3D Facial Landmark Detection under Large Yaw and Expression Variations

Panagiotis Perakis, *Member*, *IEEE Computer Society*, Georgios Passalis, Theoharis Theoharis, and Ioannis A. Kakadiaris, *Senior Member*, *IEEE*

Abstract—A 3D landmark detection method for 3D facial scans is presented and thoroughly evaluated. The main contribution of the presented method is the automatic and pose-invariant detection of landmarks on 3D facial scans under large yaw variations (that often result in missing facial data), and its robustness against large facial expressions. Three-dimensional information is exploited by using 3D local shape descriptors to extract candidate landmark points. The shape descriptors include the *shape index*, a continuous map of principal curvature values of a 3D object's surface, and *spin images*, local descriptors of the object's 3D point distribution. The candidate landmarks are identified and labeled by matching them with a *Facial Landmark Model* (FLM) of facial anatomical landmarks. The presented method is extensively evaluated against a variety of 3D facial databases and achieves state-of-the-art accuracy (4.5-6.3 mm mean landmark localization error), considerably outperforming previous methods, even when tested with the most challenging data.

Index Terms—Face models, landmark detection, shape index, spin images

1 Introduction

In a wide variety of disciplines, it is of great practical importance to measure, describe, process, and compare the shapes of objects; these tasks can be greatly facilitated by using landmark points. In biometric applications, computer vision, and computer graphics, the class of objects is often the human face. Three-dimensional facial landmark detection can be used for face registration, face recognition, facial expression recognition, facial shape analysis, segmentation and labeling of facial parts, facial region retrieval, partial face matching, facial mesh reconstruction, face relighting, face synthesis, face animation, and motion capture. Thus, in almost any application that requires processing of 3D facial data, an initial registration, based on the landmark points' correspondence, is necessary in order to make a system fully automatic [1], [2]. The landmark detector must be pose invariant in order to allow the registration of both frontal and side facial scans [3], [4], [2], [5].

- *P. Perakis and G. Passalis are with the Computer Graphics Laboratory, Department of Informatics and Telecommunications, University of Athens, 17584 Ilisia, Greece, and the Computational Biomedicine Lab, Department of Computer Science, University of Houston, 4800 Calhoun, Houston, TX 77204. E-mail: takis@antinoos.gr, passalis@di.uoa.gr.*
- *T. Theoharis is with the Computer Graphics Laboratory, Department of Informatics and Telecommunications, University of Athens, 17584 Ilisia, Greece, the Department of Computer and Information Science, NTNU, Sem Saelands vei 7-9, NO-7491 Trondheim, Norway, and the Computational Biomedicine Lab, Department of Computer Science, University of Houston, 4800 Calhoun, Houston, TX 77204. E-mail: theotheo@di.uoa.gr, theotheo@idi.ntnu.no.*
- *I.A. Kakadiaris is with the Computational Biomedicine Lab, Department of Computer Science, University of Houston, 4800 Calhoun, Houston, TX 77204. E-mail: ioannisk@uh.edu.*

Manuscript received 21 May 2012; revised 17 Oct. 2012; accepted 12 Nov. 2012; published online 20 Nov. 2012.
Recommended for acceptance by T. Cootes.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2012-05-0385.
Digital Object Identifier no. 10.1109/TPAMI.2012.247.

Even though existing 3D landmark detection methods claim pose invariance, they fail to address large pose variations (Section 2). 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 "almost frontal" scans or "reconstructed" complete facial meshes. *Side scans usually have large missing areas, due to self-occlusion, and the size of the missing areas depends on the amount of pose variation*. These scans are very common in realistic scenarios such as in the case of imaging under uncontrolled conditions.

In this paper, we present a method to automatically detect landmarks (eye and mouth corners, nose, and chin tips) on 3D facial scans that exhibit yaw and expression variations. The main contribution of the presented method is its applicability to large yaw variations (up to 82 degrees) that often result in missing (self-occluded) facial data, and its tolerance against varying facial expressions in an holistic way with high success rates.

In the training phase, a Facial Landmark Model (FLM) representing the landmark positions is created (Fig. 1h), shape index target values for each landmark are computed (Fig. 1f), and spin image templates for each landmark are generated (Fig. 1g). In the detection phase, the algorithm first detects candidate landmarks on the probe facial datasets, by exploiting the 3D geometry-based information of shape index and spin images (Figs. 1a, 1b, and 1c). The extracted candidate landmarks are then filtered out and labeled by matching them with the FLM (Figs. 1d and 1e).

