Partial Matching of Interpose 3D Facial Data for Face Recognition

P. Perakis†‡ G. Passalis†‡ T. Theoharis†‡ G. Toderici‡ I.A. Kakadiaris ‡

Abstract— **Three–dimensional face recognition has lately received much attention due to its robustness in the presence of lighting and pose variations. However, certain pose variations often result in missing facial data. This is common in realistic scenarios, such as uncontrolled environments and uncooperative subjects. Most previous 3D face recognition methods do not handle extensive missing data as they rely on frontal scans. Currently, there is no method to perform recognition across scans of different poses. A unified method that addresses the partial matching problem is proposed. Both frontal and side (left or right) facial scans are handled in a way that allows interpose retrieval operations. The main contributions of this paper include a novel 3D landmark detector and a deformable model framework that supports symmetric fitting. The landmark detector is utilized to detect the pose of the facial scan. This information is used to mark areas of missing data and to roughly register the facial scan with an Annotated Face Model (AFM). The AFM is fitted using a deformable model framework that introduces the method of exploiting facial symmetry where data are missing. Subsequently, a geometry image is extracted from the fitted AFM that is independent of the original pose of the facial scan. Retrieval operations, such as face identification, are then performed on a wavelet domain representation of the geometry image. Thorough testing was performed by combining the largest publicly available databases. To the best of our knowledge, this is the first method that handles side scans with extensive missing data (e.g., up to half of the face missing).**

I. INTRODUCTION

In recent years, 3D face recognition has received much attention because of its potential to overcome lighting and pose variations which hamper 2D face recognition methods. Most previously proposed 3D face recognition methods claim pose invariance. The main assumption of these methods is that even though the head can be rotated with respect to the sensor, the *entire* face is always visible. However, this is true only for frontal scans; side scans usually have large missing areas. These scans are very common in realistic scenarios such as uncooperative subjects or uncontrolled environments. Therefore, to take advantage of the full pose invariance potential of 3D face recognition, the partial object matching problem must be addressed.

In our previous work, we have proposed a general intraclass 3D object retrieval method [26]. We employed this method to face and ear recognition [15], [27], [34] with success (scored the top accuracy in the shape-only section of NIST's Face Recognition Vendor Test 2006). However, no interpose retrieval was performed; frontal scans were used strictly for face recognition and side scans were used strictly for ear recognition.

In this paper, we extend our previous work to handle partial matching in order to allow interpose identification. A novel 3D landmark detector based on curvature and radial distances is introduced. It is robust to missing data and works both for frontal and side scans, allowing an initial registration with an Annotated Face Model (AFM). The deformable model framework is extended to take advantage of the symmetry of the human face during fitting (Fig. 1). This allows the creation of geometry images that are independent of the initial pose, even in cases of severe missing data. Therefore, by *completing* the missing data we handle the partial matching problem, thus achieving true pose invariance. Experiments are performed on a combination of the largest publicly available databases. Since the available 3D databases for face recognition contain frontal scans, we combined them with 3D databases for ear recognition. These contain up to 80^o side scans (both left and right). Note that we only use the subset of subjects that have at least one frontal and one side scan.

The rest of this paper is organized as follows: Section II describes related work in the field, Section III presents the proposed method in detail, Section IV evaluates the method's performance, while Section V summarizes the method and proposes future directions.

II. RELATED WORK

There exists extensive work in the face recognition field, excellent surveys of which are given by Bowyer *et al.* [3] and Chang *et al.* [5]. The majority of these works use frontal scans only, and few handle the partial matching problem.

Lu *et al*. [19], [20], [21], in a series of works, have presented methods to locate the positions of eye and mouth corners, and nose and chin tips, based on a fusion scheme of 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 matching accuracy in an identification experiment, no claims where made with respect to the effects of pose variations in results.

Dibeklioglu [11] 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 proposed system is that it is

†Computer Graphics Laboratory, Department of Informatics and Telecommunications, University of Athens, Ilisia 15784, GREECE

‡Computational Biomedicine Lab, Department of Computer Science, University of Houston, Texas 77204, USA

not applicable to facial datasets with yaw rotations greater than 45°. Additionally, even though the Bosphorus database used consists of 3,396 facial scans, they are obtained from 81 subjects.

Blanz *et al*. [2], [1] presented works on 3D face reconstruction by fitting their 3D Morphable Model on 3D facial scans. Although their method is a well established approach for producing 3D synthetic faces from scanned data, it uses manually positioned landmarks for the fitting procedure. Also, 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°.

Bronstein *et al*. [4] 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. On a limited database of 30 subjects they report high recognition rates. However, the database they use has no side scans. The scans with missing data that they use are derived synthetically by randomly removing certain areas from frontal scans.

