scan is considered gallery and the rest are considered probe. The minimum and maximum right scans per subject are 1 and 6, respectively. *Total: 1075 scans.*
- UHDB7LR-M: Contains multiple left and right side scan pairs from 281 subjects. For each subject, one left and one right scan (acquired simultaneously) are considered gallery and the rest are considered probes. The minimum and maximum pairs of scans (left and right) per subject are 1 and 6, respectively. *Total: 2150 scans.*
- UHDB7LR-S: Contains a single left and right side scan pair from 281 subjects. For each subject, the left scan is considered gallery and the right scan is considered probe. The left and right scans were acquired simultaneously. *Total: 562 scans.*

In all cases there is only one gallery scan per subject with the exception of UHDB7LR-M where there are two. Also, all subjects present in a probe set are also present in the gallery set (but not vice versa).

## 4.2 Landmark Detection Evaluation

In order to evaluate the performance of the landmark detection algorithm we manually annotated landmarks on several facial scans, which can be considered as ground truth. We manually annotated landmarks on the 466 frontal scans of UND00LR (Fig. 9 (a)), on the 118 left and 118 right side scans of UND45LR (Fig. 9 (b-c)) and on the 87 left and 87 right side scans of UND60LR (Fig. 9 (d-e)).

In all cases, the overall mean distance error and standard deviation between the manually annotated landmarks and the automatically detected landmarks were computed to represent the landmarks' *localization error*. This error is expressed with the *mean Euclidian* distance, $D_E$, between the detected landmark points and the annotated landmark points.

Another metric which reflects the quality of the landmark detection algorithm for registration purposes is the *modified directed Hausdorff* distance, $D_H$, of the face model $M$ to the probe face $T$, as defined in [52]:

$$D_H(M,T) = \frac{1}{p} \sum_{i=1}^{p} \min_{t_j} |\mathbf{m}_i - \mathbf{t}_j| \; ,$$

where $|\mathbf{m}_i - \mathbf{t}_j|$ is the Euclidian distance between the face model vertices $\mathbf{m}_i$ and the probe face vertices $\mathbf{t}_j$, and $p$ is the number of face model vertices. The $D_H(M,T)$ represents the mean value of the minimum Euclidian distances $|\mathbf{m}_i - \mathbf{t}_j|$ of the vertices of the face model $M$, to which a probe facial scan $T$ is registered. In order to compute this metric, only the automatically detected

### TABLE 1
### 3D Landmark detection errors.

| Database | $D_H$ | | $D_E$ | | |
|---|---|---|---|---|---|
| | mean (mm) | std.dev (mm) | mean (mm) | std.dev (mm) | > 10 (mm) |
| UND00LR - front | 4.61 | 1.04 | 5.77 | 1.81 | 3.4 % |
| UND45LR - right | 3.90 | 0.95 | 5.83 | 2.49 | 4.2 % |
| UND45LR - left | 4.03 | 1.22 | 6.02 | 2.45 | 6.8 % |
| UND60LR - right | 4.37 | 3.11 | 5.87 | 2.47 | 6.9 % |
| UND60LR - left | 4.32 | 2.41 | 6.08 | 2.53 | 11.5 % |
| UHDB7R | 4.72 | 2.46 | - | - | - |
| UHDB7L | 4.76 | 2.86 | - | - | - |

landmarks were used for registration (without the Simulated Annealing step). To get comparative results for $D_H$ we used as a model for frontal databases all the vertices of the complete AFM, for left-side databases the left side vertices of the AFM, and for right-side databases the right side vertices of the AFM.

The results are summarized in Table 1. The last column reports the percentage of landmarks with $D_E$ more than $10 \; mm$. Note that there were no manually annotated landmarks for the UHDB7R and UHDB7L databases, threfore only the $D_H$ metric is reported. A more detailed performance evaluation of the landmark detection algorithm can be found in [5].

## 4.3 Face Recognition Performance Evaluation

Using the databases described in Section 4.1 we performed several identification experiments. UR3D-S tackles the problem of matching arbitrary facial scans (left, right or frontal). This is considerably harder than matching only frontal scans, since side scans have extensive missing data and it is not known a priori whether each scan is left, right or frontal. In all experiments the Cumulative Match Characteristic (CMC) graphs and the rank-one recognition rates are reported. The automatic landmark detector was used in all cases unless stated otherwise.