The presented method is evaluated by computing the distance between manually annotated landmarks (ground truth) and the automatically detected landmarks. The experiments have been carried out on two of the largest publicly available databases containing facial datasets: FRGC v2 [6], [7] and the UND Ear Database [8]. The first database contains frontal facial scans with varying expressions, while the second contains side facial scans (both left and right) with up to 82 degrees yaw rotation.

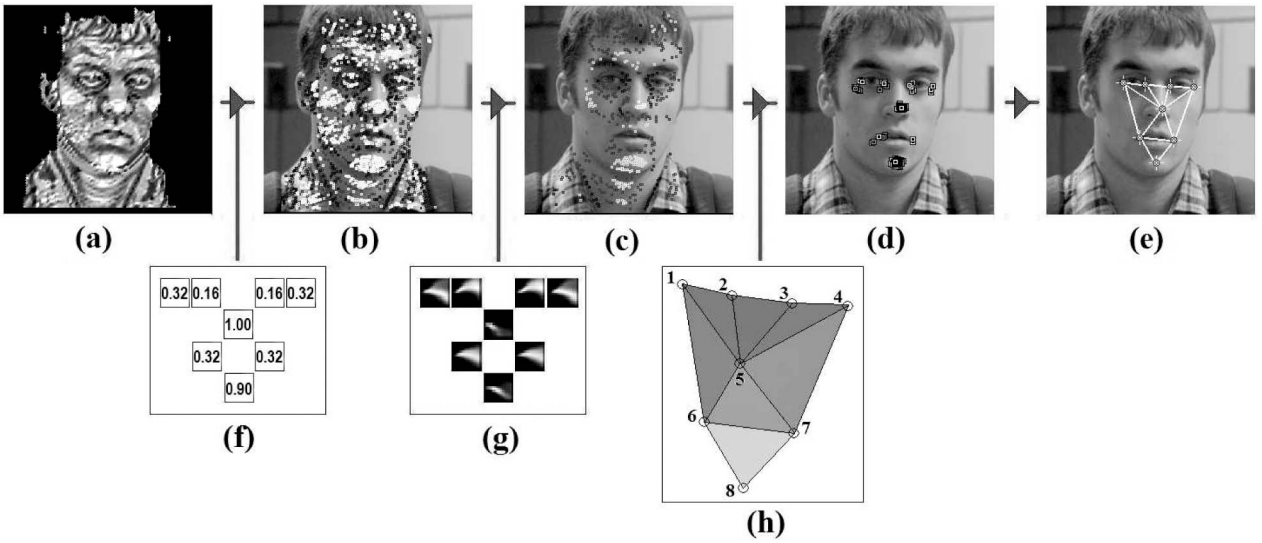


Fig. 1. Process pipeline of landmark detection: (a) shape index map, (b) shape index's candidate landmarks, (c) spin image similarity filtering, (d) extracted landmark sets consistent with FLM, (e) resulting optimal landmark set, (f) shape index target values, (g) spin image templates, and (h) FLM.

In previous work, we have introduced methods for detecting landmarks on 3D facial scans [3], [4], [2], but the aim of these papers was face segmentation, registration, and recognition; 3D landmark detection was thus a very small part which was briefly presented and not thoroughly evaluated. This paper builds on our previously published work, but focuses solely on 3D facial landmark detection, introducing several novelties that lead to state-of-the-art landmark detection results. The method presented in this paper achieves high landmark detection rates in databases that contain faces with large yaw and expression variations (Tables 5 and 2). We also present detailed analytical comparative results against existing state-of-the-art methods (Tables 3 and 4).

In addition to previously published work, here we use FLMs, shape index target values, and spin image templates that are trained from a specific subset of the FRGC v2 database that contains 300 facial scans with varying expressions that are not used in the evaluation experiments (Fig. 11). The inclusion of facial expressions into the FLMs and the use of separate shape index target values for each individual landmark resulted in an improved landmark detection accuracy (by up to 28 percent) and an improved landmark detection rate (by up to 16 percent) compared to the results that were obtained in our previous work [2] (Tables 3 and 4).

The rest of this paper is organized as follows: Section 2 describes related work in the field, Sections 3 and 4 detail the presented method, Section 5 presents our results, and Section 6 summarizes our method.

2 Related Work

Development of 3D modeling and digitizing techniques has sparked research interest in 3D facial feature extraction for landmark detection and is reported in a number of publications.