In Nair's [23] work on partial 3D face matching, the face is divided into areas and only certain areas are used for registration and matching. The assumption is 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 yields higher recognition rates. They mention, however, that their method has problems when holes exist around the nose region.

Lin *et al*. [17] introduced a coupled 2D and 3D feature extraction method to determine the positions of eye sockets by 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 tested on only 27 human faces with various poses and expressions.

Methods that are focused mainly on the detection of 3D facial landmarks include that of Segundo *et al*. [30]. They introduced 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% correct detections. However, nose and eye corner detection presented problems when the face had a significant pose variation ($> 15^\circ$ around the y and z-axes).

Finally, Wei *et al*. [35] introduced a nose tip localization method to determine the facial pose. The method was based on a Surface Normal Difference algorithm and Shape Index estimation, and used as a preprocessing step in pose-variant systems to determine the pose of the face. No claims where made with respect to invariance to pose.

III. Method

The proposed method processes each raw facial scan through a common pipeline of algorithms. The output of this pipeline is a wavelet domain representation of the face. The coefficients of this representation are directly comparable using a L^1 distance metric, thus allowing efficient matching and retrieval operations. Moreover, the representation is independent of the initial pose due to the symmetric fitting algorithm (Fig. 1). Therefore, by automatically completing missing data, the partial face matching problem is converted to a full face matching problem. The individual steps of our method are the following:

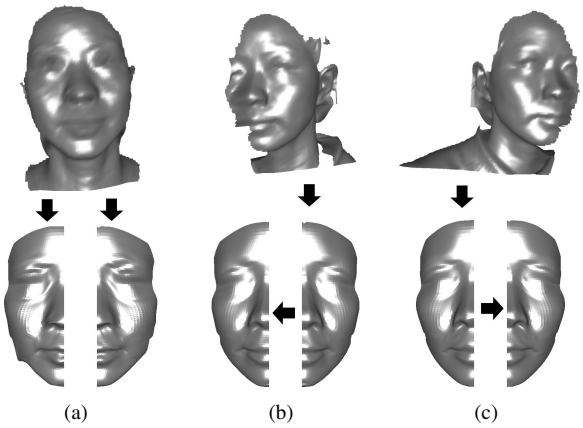


Fig. 1. Raw data (top row) and fitted AFM (bottom row) of a subject: (a) Frontal scan has both sides fitted; (b) left scan where left side is fitted (and the right is mirrored); and (c) right scan where the right side is fitted (and the left is mirrored).

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

Step 2. *3D Landmark Detection:* A novel landmark detector is used to estimate the rough pose (thus determining if it is a frontal, left or right scan).

Step 3. *Registration:* The raw data are registered to the AFM.

Step 4. *Deformable Model Fitting:* The Annotated Face Model is fitted to the data. The fitted model is then converted to a geometry image (a normal image is also computed).

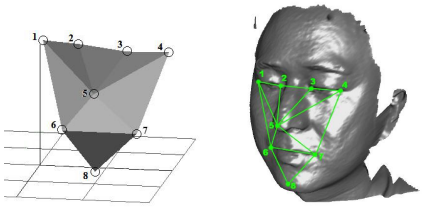
Step 5. *Wavelet Analysis:* A wavelet transform is applied on the geometry and normal image and the wavelet coefficients are exported and stored.

A. Preprocessing

The purpose of preprocessing is mainly to eliminate sensor-specific problems. 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, certain preprocessing algorithms (*Median Cut*, *Hole Filling*, *Smoothing* and *Subsampling*) operate directly on the range data before the conversion to polygonal data [15].

B. 3D Landmark Detection

Our method for 3D Landmark Detection and Pose Estimation utilizes 3D information to extract candidate interest points, which are identified and labeled as anatomical landmarks by matching them with a Facial Landmark Model



(a) (b)

Fig. 2. Landmark models: (a) landmark model as a 3D object; and (b) landmark model overlaid on a 3D facial dataset. [28]

(FLM) [28]. Once anatomical landmarks are localized, the corresponding rigid transformation is computed in order to register the facial datasets.

We used a set of eight anatomical landmarks (Fig. 2): (1) the right eye outer corner, (2) right eye inner corner, (3) left eye inner corner, (4) left eye outer corner, (5) nose tip, (6) mouth right corner, (7) mouth left corner, and (8) chin tip.

Note that five of these landmarks are visible in side scans (right side contains landmarks 1, 2, 5, 6, 8 and left side contains 3, 4, 5, 7, 8). These sets of landmarks constitute a Facial Landmark Model (FLM). In the following, 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 steps to create the FLMs are:

- A statistical Mean Shape for each landmark set (FLM8, FLM5L and FLM5R) is estimated from a manually annotated training set (150 frontal faces with neutral expressions) using Procrustes Analysis.
- Variations of each Facial Landmark Model are computed using Principal Component Analysis (PCA).