### 4.3.1 Matching facial scans of the same side

In this experiment the performance of UR3D-S using scans of the same side for both gallery and probe sets was evaluated. This is not a realistic scenario, but allows for the evaluation of our method without the need to use facial symmetry. The only database suitable for this purpose is the UH Database as it has multiple left and multiple right side scans of each subject. We measured the performance for matching left (gallery) with left (probe) side scans (for UHDB7L) and right (gallery) with right (probe) side scans (for UHDB7R). The rank-one rate is 85.8 % and 86.8 % for UHDB7L and UHDB7R respectively. The CMC graphs are presented in Fig. 11.

### 4.3.2 Matching facial scans of arbitrary side

In this experiment the performance of UR3D-S using scans of arbitrary sides for gallery and probe sets is
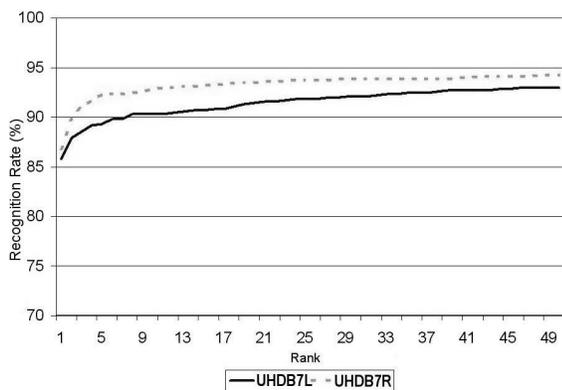
Fig. 11. CMC graphs for matching left (gallery) with left (probe) side scans (for UHDB7L) and right (gallery) with right (probe) side scans (for UHDB7R).
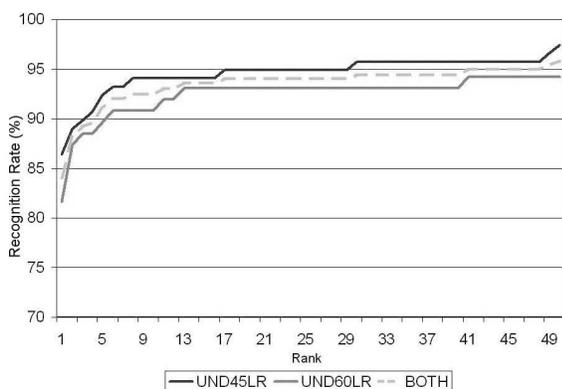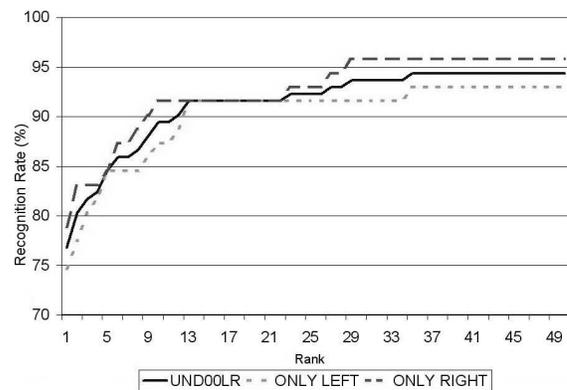


Fig. 13. CMC graphs for matching frontal (gallery) with left, right and both (probe) side scans using UND00LR.



Fig. 12. CMC graphs for matching left (gallery) with right (probe) side scans using UND45LR, UND60LR and the combination of the two.
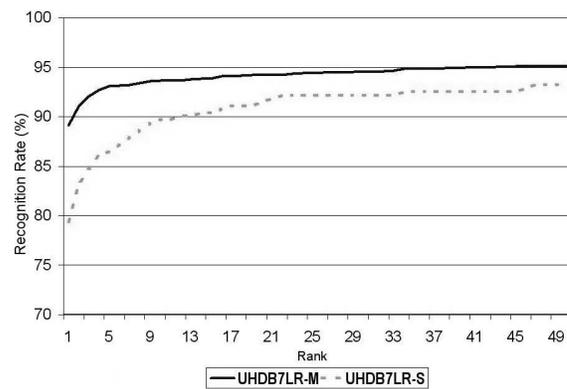


Fig. 14. UHDB7LR-M: matching left and right (gallery) with left and right (probe) side scans. UHDB7LR-S: matching left (gallery) with right (probe) side scans.

evaluated. This is a realistic scenario, as side scans (with extensive occlusions that lead to missing data) are very common in real world applications with unconstrained acquisition. Our method can match any combination of left, right or frontal facial scans with the use of facial symmetry. Moreover, our method automatically detects the side of the scan. For this experiment we used the UND45LR, UND60LR, UND00LR, UHDB7LR-M and UHDB7LR-S databases and the rank-one rates are reported in Table 2.