Professor Jain's group [9], [10], [11], [12], [13] presented methods to locate the positions of eye and mouth corners, and nose and chin tips, based on a fusion scheme of shape index [14] on range maps and the "cornerness" response [15] on intensity maps. They also developed a heuristic method based on cross-profile analysis to locate the nose tip more robustly. In contrast to our approach, candidate landmark points were filtered out using a static (non-deformable) statistical model of landmark positions. The 3D feature extraction method presented in [10] addressed the problem of pose variations and was tested against a composite database consisting of 953 scans from the FRGC database and 160 scans from a proprietary database with frontal scans, extended with variations of pose, expressions, occlusions, and noise. Their multimodal algorithm [9] used 3D+2D information and was applicable to almost-frontal scans (< 5 degrees yaw rotation). It was tested against the FRGC database with 946 near frontal scans. The 3D feature extraction method presented in [11] also addressed the problem of pose variations, and was tested against the FRGC database with 953 near frontal scans along with their proprietary MSU database consisting of 300 multiview scans (0, ± 45 degrees) from 100 subjects. Results of methods [9], [11], [13] are presented in Table 3, and of method [11] in Table 4.

Conde et al. [16] introduced a global face registration method by combining clustering techniques over discrete curvature and spin images for the detection of eye inner corners and of the nose tip. The method was tested on a proprietary database of 714 scans (51 subjects with 14 captures each), with small pose variations (< 15 degrees yaw rotation). Although they presented a feature localization success rate of 99.66 percent on frontal scans and 96.08 percent on side scans, they did not define what constitutes a successful localization.

Xu et al. [17] presented a feature extraction hierarchical scheme to detect the positions of the nose tip and nose ridge. They introduced the "effective energy" notion to describe the local distribution of neighboring points of nose tips and a Support Vector Machine classifier to select the correct nose tips. Although it was tested against various

databases, exact landmark localization results were not provided.

Lin et al. [18] 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 to be the extreme vertex along the normal direction of eye sockets. The method was tested on 27 faces with various poses and expressions in an automatic 3D face authentication system.

Segundo et al. [19] introduced a face and facial feature detection method by combining 2D face segmentation on depth images with surface curvature information in order to detect the eye corners, nose tip, nose base, and nose corners. Although they claimed over 99.7 percent correct detections on the FRGC v2 database, they did not provide a definition of what a correct detection is. Additionally, nose and eye corner detection was not robust under significant pose variations (>15 degrees yaw and roll).

Wei et al. [20] introduced a nose tip and nose bridge localization method, based on a Surface Normal Difference algorithm and shape index estimation in order to determine the facial pose in pose-variant systems. They reported an angular error of the nose tip—nose bridge segment less than 15 degrees in 98 percent of the 2,500 datasets of the BU-3DFE database.

Mian et al. [21] introduced a heuristic method for nose tip detection that was based on a geometric analysis of the nose ridge contour. It was used in a face recognition system to pose correct the facial data. However, no clear localization error results were presented. Additionally, their nose tip detection algorithm had limited applicability to near frontal scans (<15 degrees yaw and pitch).

Faltemier et al. [22] introduced a heuristic method for nose tip detection, a fusion of curvature and shape index analysis, and of template matching using the Iterative Closest Point registration algorithm. The nose tip detector had a localization error less than 10 mm in 98.2 percent of the 4,007 facial datasets of FRGC v2 where it was tested. However, no exact landmark localization results were provided. They also introduced a method called "Rotated Profile Signatures" [23], 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 [8], which contains 7,317 facial scans, 406 frontal and 6,911 in various yaw and pitch angles. They reported a 96 to 100 percent success rate, with distance error threshold 10 mm, under significant yaw and pitch variations. Although their method achieved high success rates, it was limited to the detection of the nose tip only, for which exact localization distance error results were not presented. In addition, it is a 2D-assisted 3D method since it uses skin segmentation to eliminate outliers.

Dibeklioğlu et al. [24], [25] presented methods for detecting facial features on 3D facial datasets to enable pose correction under significant pose variations. They introduced a statistical method to detect facial features, based on training a model of local features, from the gradient of the depth map. The method was tested against the FRGC v1 and the Bosphorus databases, but data with pose variations were not considered. They also introduced a nose tip localization and segmentation method using curvature-based heuristic analysis that was tested against the Bosphorus database, which consists of 3,396 facial scans obtained from 81 subjects. However, the proposed system exhibited limited capabilities on facial datasets with yaw rotations greater than 45 degrees. In addition, no exact landmark localization distance error results were presented.

Yu and Moon [26] presented a nose tip and eye inner corners detection method on 3D range maps. The landmark detector was trained from example facial data using a genetic algorithm and was applied on 200 almost-frontal scans from the FRGC v1 database. However, a limitation of that system is that it is not applicable to facial datasets with large yaw rotations since the three aforementioned control points-landmarks (nose tip and eye inner corners) that were used are not always visible. Results of the method are presented in Table 3.