For each facial dataset the procedure for landmark detection has the following steps (Fig. 5):

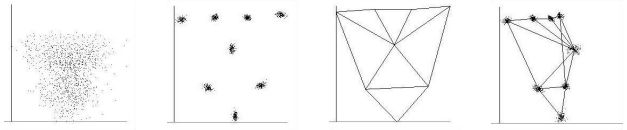
Step 2.1. Extract candidate landmarks from the Shape Index map and the Extrusion map.

Step 2.2. Compute the rigid transformation that best aligns combinations of eight or five candidate face landmarks with the corresponding FLMs.

Step 2.3. Discard combinations of candidate landmark sets that are not consistent with the FLMs.

Step 2.4. Select the best combination of candidate landmarks (based on the minimum Procrustes distance) and the corresponding rigid transformation for registration.

The Landmark Mean Shape: To obtain a true representation of landmark shapes, location and rotational effects need to be filtered out. Since, for our purposes, the size of the shape is of great importance, it is not filtered out by scaling shapes to unit size. This is carried out by establishing a common coordinate reference to which all shapes are aligned. Alignment is performed by minimizing the Procrustes distance $D^2 = |\mathbf{x}_i - \mathbf{x}_m|^2$ of each shape ($\mathbf{x}_i$) to the mean ($\mathbf{x}_m$). The alignment procedure is commonly known as Procrustes Analysis [12], [32], [8], and is used to compute the Mean Shape of landmark shapes (Fig. 3). The



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

Fig. 3. Landmark Mean Shape estimation: (a) unaligned landmarks; (b) aligned landmarks; (c) landmark mean shape; and (d) landmark cloud & mean shape rotated 60° around the y-axis. [28]

Mean Shape is the Procrustes mean: $\mathbf{x}_m = \frac{1}{N}\sum \mathbf{x}_i$ for all example shapes $\mathbf{x}_i$ after alignment.

Landmark Shape Variations: After bringing landmark shapes into a common frame of reference and estimating the landmarks' Mean Shape, further analysis can be carried out to describe the shape variations. This shape decomposition is performed by applying PCA to the aligned shapes.

Aligned shape vectors form a distribution in the nd-dimensional shape space, where n is the number of landmarks and d the dimension of each landmark. We can model this distribution by estimating a vector $\mathbf{b}$ of parameters that describes the shape's deformations [10], [8], [9]. We compute the eigenvectors $\mathbf{A_i}$, and corresponding eigenvalues λ_i of the covariance matrix of the shape vectors, sorted in descending order.

If $\mathbf{A}$ contains (in columns) the p eigenvectors $\mathbf{A}_i$ corresponding to the p largest eigenvalues, then we can approximate any example shape $\mathbf{x}$ using: $\mathbf{x}' \approx \mathbf{x}_m + \mathbf{A} \cdot \mathbf{b}$, where $\mathbf{b}$ is a p-dimensional vector given by: $\mathbf{b} = \mathbf{A}^T \cdot (\mathbf{x} - \mathbf{x}_m)$.

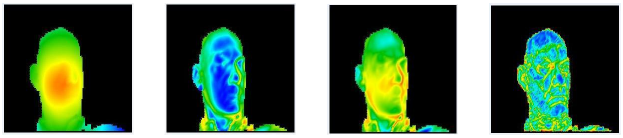
The vector $\mathbf{b}$ is the projection of $\mathbf{x}$ onto the subspace spanned by the p most significant eigenvectors of the eigenspace (*principal components*). By selecting the p largest eigenvalues, the mean square error between $\mathbf{x}$ and its approximation $\mathbf{x}'$ is minimized. By applying limits to each b_i (i.e., $|b_i| \leq 3\sqrt{\lambda_i}$) we can create marginal mean shape deformations. Thus, the *Facial Landmark Model* (FLM) is created [28].

The number p of eigenvectors and eigenvalues to retain (*modes of variations*) can be chosen so that the model represents a given proportion of the total variance of the data. For our purposes, a factor of 99% was chosen.

Fitting Landmarks to the Model: General-purpose feature detection methods are not able to identify and label the detected candidate landmarks. It is clear that some topological properties of faces need to be taken into consideration. To address this problem, we use the FLM. Candidate landmarks, irrespectively of the way they are produced, have to be consistent with the corresponding FLM. This is done by fitting a candidate landmark set to the FLM and checking the deformation parameters $\mathbf{b}$ to be within certain margins.