In UND45LR and UND60LR, for each subject, the gallery set contains a single left side scan while the probe set contains a single right side scan. Therefore, facial

symmetry is always used for identification. As expected, the $60^o$ side scans yielded lower recognition rates as they are considered more challenging compared to the $45^o$ side scans (Fig. 12).

In UND00LR, the gallery set contains a frontal scan for each subject, while the probe set contains left and right side scans. This scenario is very common when the enrollment of subjects is controlled but the identification is uncontrolled. Figure 13 depicts the CMC graph (UND00LR's probe set is also split in left-only and right-only subsets). Compared to UND45LR and UND60LR, there is a decrease in the performance of our method in UND00LR. One could argue that since the gallery set consists of frontal scans (without missing data), there should be an increase in the performance. However, UND00LR has the largest gallery set (it includes all of the 466 subjects found in the FRGC v2 database) making it the most challenging database in our experiments.

In UHDB7LR-M, for each subject the gallery set contains a left and right side scan pair, while the probe set contains multiple left and right side scan pairs. Since the gallery set has two scans per subject, the performance on this database is the highest among all databases, as expected. The performance difference is substantial compared to UND00LR (89.1 % versus 76.8 % rank-one).

TABLE 2
Rank-one recognition rates per database.

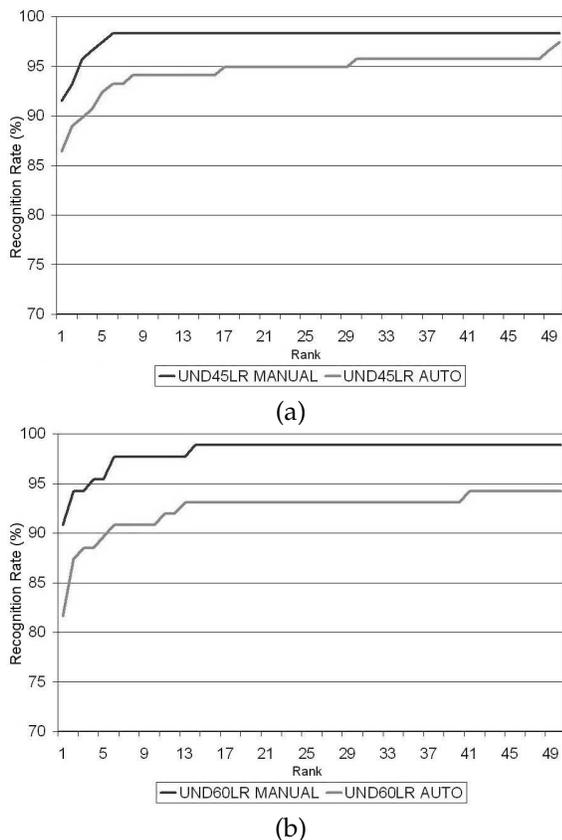|  | Rank-one Rate |
|---|---|
| UND45LR | 86.4 % |
| UND60LR | 81.6 % |
| UND00LR | 76.8 % |
| UHDB7LR-M | 89.1 % |
| UHDB7LR-S | 79.4 % |

(a)



(b)

Fig. 15. CMC graphs for matching left (gallery) and right (probe) side scans using automatic and manual landmarks: (a) UND45LR and (b) UND60LR

TABLE 3
Rank-one recognition rates per database using automatic and manual landmarks.

| | Rank-one Rate | |
|---|---|---|
| | *Automatic Landmarks* | *Manual Landmarks* |
| UND45LR | 86.4 % | 91.5 % |
| UND60LR | 81.6 % | 90.8 % |

This indicates that a pair of left and right side scans is more descriptive than one frontal scan.