Romero-Huertas and Pears [27] presented a graph matching approach to locate the positions of nose tip and inner eye corners. They introduced the "distance to local plane" notion to describe the local distribution of neighboring points and detect convex and concave areas of the face. After the graph matching algorithm eliminated false candidates, the best combination of landmark points was selected, based on the minimum Mahalanobis distance to the trained landmark graph model. The method was tested against the FRGC v1 (509 scans) and FRGC v2 (3,271 scans) databases. They reported a success rate of 90 percent with thresholds for the nose tip at 15 mm and for the inner eye corners at 12 mm, but exact landmark localization distance error results were not presented.

Nair and Cavallaro [28] presented a method for detecting facial landmarks on 2.5D scans. Their method used the shape index and the curvedness index to extract candidate feature points. A statistical shape model (Point Distribution Model) of feature points is fitted to the facial dataset dataset by using three landmark points (nose tip and left and right inner eye corners) for coarse registration, and the rest for fine registration. The localization accuracy of the landmark detector was assessed using the BU-3DFE facial database, which contains only complete frontal facial datasets reconstructed from scans captured at ± 45 degrees of yaw [29]. Furthermore, their method is not applicable to missing data resulting from pose self-occlusion since it always uses the aforementioned three landmark points (nose tip and eye inner corners) for model fitting, which are not always visible. Results of the method are presented in Table 3.

3 3D Facial Landmark Model

We use a set of eight anatomical landmarks (Fig. 1h):

1. right eye outer corner (REOC),
2. right eye inner corner (REIC),
3. left eye inner corner (LEIC),
4. left eye outer corner (LEOC),
5. nose tip (NT),
6. mouth right corner (MRC),
7. mouth left corner (MLC), and
8. chin tip (CT).

Note that five of these points are visible on profile and semiprofile face scans. Hence, the complete set of eight landmarks can be used for frontal and almost-frontal faces

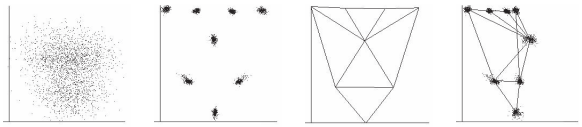


Fig. 2. Depiction of: (a) unaligned landmarks, (b) aligned landmarks, (c) landmarks' mean shape, and (d) landmark clouds and mean shape at 60 degrees.

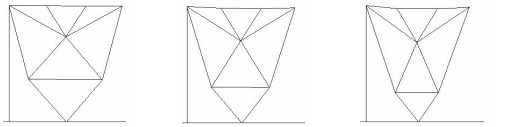


Fig. 3. First mode of FLM8 deformations at 0 degrees.

and two reduced sets of five landmarks (right and left) for semiprofile and profile faces. The right side landmark set and the left side landmark set contain the points (1, 2, 5, 6, 8) and (3, 4, 5, 7, 8), respectively.

Each of these sets of landmarks constitutes a corresponding FLM. Henceforth, 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.

3.1 The Landmark Mean Shape

The mathematical representation of an n-point landmark shape in d dimensions can be defined by concatenating all landmark point coordinates into a $k = nd$ vector and establishing a *Shape Space* [30], [31], [32]. The *vector representation* for 3D landmark shapes would then be

$$\mathbf{x} = [p_{x,1}, \ldots, p_{x,n}, p_{y,1}, \ldots, p_{y,n}, p_{z,1}, \ldots, p_{z,n}]^T, \quad (1)$$

where $(p_{x,i}, p_{y,i}, p_{z,i})$ represent the 3D coordinates of n landmark points.

Since shape has to be invariant to 3D euclidean similarity transformations, translational, scale, and rotational effects need to be filtered out. This procedure is commonly known as *Procrustes Analysis*, and is performed by minimizing the *Procrustes distance* D_P:

$$D_P^2 = |\mathbf{x}_i - \mathbf{x}_m|^2 = \sum_{j=1}^{k} (x_{ij} - x_{mj})^2, \quad (2)$$

between each example shape $\mathbf{x}_i$ and the mean shape $\mathbf{x}_m$. Although there are analytic solutions, a typical iterative approach is used [5] by which the mean shape of landmark shapes (Fig. 2) is computed and example shapes are aligned to it. The resulting mean shape $\mathbf{x}_m$ is the *Procrustes mean*:

$$\mathbf{x}_m = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i \quad (3)$$

of all N example shapes $\mathbf{x}_i$. Since this is an iterative process, the example shapes $\mathbf{x}_i$ are aligned in each iteration step to the current mean shape $\mathbf{x}_m$ [5].

The mean shape for each landmark model (FLM8, FLM5L, and FLM5R) is computed from a manually annotated training set of 300 frontal facial scans of different subjects with varying expressions, which are chosen from the FRGC v2 database subset I (Fig. 11). Training the FLMs with expressions allows the fitting procedure (Section 3.3) to capture candidate landmarks on faces exhibiting expressions.