Fitting a set of points $\mathbf{y}$ to the FLM $\mathbf{x}$ is accomplished by minimizing the Procrustes distance in a simple iterative approach. We also consider a landmark shape as plausible if it is consistent with marginal shape deformations [8], [9].

Landmark Detection & Selection: Shape Index is extensively used for 3D landmark detection [6], [7], [19], [20], [21]. It is a continuous mapping of principal curvature values



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

Fig. 4. 2D image maps: (a) radial map; (b) tangent map; (c) extrusion map; and (d) Shape Index map.

(k_{max}, k_{min}) of a 3D object point $\mathbf{p}$ into the interval [0,1], according to the formula:

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

Its value represents the type of local curvature of shapes (Cup = 0.0, Rut = 0.25, Saddle = 0.5, Ridge = 0.75, Cap = 1.0).

After computing Shape Index values on a 3D facial dataset, a mapping to 2D space is performed (using the native UV parameterization of the facial scan) in order to create a *Shape Index map* (Fig. 4(d)). Local maxima and minima are identified on the Shape Index map. Local maxima (Cap = 1.0) are candidate landmarks for nose tips and chin tips and local minima (Cup = 0.0) for eye corners and mouth corners. The Shape Index's located maxima and minima are sorted in descending order of significance according to their corresponding Shape Index values. The most significant subset of points for each group (Caps and Cups) is retained. In Fig. 5(a), black boxes represent Caps, and white boxes Cups.

Our experiments indicated that the Shape Index is not sufficiently robust for detecting the nose and chin tips. Thus, we propose a novel method based on two common attributes for locating these two landmarks. The first attribute is that they extrude from the rest of the face. To encode this feature we used the *radial map* (Fig. 4(a)). The radial map is a 2D map that represents, at each u, v pixel, the distance of the corresponding (x, y, z) point from the centroid of the object. The second attribute is that most of the normals at nose and chin regions have an outward direction (with respect to the centroid). The *tangent map* (Fig. 4(b)) encodes this feature. It is a 2D map that represents, at each u, v pixel, the cosine value of the angle between the normal vector at the corresponding (x, y, z) point and the radial vector from the centroid of the object. Both maps are subsequently normalized to $[0, 1]$. Their product constitutes the *extrusion map* that represents the conjunction of the above two attributes (Fig. 4(c)). Since the extrusion map depends only on the position of the centroid, it is considered pose invariant. Local maxima of the extrusion map that are also Shape Index's maxima (Cap = 1.0) are candidate landmarks for nose tips and chin tips. Located candidate nose and chin tips are sorted in descending order of significance according to their corresponding extrusion map values. The most significant subset of points is retained. In Fig. 5(b), crossed circles represent candidate nose and chin tips.

By using the extrusion map, the candidate landmarks of the nose and chin tips are decreased significantly. We retain

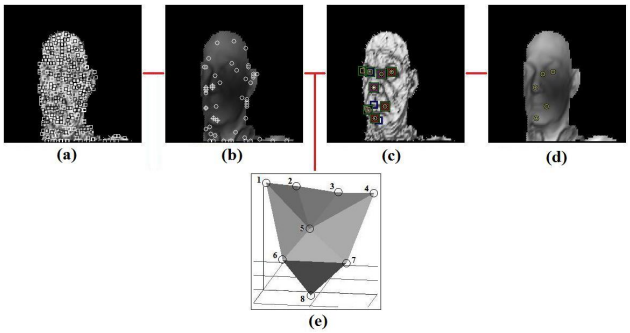


Fig. 5. Landmark detection and selection process: (a) Shape Index's maxima & minima; (b) candidate nose and chin tips; (c) extracted best landmark sets; (d) resulting landmarks; and (e) facial landmark model.

the Shape Index's minima as candidate landmarks for eye and mouth corners (Fig. 5(a)) and extrusion map maxima as candidate landmarks of the nose and chin tips (Fig. 5(b)).

Using the candidate landmark points, we create combinations of five landmarks. Since an exhaustive search of all possible combinations of the candidate landmarks is not feasible, simple length constraints from the shape model and its deformations (FLM) are used to reduce the search space (pruning). From all the feasible candidate five-landmark sets, the ones that do not conform with either FLM5L or FLM5R are filtered out. This is accomplished by applying the fitting procedure that has been previously described.

The final step is to fuse them in complete landmark sets of eight landmarks that conform with the FLM8. From the three available sets (FLM5R, FLM5L, FLM8), the one that has the minimum Procrustes distance to the corresponding model is considered the final solution.