UHDB7LR-S, even though it is a subset of UHDB7LR-M, is considered more challenging (Fig. 14). This is because the gallery set contains a single left side scan while the probe set contains a single right side scan. Compared to UND45LR and UND60LR (which have a similar probe/gallery setup), the performance on UHDB7LR-S is lower (79.4 % versus 86.4 % and 81.6 % rank-one). However, UHDB7LR-S is considerably larger, consisting of 281 subjects versus 118 and 87 subjects for UND45LR and UND60LR, respectively.

### 4.3.3 Automatic versus manual landmarks

In the final experiment, we evaluated the performance of our method with an ideal landmark detector. To this end, we used the manually annotated landmarks for the UND45LR and UND60LR databases described in Section 4.2. The resulting CMC graphs are depicted in Figure 15. As expected, there is an increase in the rank-one recognition rate (given in Table 3) for both databases (5.1 % and 9.2 % increase for UND45LR and UND60LR, respectively). Interestingly, the percentage of automatic landmarks with mean localization error more than $10$ $mm$ (compared to manually annotated landmarks) is on average 5.5 % and 9.2 % on UND45LR and UND60LR, respectively (see last column of Table 1). This indicates that the proposed method can tolerate landmark mean localization errors up to $10$ $mm$. If this error is significantly above this threshold then the pose estimation is probably invalid and the subsequent fitting fails to extract meaningful geometrical information.

### 4.4 Computational Efficiency

For the evaluation of the proposed method's computational efficiency an Intel Core 2 Duo 2.2 GHz with 2 GB RAM and a NVIDIA 8600 GTS was used. Using this PC, an average time of 18 *secs* is required to process a facial scan: 9 *secs* to localize the facial landmarks plus 9 *secs* to extract the biometric signature (geometry and normal images). The procedures of determining the optimal rotation for the alignment of the landmark shapes to the FLMs require up to 8 iterations to converge. The Simulated Annealing step may take up to 2,000 iterations to converge, but the computation is very efficient (requires 2 *secs*) as the z-buffers are created using the GPU. The fitting step takes 64 iterations to converge (and requires 7 *secs*). The creation of the signature from the deformed AFM requires just a few milliseconds and can be matched at a rate of 15,000 matches per second.

### 4.5 Discussion

Most of the face recognition methods that have been previously proposed do not handle data with significant pose variations along the yaw axis. On the contrary, UR3D-S can handle extensive occlusions (that result in missing data) caused by pose variations (such as $60^o$ side scans). The only limitation of the proposed method is that at least half of the face is visible (so that facial symmetry can be used). The flexibility to handle (seamlessly) left, right and frontal scans is important in an unconstrained acquisition scenario. Therefore, we consider that our focus on data with arbitrary pose variations is a necessity for real-world applications.

For evaluation purposes we used the most challenging databases in terms of pose variations and missing data. We demonstrated that UR3D-S can match frontal and left or right side scans by using facial symmetry. Unavoidably, the use of facial symmetry has an impact on recognition rates, as human faces are not completely symmetric. In the above experiments, the average rank-one recognition rate was 83.7 %. We anticipate that the average recognition rate can be increased by improving certain steps of the proposed method (e.g., landmark

detection) before the limit imposed by facial asymmetry is reached.

Compared to our previous work with the combined UND databases [3], UR3D-S offers significantly better results. When using automatically detected landmarks, the rank-one recognition rates increased by 19 % and 17 %, for UND45LR and UND60LR, respectively. This indicates that the landmark detector used in this work (see [5]) is far more robust and accurate compared to the detector proposed in [3]. When using manually selected landmarks, the rank-one recognition rates increased by 9 % and 22 %, for UND45LR and UND60LR, respectively (with respect to [3]). Thus, the improvements on the other steps of the method also offer increased accuracy. Another important difference compared to previous work [3] is that each frontal scan is now handled as a pair of left and right side scans (it produces two independent signatures). This is why the largest performance increase was on a database with frontal scans in the gallery set and side scans in the probe set (20 % increased recognition rate on UND60LR compared to the average recognition rate of DB60F and DB45F in [3]).

Regarding the computational efficiency of UR3D-S, there are a number of independent tasks that can be exploited to increase the computational efficiency of the method (parallel processing techniques), but we have not specifically addressed this issue. Nevertheless, the combination of a signature creation step with a reasonable computational cost (18 *secs*) and a signature matching step with an extremely low computational cost (15,000 matches per second) makes the proposed method applicable to real-world scenarios.