Note that in our case, where the size of the facial landmark shape is of great importance, scaling shapes to unit size is omitted. In such cases, the shapes are aligned by performing only the translational and rotational transformations. Thus, we use facial landmark distances as constraints that are incorporated during the training phase into the model.

3.2 Landmark Shape Variations

After bringing landmark shapes into a common frame of reference and computing the landmarks' mean shape, further analysis can be done to describe the shape variations. Since facial landmark points represent a certain class of shapes, aligned shape vectors form a specific distribution in the nd-dimensional shape space, which can be modeled by applying PCA to the aligned shapes [30], [31], [32], [33].

Hence, if $\mathbf{A}$ contains (in columns) the p eigenvectors $\mathbf{A}_i$ corresponding to the p largest eigenvalues, λ_i, of the covariance matrix $\mathbf{C_x}$ of the aligned original example shape vectors, the *Facial Landmark Model (FLM)* is created [3], [4], [2], [5] and is represented by the set $\{\mathbf{x}_m, \mathbf{A}_i, \lambda_i\}$, with $i \in \{1, \ldots, p\}$.

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 f of the total variance of the data V_t:

$$\sum_{i=1}^{p} \lambda_i \geq f \cdot V_t, \; V_t = \sum_{i=1}^{k} \lambda_i. \quad (4)$$

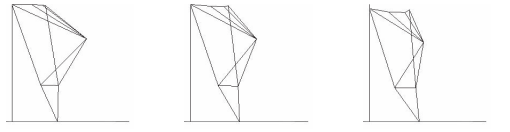
Shape deformations $\mathbf{x}'$ can be modeled by a p-dimensional vector $\mathbf{b}$ of parameters, which represents the *principal modes of variation*:

$$\mathbf{x}' = \mathbf{x}_m + \mathbf{A} \cdot \mathbf{b}. \quad (5)$$

By setting $b_i = \pm 3\sqrt{\lambda_i} = \pm 3\sigma_i$ and all the other $b_j = 0$ we obtain the extreme shape deformations for each mode of variation i [5], which represents $f_i = \frac{\lambda_i}{V_t}$ of the total shape variations of the training datasets [30], [32], [33].

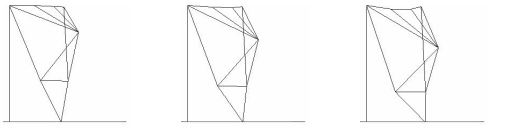
The first mode captures the face size and shape (circular versus oval) and represents 30.6 percent of the total shape variations of FLM8 (Fig. 3). The second mode captures the nose shape (peaked versus flat) and represents 18.8 percent of the total shape variations of FLM8 (Fig. 4). The third mode captures the chin movement (down versus up) due to open mouth and close mouth expressions and represents 9.6 percent of the total shape variations of FLM8 (Fig. 5).

We incorporated 14 eigenvalues (out of the total 24) in FLM8, and seven eigenvalues (out of the total 15) in FLM5L and FLM5R, which represent 99.0 percent of the total shape



(a) $b_2 = -3\sqrt{\lambda_2}$ (b) $b_2 = 0$ (c) $b_2 = +3\sqrt{\lambda_2}$

Fig. 4. Second mode of FLM8 deformations at 70 degrees.



(a) $b_3 = -3\sqrt{\lambda_3}$ (b) $b_3 = 0$ (c) $b_3 = +3\sqrt{\lambda_3}$

Fig. 5. Third mode of FLM8 deformations at 60 degrees.

variations of each model. The least significant eigenvalues that are not incorporated into the FLMs are considered to represent noise [32], [34].

The principal modes represent the marginal deformations of the landmark model (FLM), which are described by the deformation parameters b_i. These are used to establish whether a detected landmark shape is plausible or not (Section 3.3) and for computing the distance constraints of every pair of landmarks (Section 4.3). Note that facial size is incorporated into the FLM by the first deformation parameter b_1. If scale normalization was applied, then size would not be incorporated into the FLM. Thus, at the detection phase, candidate landmark shapes consisting of outlier points (located on the hair or shirt), which are of "small sizes," would eventually be considered as plausible, resulting in more false detections.

3.3 Fitting Landmarks to the FLM

General-purpose feature detection methods are not capable of identifying and labeling the detected candidate landmarks; some topological properties of faces must be taken into consideration. To address the problem of labeling the detected landmarks, we use the FLMs. Candidate landmarks, irrespective of the way they are produced, must be consistent with the corresponding FLM. This is accomplished by fitting a candidate landmark set to the FLM and checking if the deformation parameters $\mathbf{b}$ fall within certain margins [32], [33].