In Fig. 5(c), blue boxes represent landmark sets consistent with the FLM5R, red boxes with the FLM5L, green boxes with the FLM8, and yellow boxes the best landmark set. 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 the best solutions for side facial datasets (Fig. 5(d)). Finally, using the best solution, the pose is estimated, and the facial dataset is classified as frontal, left side or right side (based on the rotation angle with respect to the vertical axis).

C. Annotated Face Model

The *Annotated Face Model* (AFM) is an anthropometrically correct 3D model of the human face [13]. It is constructed only once and is used in the alignment, fitting, and metadata generation [15]. The AFM defines the control points of subdivision surfaces and it is annotated into different areas (e.g., mouth, nose, eyes). Using a global UV parameterization of the AFM, we can convert the polygonal representation of the model to an equivalent geometry image representation.

A *geometry image* is the result of mapping all vertices of a 3D object (x, y and z coordinates) to a 2D grid representation (u, v coordinates) [14]. Thus, a geometry image is a regular continuous sampling of a 3D model represented as a 2D

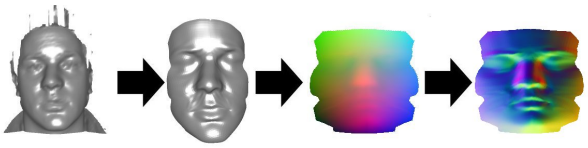
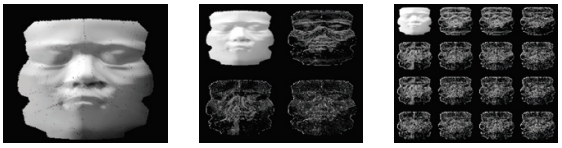


Fig. 6. From left to right, for a frontal facial scan: Raw data → Fitted AFM → Extracted geometry image → Computed normal image.



(a) (b) (c)

Fig. 7. Wavelet analysis of a frontal facial geometry image. For visualization purposes, the coefficient's magnitude is mapped as greyscale intensity: (a) Original image; (b) First level Walsh transform; (c) Second level Walsh transform. [15]

image, with each u, v pixel corresponding to the original x, y, z coordinates. 2D 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).

D. Registration

The landmarks provide a common frame of reference for the raw 3D data. However, in order to fit the AFM, a finer registration is needed. In contrast with the previous step that computes only an initial pose, this step provides a tight rigid registration between the raw data and the AFM.

We utilize the registration algorithm presented by Papaioannou *et al.* [25] that uses a global optimization technique (Simulated Annealing [16], [31]) applied to depth images. The Simulated Annealing process minimizes the following objective function:

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

where R is the spatial resolution of the buffer and D is the z-buffer (normalized to $[0,1]$).

Note that we assume that the initial pose is roughly correct, thus limiting the translation and rotation limits of Simulated Annealing. This step, therefore, can only fine tune the registration – it cannot alleviate errors caused by completely wrong landmark detection. Also, 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 have missing data.

E. Deformable Model Fitting

The purpose of fitting the model to the data is to capture the geometric information of the desired object. In order to fit the AFM to the raw data, a deformable model framework [15] 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 deform 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.

In this paper, we extend this technique to handle partial data by incorporating the notion of *symmetric fitting*. In areas of missing data, the computed external forces will distort the deformation. Therefore, in these cases, mirrored external forces from the model's symmetric area are used. The internal forces are not affected and remain unmodified in order to ensure the continuity of the fitted surface. In our implementation, facial scans that are classified as *frontal* from the landmark detector do not use any symmetric fitting. Facial scans that are classified as *left* use symmetric fitting for the whole right side and vice versa.

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

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

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 $\vec{f}_q$ are the external forces. For data fitting purposes we used $\mathbf{M} = O$ and $\mathbf{D} = O$. 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:

$$\begin{aligned} E_{fo} &= \tfrac{1}{2}\kappa_{fo}\vec{q}^T\mathbf{K}_{fo}\vec{q}, \\ E_{so} &= \tfrac{1}{2}\kappa_{so}\vec{q}^T\mathbf{K}_{so}\vec{q}, \\ E_{sp} &= \tfrac{1}{2}\kappa_{sp}\vec{q}^T\mathbf{K}_{sp}\vec{q} \end{aligned}$$

where κ_{fo} ,κ_{so}, κ_{sp} are the individual weights.

Finite Element Method: In our implementation, we employed the subdivision-based Finite Element Method approximation proposed by Mandal [22]. This approximation solves the above equations in an iterative way. We build a Loop subdivision surface using the AFM as the control mesh. The Loop subdivision scheme [18] is used here 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.

When the deformation stops, 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. We also compute the normal image (equivalent to the first derivative of the geometry data). This process is depicted in Fig. 6 for a frontal facial scan.