## 5 CONCLUSION

A novel 3D face recognition method suitable for real-world biometric applications was proposed. Unlike most previously proposed methods that require frontal scans, the proposed method can perform partial matching among interpose facial scans, even when extensive data are missing. By using facial symmetry it can handle seamlessly frontal and side facial scans. Competitive results were presented on databases with challenging pose variations. Compared to our previous work, we introduced a series of novelties that resulted approximately in a 20 % increased recognition rate. Future work will be directed towards improving the robustness of the landmark detector and evaluating the performance of the proposed method on additional databases.

## 6 ACKNOWLEDGMENT

## REFERENCES

[1] G. Passalis, I. Kakadiaris, and T. Theoharis, "Intraclass retrieval of non-rigid 3D objects: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 218–229, 2007.

[2] I. Kakadiaris, G. Passalis, G. Toderici, M. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis, "Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 640–649, Apr. 2007.

[3] P. Perakis, G. Passalis, T. Theoharis, G. Toderici, and I. Kakadiaris, "Partial matching of interpose 3D facial data for face recognition," in *Proc. 3rd IEEE International Conference on Biometrics: Theory, Applications and Systems*, Arlington, VA, Sep. 28-30 2009, pp. 439–446, Best Reviewed Paper Award.

[4] P. Perakis, T. Theoharis, G. Passalis, and I. Kakadiaris, "Automatic 3D facial region retrieval from multi-pose facial datasets," in *Proc. Eurographics Workshop on 3D Object Retrieval*, Munich, Germany, Mar. 30 - Apr. 3 2009, pp. 37–44.

[5] P. Perakis, T. Theoharis, G. Passalis, and I. Kakadiaris, "3D facial landmark detection & face registration: A 3D facial landmark model & 3D local shape descriptors approach," Computer Graphics Laboratory, University of Athens, Tech. Rep., January 2011. [Online]. Available: http://graphics.di.uoa.gr/Research/TechReports/3D_Lmk_v1.pdf

[6] K. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *Computer Vision and Image Understanding*, vol. 101, no. 1, pp. 1–15, Jan. 2006.

[7] K. Chang, K. Bowyer, and P. J. Flynn, "An evaluation of multi-modal 2D+3D face biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 619–624, April 2005.

[8] X. Lu and A. Jain, "Multimodal facial feature extraction for automatic 3D face recognition," Michigan State University, Tech. Rep. MSU-CSE-05-22, Oct. 2005.

[9] ——, "Automatic feature extraction for multiview 3D face recognition," in *Proc. 7th International Conference on Automatic Face and Gesture Recognition*, Southampton, UK, Apr. 2-6 2006, pp. 585–590.

[10] X. Lu, A. Jain, and D. Colbry, "Matching 2.5D face scans to 3D models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 31–43, 2006.

[11] H. Dibeklioglu, "Part-based 3D face recognition under pose and expression variations," Master's thesis, Bogazici University, 2008.

[12] H. Dibeklioglu, A. Salah, and L. Akarun, "3D facial landmarking under expression, pose and occlusion variations," in *Proc. 2nd International Conference on Biometrics Theory, Applications and Systems*, Arlington, VA, Sep. 29 - Oct. 1 2008, pp. 1–6.

[13] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063–1074, 2003.

[14] V. Blanz, K. Scherbaum, and H.-P. Seidel, "Fitting a morphable model to 3D scans of faces," in *Proc. 11th IEEE International Conference on Computer Vision*, Rio de Janeiro, Brazil, Oct. 14-20 2007, pp. 1–8.

[15] A. Bronstein, M. Bronstein, and R. Kimmel, "Robust expression-invariant face recognition from partially missing data," in *Proc. European Conference on Computer Vision*, Graz, Austria, 2006, pp. 396–408.

[16] P. Nair and A. Cavallaro, "Matching 3D faces with partial data," in *Proc. British Machine Vision Conference*, Leeds, UK, Sep. 1-4 2008.

[17] ——, "3-D face detection, landmark localization, and registration using a point distribution model," *IEEE Transactions on Multimedia*, vol. 11, no. 4, pp. 611–623, Jun. 2009.

[18] T. Lin, W. Shih, W. Chen, and W. Ho, "3D face authentication by mutual coupled 3D and 2D feature extraction," in *Proc. 44th ACM Southeast Regional Conference*, Melbourne, FL, Mar. 10-12 2006, pp. 423–427.