Fitting a set of landmark points $\mathbf{y}$ to the FLM $\{\mathbf{x}_m, \mathbf{A}_i, \lambda_i\}$ is done by minimizing the Procrustes distance $|\mathbf{y} - \mathbf{x}_m|$ in a simple iterative approach [5]. Then, by projecting $\mathbf{y}$ onto the shape eigenspace, its deformation parameters $\mathbf{b}$ are determined as

$$\mathbf{b} = \mathbf{A}^T \cdot (\mathbf{y} - \mathbf{x}_m). \tag{6}$$

We consider a landmark shape $\mathbf{y}$ as plausible if it is consistent with the marginal FLM deformations. Considering that certain b_i of $\mathbf{y}$ satisfy the deformation constraint $|b_i| \leq 3\sqrt{\lambda_i}$, then the candidate landmark shape $\mathbf{y}$ belongs to the shape class with probability

$$Pr(\mathbf{y}) = \frac{\sum \lambda_i}{V_p}, \tag{7}$$

where λ_i are the eigenvalues that satisfy the deformation constraints and V_p is the sum of the eigenvalues that are incorporated into the FLM. If $Pr(\mathbf{y})$ exceeds a certain threshold value, the landmark shape is considered plausible; otherwise it is rejected as a member of the class. The threshold value is set to 0.99 so that only the weakest eigenvalue deformations may not be satisfied since they can be considered as noise.

4 Landmark Detection and Labeling

To detect landmark points, we have used two 3D local shape descriptors that exploit the 3D geometry-based information of facial datasets: shape index and spin images. A facial scan belongs to a subclass of 3D objects which can be considered as a surface S expressed in a general parametric form with native u, v parameterization that allows us to map 3D information into 2D space. Since differential geometry is used for describing the local behavior of surfaces (such as surface curvature and surface normals), we assume that the surface S can be adequately modeled as being at least piecewise smooth. Therefore, to eliminate sensor-specific problems such as white noise, spikes, and holes (especially in areas like the eyebrows and the eyes), certain preprocessing algorithms (*median cut*, *hole filling*, *smoothing*, and *subsampling*) operate directly on the range data before the conversion to polygonal data [1], [2].

4.1 Shape Index

The *Shape Index* [14], [35] is extensively used for 3D landmark detection [13], [11], [12], [10], [9]. It is a continuous mapping of principal curvature values (k_{max}, k_{min}) of a 3D object point $\mathbf{p}$ into the interval $[0, 1]$, and is computed as

$$SI(\mathbf{p}) = \frac{1}{2} - \frac{1}{\pi} tan^{-1} \frac{k_{max}(\mathbf{p}) + k_{min}(\mathbf{p})}{k_{max}(\mathbf{p}) - k_{min}(\mathbf{p})}. \tag{8}$$

The shape index captures the intuitive notion of "local" shape of a surface. Five well-known shape types and their shape index values are: $\mathrm{Cup} = 0.0$, $\mathrm{Rut} = 0.25$, $\mathrm{Saddle} = 0.5$, $\mathrm{Ridge} = 0.75$, and $\mathrm{Cap} = 1.0$.

The shape index is computed from the principal curvature values of the surface spanned by the nearest neighbors of each vertex, a region of 5.5 mm radius on average. After computing the shape index values on a 3D facial dataset, a u, v mapping is performed in order to create a *shape index map* SI_{map} (Fig. 6a):

$$SI_{map}(u, v) \leftarrow SI(x, y, z). \tag{9}$$

To locate interest points on the shape index map, we compute shape index target values that represent the landmarks used. Due to the symmetric nature of the face, shape index target values can represent only five landmark classes (without the distinction of left/right): the eye outer corner, eye inner corner, nose tip, mouth corner, and chin tip landmarks. Shape index target values are statistically generated from 300 manually annotated frontal face scans of different subjects from the FRGC v2 database, subset I (Fig. 11), with varying expressions. The shape index target values for each landmark class are obtained from the mode of the distribution of the shape index values of the associated landmark (Fig. 1f). These values are: 1.00 for

Fig. 6. Depiction of: (a) shape index map (black denotes regions of minimum values and white denotes regions of maximum values, in a gray scale mapping of $[0, 1]$), and (b)-(f) spin image similarity maps: (b) eye outer corner, (c) eye inner corner, (d) nose tip, (e) mouth corner, and (f) chin tip (black denotes regions of low similarity values (-1) and white denotes regions of high similarity values $(+1)$, in a gray scale mapping of $[-1, +1]$).