F. 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 above, even if half of the

face is missing, the derived geometry and normal images describe the full face. Clearly, there is redundant information for side scans, as half of the geometry and normal image 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 the wavelet analysis. The Walsh wavelet transform [33] 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 filterbank, a two-channel multirate filterbank based on the Haar conjugate mirror filter. The choice of Haar wavelets was based on their properties. The transform is conceptually simple and computationally efficient. The Haar wavelet transform is performed by applying a low-pass filter and a high-pass filter on a one-dimensional input, then repeating the process on the two resulting outputs. Since we are working with images, there will be four outputs for each level of the Haar wavelet (Low-Low, Low-High, High-High, High-Low). We compute a level 4 decomposition, meaning that we apply the filters four times, which yields 256 16×16 wavelet packets (Fig. 7).

Each packet contains a different amount of energy from the initial image. It is possible to ignore most of the packets without losing significant information and store the same subset of the most significant coefficients as metadata. This allows an efficient direct comparison of the selected coefficients of two images (approximately 15%) without the need for reconstruction, by using a weighted L^1 distance metric. The weights are empirically selected and depend on the annotation of the face model.

IV. Results

A. Databases

For performance evaluation we combined the largest publicly available 3D face and ear databases. For frontal facial scans, we use the FRGC v2 database [29]. 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 640×480. These data were obtained from 466 subjects and contain various facial expressions (e.g., happiness, surprise). For side facial scans, we used the Ear Database from the University of Notre Dame (UND) [24], collections F and G. This database (which was created for ear recognition purposes) contains side scans with a vertical rotation of 45^o, 60^o and 90^o. In the 90^o side scans, both sides of the face are occluded from the sensor, therefore they contain no useful information for face recognition purposes. We use only the 45^o side scans (119 subjects, 119 left and 119 right) and the 60^o side scans (88 subjects, 88 left and 88 rights). Note that even though the creators of the database marked these side scans as 45^o and 60^o, the measured average angle of rotation is 65^o and 80^o respectively (Fig. 8). However, when we refer to these scans we will use the database notation (45^o and 60^o).

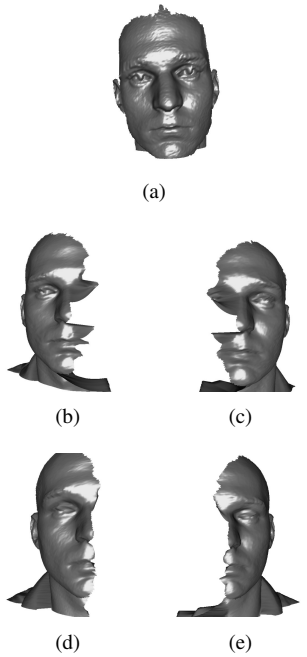


Fig. 8. Scans from the database we constructed: (a) frontal, (b) 45^o right, (c) 45^o left, (d) 60^o right, and (e) 60^o left. Note the extensive missing data in (b-e).

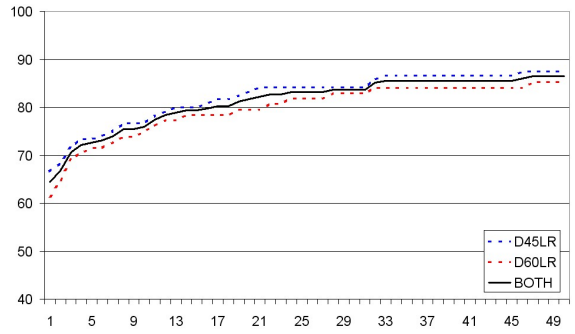
Unfortunately, not all subjects exist in both databases. The number of common subjects between the frontal scans and the 45^o side scans are 39 and between the frontal scans and the 60^o side scans are 33. For our experiments we define the following test databases:

- DB45LR: 45^o side scans from 119 subjects. For each subject, the left scan is considered gallery and the right is considered probe.
- DB60LR: 60^o side scans from 88 subjects. For each subject, the left scan is considered gallery and the right is considered probe.
- DB45F: 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.
- DB60F: Gallery set has one frontal scan for each of the 466 subjects. Probe set has two 60^o side scans (left and right) for each of the 33 subjects.

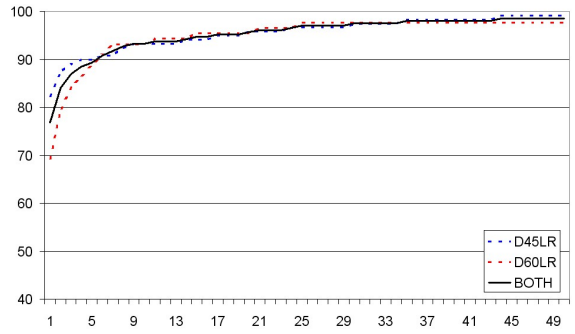
In all cases there is only one gallery scan per subject. Also, all subjects present in the probe set are also present in the gallery set (the opposite is not always true).