[19] A. Mian, M. Bennamoun, and R. Owen, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1927 – 1943, Nov. 2007.

[20] M. Segundo, C. Queirolo, O. Bellon, and L. Silva, "Automatic 3D facial segmentation and landmark detection," in *Proc. 14th International Conference on Image Analysis and Processing*, Modena, Italy, Sep. 10-14 2007, pp. 431 – 436.

[21] X. Wei, P. Longo, and L. Yin, *LNCS, Advances in Biometrics*. Springer, 2007, ch. Automatic Facial Pose Determination of 3D Range Data for Face Model and Expression Identification, pp. 144–153.

[22] T. Faltemier, K. Bowyer, and P. Flynn, "Rotated profile signatures for robust 3D feature detection," in *Proc. $8^{th}$ IEEE International Conference on Automatic Face and Gesture Recognition*, Amsterdam, The Netherlands, Sep. 17-19 2008, pp. 1–7.

[23] C. Xu, T. Tan, Y. Wang, and L. Quan, "Combining local features for robust nose location in 3D facial data," *Pattern Recognition Letters*, vol. 27, no. 13, pp. 62–73, 2006.

[24] Y. Liu and J. Palmer, "A quantified study of facial asymmetry in 3D faces," in *Proc. IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, Oct. 17 2003, pp. 222–229.

[25] L. Kompanets, "Biometrics of asymmetrical face," in *Proc. $1^{st}$ International Conference on Biometric Authentication*, Hong Kong, July 15-17 2004, pp. 67–73.

[26] Y. Liu, R. L. Weaver, K. Schmidt, N. Serban, and J. Cohn, "Facial asymmetry: A new biometric," Carnegie Mellon University - Robotics Institute, Tech. Rep. CMU-RI-TR-01-23, August 2001.

[27] S. Mitra, M. Savvides, and B. V. Kumar, "Face identification using novel frequency-domain representation of facial asymmetry," in *Proc. IEEE Transactions on Information Forensics and Security*, Sep. 2006, pp. 350–359.

[28] S. Gutta, V. Philomin, and M. Trajkovic, "An investigation into the use of partial-faces for face recognition," in *Proc. $5^{th}$ IEEE international Conference on Automatic Face and Gesture Recognition*, Washington DC, May 20-21 2002, pp. 28–33.

[29] I. Dryden and K. Mardia, *Statistical shape analysis*. Wiley, 1998.

[30] M. B. Stegmann and D. D. Gomez, "A brief introduction to statistical shape analysis," Technical University of Denmark, Tech. Rep., Mar. 2002.

[31] T. Cootes and C. Taylor, "Statistical models of appearance for computer vision," University of Manchester, Tech. Rep., Oct 2001.

[32] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active shape models - their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, Jan. 1995.

[33] T. Cootes, C. Taylor, H. Kang, and V. Petrovic, "Modeling facial shape and appearance," in *Handbook of Face Recognition*. Springer, 2005, pp. 39 – 63.

[34] D. Colbry, "Human face verification by robust 3D surface alignment," Ph.D. dissertation, Michigan State University, 2006.

[35] D. Colbry, G. Stockman, and A. Jain, "Detection of anchor points for 3D face verification," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, Jun. 20-25 2005, p. 118.

[36] A. Johnson, "Spin-Images: A Representation for 3-D Surface Matching," Ph.D. dissertation, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, Aug. 1997.

[37] L. Farkas, *Anthropometry of the head and face*, $2^{nd}$ ed., L. G. Farkas, Ed. Raven Press, 1994.

[38] X. Gu, S. Gortler, and H. Hoppe, "Geometry images," in *Proc. SIGGRAPH*, San Antonio, TX, Jul. 2002, pp. 355–361.

[39] E. Praun and H. Hoppe, "Spherical parametrization and remeshing," in *Proc. SIGGRAPH*, San Diego, CA, July 2003, pp. 340–349.

[40] G. Papaioannou, E. Karabassi, and T. Theoharis, "Reconstruction of three-dimensional objects through matching of their parts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 114–124, Jan. 2002.

[41] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by simulated annealing," *Science*, vol. 22, no. 4598, pp. 671–680, 1983.