Fig. 7. Depiction of detected candidate landmarks on texture image (for viewing purposes only): Top: located landmarks according to similarity with shape index target values, and bottom: filtered landmarks according to similarity with spin image templates: (a) eye outer corner, (b) eye inner corner, (c) nose tip, (d) mouth corner, and (e) chin tip.

nose tips, 0.90 for chin tips, 0.32 for mouth corners, 0.32 for eye outer corners, and 0.16 for eye inner corners. The shape index candidate landmarks that are located for each class are kept in five lists sorted in descending order of significance according to their absolute difference from the corresponding shape index target values. The most significant subset of points from each list is retained (a maximum of 1,024 points for each landmark class) (Fig. 7).

However, our experiments indicated that the shape index alone is not sufficiently robust for detecting landmarks on facial datasets in a variety of poses and expressions (the candidate landmarks are too many, having a large number of outliers that lead to false detections). Thus, candidate landmarks located from the shape index values serve as a basis, but are further classified and filtered out (Fig. 7).

4.2 Spin Images

A *Spin Image* [36] 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 surface. 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 degrees. Thus, a spin image is a descriptor of the global or local shape of the object, invariant under rigid transformations. Locality is expressed by the size of the spin image grid and the size of the grid cells (bins). For the purpose of representing facial features on 3D facial datasets, it was experimentally determined that a 16×16 spin image grid with 2 mm bin size should be used. This represents the local shape of the neighborhood of each landmark, spanned by a cylinder of 3.2 cm height and 3.2 cm radius.

To identify interest points on 3D facial datasets, we create spin image templates that represent the classes of the landmarks used. Due to the symmetric nature of the face, spin image templates can represent only five classes (without the distinction of left/right): the eye outer corner, eye inner corner, nose tip, mouth corner, and chin tip landmarks. Spin image templates are statistically generated from 300 manually annotated frontal face scans of different subjects, from the FRGC v2 database, subset I (Fig. 11) with varying expressions. They represent the mean spin images associated with the five classes of the landmarks (Figs. 1g and 8).

Landmark points can be identified according to a similarity measure of their spin images P with the five spin image templates Q that represent each landmark class. This similarity measure is expressed by the normalized linear correlation coefficient:

$$S(P,Q) = \frac{N\sum p_i q_i - \sum p_i \sum q_i}{\sqrt{\left[N\sum p_i^2 - \left(\sum p_i\right)^2\right]\left[N\sum q_i^2 - \left(\sum q_i\right)^2\right]}}, \quad (10)$$

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

The *spin image similarity maps* S_{map} (Figs. 6b, 6c, 6d, 6e, and 6f) provide an insight into the discriminating power of

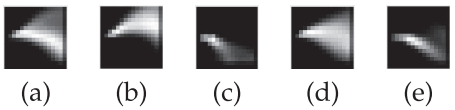


Fig. 8. Depiction of spin image templates: (a) eye outer corner, (b) eye inner corner, (c) nose tip, (d) mouth corner, and (e) chin tip.

each spin image template. They are a u, v mapping of the $S(P, Q)$ value between the spin image P of every facial dataset point and each spin image template Q:

$$S_{map}(u, v) \leftarrow S(P(x, y, z), Q). \quad (11)$$

Spin image templates for the eye inner corner and the nose tip have the highest discriminating power since high similarity areas are located at the expected facial regions, even though the nose tip template has some similarity with eyebrows and chin regions. The spin image template for the chin tip has a medium discriminating power since it has similarity with eyebrows and nose regions. Finally, the spin image templates for the eye outer corner and the mouth corner have the lowest discriminating power since there is high similarity between them and also with other regions of the face, such as the cheeks and forehead. These error-prone regions can be filtered out by using the shape index values.

Therefore, instead of searching all points of a facial dataset to determine the correspondence with the spin image templates, we use the shape index's candidate landmark points. Thus, the candidate landmark points of the five landmark classes (eye outer corner, eye inner corner, nose tip, mouth corner, and chin tip) that are obtained from the shape index map are further filtered out according to the similarity $S(P, Q)$ of their spin images with the spin image templates representing each landmark class. These classified filtered landmarks are sorted in descending order of significance according to their similarity measure with their corresponding spin image template and kept in five lists, one for each landmark class. The most significant subset from each list is retained (a maximum of 160 eye outer corners, 64 eye inner corners, 24 nose tips, 320 mouth corners, and 128 chin tips) (Fig. 7). By using the spin images, the total number of candidate landmarks resulting from the shape index values are significantly decreased and are more robustly localized.

4.3 Landmark Labeling and Selection

The procedure for landmark detection, labeling, and selection is described in Algorithm 1. The detected and classified geometric candidate landmarks from the shape index and the spin image maps are used as the candidate landmarks for eye outer corner, eye inner corner, nose tip, mouth corner, and chin tip (Figs. 1 and 7).