B. Performance Evaluation

In order to evaluate performance separately for the landmark detection method and the face retrieval (recognition) method, we manually annotated all scans. Even though our manual annotation may contain inaccuracies, for the purposes of these experiments it will be considered the ground truth. On average, the manually placed landmarks boost the face recognition approach by approximately 10%. Note that 10% is the approximate rate of total failures for the automatic landmark detector, indicating a lack of robustness

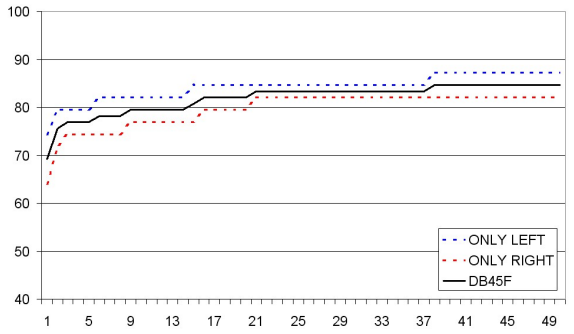


(a)

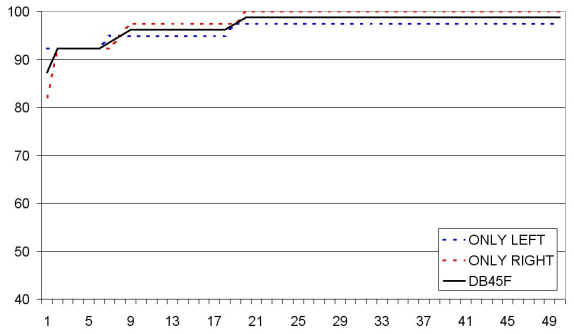


(b)

Fig. 9. CMC graphs for matching left side scans (gallery) with right scans (probe) using DB45LR and DB60LR: (a) automatic landmarks and (b) manually placed landmarks.



(a)



(b)

Fig. 10. CMC graphs for matching frontal scans (gallery) with left and right scans (probe) using DB45F: (a) automatic landmarks and (b) manually placed landmarks.

rather than a lack of accuracy. In all experiments Cumulative Match Characteristic (CMC) graphs are used.

The first experiment is designed to evaluate the performance when a left side scan must be matched with a right side scan. The CMC graphs are given in Fig. 9. The rank–one rate for DB45LR is 67% and 82% for the automatic and manually placed landmarks, respectively. The rank–one rate for DB60LR is 64% and 69% for the automatic and manually placed landmarks, respectively. As expected, the 60^o side scans yield lower results as they are considered more challenging compared to the 45^o side scans.

The second experiment is designed to evaluate the performance when a left or right side scan has to be matched with a frontal scan. This is a common scenario, if gallery acquisition is done in a controlled environment and probe acquisition is uncontrolled. The CMC graphs are given in Fig. 10 and Fig. 11. The rank–one rate for DB45F is 69% and 87% for the automatic and manually placed landmarks, respectively. The rank–one rate for DB60F is 44% and 41% for the automatic and manually placed landmarks, respectively. Note that, for DB60F, using manually placed landmarks decreases the rate for rank–one but increases the rate for all other ranks.

Interestingly, in all experiments, using left side scans for matching performed better than using right scans. Since there is no objective reason for this to happen, we suspect that our implementation of the registration algorithm is slightly biased. Also, the same reason could explain why 60^o left versus 60^o right matching performs better than 60^o left/right versus frontal matching. This contradicts the fact that frontal scans are generally of better quality and hence less challenging.

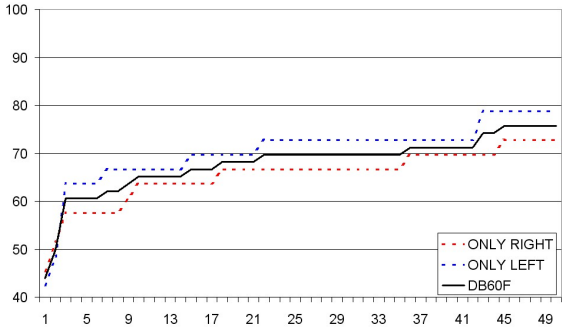
V. Conclusion

We have presented a 3D face recognition (retrieval) method that can handle missing data and offers pose invariance. The proposed method introduced a novel 3D landmark detector and employed a deformable model framework that supports symmetric fitting. It has been evaluated using the most challenging databases available that include pose variations of up to 80^o along the vertical axis. All important steps of the method (landmark detection, registration, fitting) work even when half of the face is missing. Moreover, all scans are represented in a uniform way, allowing partial matching and interpose retrieval. Future work will be directed toward increasing the robustness and accuracy of the landmark detector [28]. Additionally, the registration algorithm will be improved for unbiased registration of both frontal and side facial scans.