[42] P. Siarry, G. Berthiau, F. Durbin, and J. Haussy, "Enhanced simulated annealing for globally minimizing functions of many-continuous variables," *ACM Transactions on Mathematical Software*, vol. 23, no. 2, pp. 209–228, 1997.

[43] C. Mandal, "A dynamic framework for subdivision surfaces," Ph.D. dissertation, University of Florida, 1998.

[44] C. Loop, "Smooth subdivision surfaces based on triangles," Master's thesis, Department of Mathematics, University of Utah, 1987.

[45] A. Klinger, "Pattern and search statistics," *Optimizing Methods in Statistics*, pp. 303–337, 1971.

[46] A. Finkel and J. Bentley, "Quad trees: A data structure for retrieval of composite keys," *Acta Informatica*, vol. 4, no. 1, pp. 1–9, 1974.

[47] E. Stollnitz, T. DeRose, and D. Salesin, *Wavelets for computer graphics: Theory and applications*, I. Morgan Kaufmann Publishers, Ed. Morgan Kaufmann Publishers, Inc, 1996.

[48] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, 2005, pp. 947–954.

[49] P. Phillips, W. Scruggs, A. OToole, P. Flynn, K. Bowyer, C. Schott, and M. Sharpe, "FRVT 2006 and ICE 2006 Large-Scale Experimental Results," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 2, pp. 831–846, 2010.

[50] University of Notre Dame, "University of Notre Dame Biometrics Database," http://www.nd.edu/~cvrl/UNDBiometricsDatabase.html, 2008.

[51] "UH Biometrics Databases" [Online]. Available: http://cbl.uh.edu/URxD/datasets/

[52] Y. Gao, "Efficiently comparing face images using a modified Hausdorff distance," in *Proc. IEEE Conference on Vision, Image and Signal Processing*, Madison, WI, USA, Dec. 2003, pp. 346–350.

**Georgios Passalis** received his B.Sc. from the Department of Informatics and Telecommunications of University of Athens in 2003. Subsequently, he received his M.Sc. from the Department of Computer Science at University of Houston in 2004. He received his Ph.D. in 2008 from the University of Athens with focus on the domains of computer graphics and computer vision. His is currently working in the video game industry.

**Panagiotis Perakis** received the B.Sc. degree in physics, in 1986, and the M.Sc. degree in ICT, in 2008, from the University of Athens, Greece. Currently, he is a Ph.D. candidate at the Department of Informatics and Telecommunications, University of Athens. His thesis is focused on the domains of computer graphics and computer vision. His research interests include computer graphics, computer vision and physics-based modeling. He is also a co-owner of a Greek software development company, since 1993.

**Theoharis Theoharis** received his D.Phil. in computer graphics and parallel processing from the University of Oxford, U.K. in 1988. He subsequently served as a research fellow (postdoc) at the University of Cambridge, U.K. and as a consultant with Andersen Consulting. He is with the University of Athens, Greece (since 1993) and is also an Adjunct Professor at the Department of Computer Science of the University of Houston. During the 2002/3 academic year he was on Sabbatical leave at the Computational Biomedicine Lab, University of Houston, Texas. His main research interests lie in the fields of Computer Graphics, Visualization, Biometrics and Archaeological Reconstruction.

**Ioannis A. Kakadiaris** is an Eckhard Pfeiffer Professor of Computer Science, Electrical & Computer Engineering, and Biomedical Engineering at the University of Houston. He joined UH in August 1997 after a postdoctoral fellowship at the University of Pennsylvania. Ioannis earned his B.Sc. in physics at the University of Athens in Greece, his M.Sc. in computer science from Northeastern University and his Ph.D. at the University of Pennsylvania. He is the founder of the Computational Biomedicine Lab (www.cbl.uh.edu) and in 2008 he directed the Methodist-University of Houston-Weill Cornell Medical College Institute for Biomedical Imaging Sciences (IBIS, ibis.uh.edu) (position rotates annually among the institutions). His research interests include biomedical image analysis, biometrics, computer vision, and pattern recognition. Dr. Kakadiaris is the recipient of a number of awards, including the NSF Early Career Development Award, Schlumberger Technical Foundation Award, UH Computer Science Research Excellence Award, UH Enron Teaching Excellence Award, and the James Muller Vulnerable Plaque Young Investigator Prize. His research has been featured on The Discovery Channel, National Public Radio, KPRC NBC News, KTRH ABC News, and KHOU CBS News.