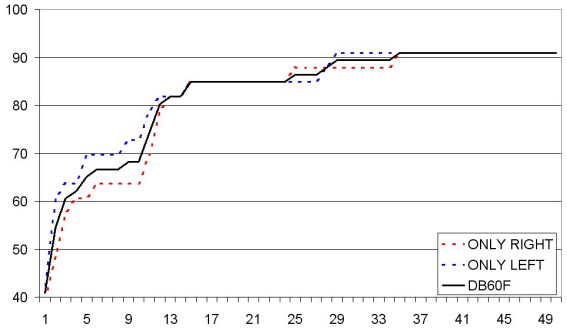
References

[1] V. Blanz, K. Scherbaum, and H-P. Seidel, *Fitting a morphable model to 3D scans of faces*, Proc. 11^{th} IEEE International Conference on Computer Vision, 2007.

[2] V. Blanz and T. Vetter, *Face recognition based on fitting a 3D morphable model*, IEEE Transactions on Pattern Analysis and Machine Intelligence (2003).



(a)



(b)

Fig. 11. CMC graphs for matching frontal scans (gallery) with left and right scans (probe) using DB60F: (a) automatic landmarks and (b) manually placed landmarks.

[3] K.W. Bowyer, K.I. Chang, and P.J. Flynn, *A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition*, Computer Vision and Image Understanding **101** (2006), no. 1, 1–15.

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

[5] K.I. Chang, K.W. Bowyer, and P. J. Flynn, *An evaluation of multi-modal 2D+3D face biometrics*, IEEE Transactions on Pattern Analysis and Machine Intelligence **27** (2005), no. 4, 619–624.

[6] D. Colbry, *Human face verification by robust 3D surface alignment*, Ph.D. thesis, Michigan State University, 2006.

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

[8] T. Cootes and C. Taylor, *Statistical models of appearance for computer vision*, Tech. report, University of Manchester, Oct 2001.

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

[10] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, *Active shape models - their training and application*, Computer Vision and Image Understanding **61** (1995), no. 1, 38–59.

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

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

[13] L.G. Farkas, *Anthropometry of the head and face*, second ed., Raven Press, 1994.

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

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

[16] S. Kirkpatrick, C. Gelatt, and M. Vecchi, *Optimization by simulated annealing*, Science **22** (1983), no. 4598, 671–680.

[17] T. Lin, W. Shih, W. Chen, and W. Ho, *3D face authentication by mutual coupled 3D and 2D feature extraction*, Proc. 44^{th} ACM Southeast Regional Conference (Melbourne, FL), Mar. 10-12 2006.

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

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

[20] X. Lu and A.K. Jain, *Automatic feature extraction for multiview 3D face recognition*, Proc. 7^{th} International Conference on Automatic Face and Gesture Recognition (Southampton, UK), Apr. 10-12 2006.

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

[22] C. Mandal, *A dynamic framework for subdivision surfaces*, Ph.D. thesis, University of Florida, 1998.

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

[24] University of Notre Dame, *University of Notre Dame Biometrics Database*, http://www.nd.edu/@cvrl/UNDBiometricsDatabase.html, 2008.

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

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

[27] G. Passalis, I.A. Kakadiaris, T. Theoharis, G. Toderici, and T. Papaioannou, *Towards fast 3D ear recognition for real-life biometric applications*, Proc. IEEE Advanced Video and Signal based Surveillance (London, UK), Sep. 5-7 2007, pp. 39–44.

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

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

[30] M.P. Segundo, C. Queirolo, O.R.P. Bellon, and L. Silva, *Automatic 3D facial segmentation and landmark detection*, Proc. 14^{th} International Conference on Image Analysis and Processing (Modena, Italy), Sep. 10-14 2007, pp. 431 – 436.

[31] 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 **23** (1997), no. 2, 209–228.

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

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

[34] T. Theoharis, G. Passalis, G. Toderici, and I.A. Kakadiaris, *Unified 3D face and ear recognition using wavelets on geometry images*, Pattern Recognition **41** (2008), no. 3, 796–804.

[35] X. Wei, P. Longo, and L. Yin, *Automatic facial pose determination of 3D range data for face model and expression identification*, Springer, 2007.