Algorithm 1. Landmark Labeling & Selection

1: Extract candidate landmarks from the geometric properties of the facial scans, using Shape Index and Spin Images (Sections 4.1 and 4.2).
2: Create feasible combinations of 5 landmarks from the candidate landmark points, by using landmark constraints.
3: Compute the rigid transformation that best aligns the combinations of 5 candidate landmarks with the FLM5L and FLM5R.
4: Filter out those combinations that are not consistent with FLM5L or FLM5R, by applying the fitting procedure (Section 3.3).
5: Sort consistent left (FLM5L) and right (FLM5R) landmark sets in descending order according to a distance metric from the corresponding FLM.
6: Fuse accepted combinations of 5 landmarks (left and right) in complete sets of 8 landmarks.
7: Compute the rigid transformation that best aligns the combinations of 8 landmarks with FLM8.
8: Discard combinations of landmarks that are not consistent with FLM8, by applying the fitting procedure (Section 3.3).
9: Sort consistent complete landmark sets in descending order according to a distance metric from FLM8.
10: Select the best combination of landmarks (consistent with FLM5L, FLM5R or FLM8) based on the distance metric to the corresponding FLM.

From the candidate landmark points we create combinations of five landmarks, one from each class. Since an exhaustive search of all possible combinations of the candidate landmarks is not feasible, two types of landmark position constraints are used to reduce the search space (pruning) by removing obvious outliers, thus speeding up the search algorithm.

The *Absolute Distance constraint* captures the fact that the distances between two landmark points must be within certain margins consistent with the absolute face dimensions. Distance constraints are created from the marginal shape variations of FLM8.

The *Relative Position constraint* captures the fact that the relative positions of landmark points must be consistent with the face shape. Considering the nose tip as a center, all other landmarks must lie in a counterclockwise direction for FLM5L and in a clockwise direction for FLM5R.

Note that the use of candidate landmark sets with five landmarks has a dual purpose: 1) It is the potential solution for semiprofile and profile faces, and 2) it reduces the combinatorial search space for creating the complete landmark sets in a divide-and-conquer manner. Instead of creating 8-tuples of landmarks out of N candidates, which generates N^8 combinations to be checked for consistency with FLM8, we create 5-tuples of landmarks and check $2N^5$ combinations for consistency with FLM5L and FLM5R. We retain 512 landmark sets consistent with FLM5L and 512 landmark sets consistent with FLM5R. By fusing them and checking consistency with FLM8 we obtain an extra 512×512 combinations to be checked. Thus, by this approach, $2N^5 + 512^2 \ll N^8$ combinations are checked, with $O(N^5) \ll O(N^8)$. For $N = 128$, we obtain approximately 69×10^9 instead of 72×10^{15} combinations to be checked.

To find the optimal solution, the three available consistent lists of landmark sets (left, right, and complete) are sorted in descending order according to a distance measure from the corresponding model (FLM5L, FLM5R, FLM8). The landmark set (left, right, or complete) that has the minimum distance measure is identified as the optimal solution (Figs. 1 and 9).

Since FLM5L, FLM5R, and FLM8 have different dimensions k in shape space, Procrustes distances D_P (2) cannot be used as a distance measure as they are not directly comparable. Thus, we must use alternative measures for the distance between two landmark shapes that can be comparable irrespective of their dimensions.

An intuitive *normalized Procrustes distance* D_{NP} that takes into consideration the shape space dimensions k is

(a) (b) (c)

Fig. 9. METHOD SISI-NPSS: Results of landmark detection and selection process: Top: extracted landmark sets consistent with FLM8 (white) and FLM5L or FLM5R (gray) and bottom: resulting optimal landmark set. (a) Face at 45 degrees yaw, (b) face at 60 degrees yaw, and (c) frontal face with extreme expression.

$$D_{NP} = \frac{D_P}{k^2}. \tag{12}$$

The division by k^2 instead of k is preferred to give a bias to the complete solution.

A nongeometric measure of the quality of a landmark shape is its *mean spin image similarity* D_{SS} normalized to $[0, 1]$ (0 for high and 1 for low similarity):

$$D_{SS} = \frac{1}{2}\left[1 - \frac{\sum_{i=1}^{n} S(P_i, Q_i)}{n}\right], \tag{13}$$

where $S(P_i, Q_i)$ is the similarity measure between the landmark spin image P_i and the corresponding template Q_i, and n is the number of landmarks.

Thus, an intuitive *normalized Procrustes* $\times$ *mean spin similarity distance* D_{NPSS} that takes into consideration the geometric distance and the spin image similarities can be defined as:

$$D_{NPSS} = D_{NP} \cdot D_{SS}. \tag{14}$$

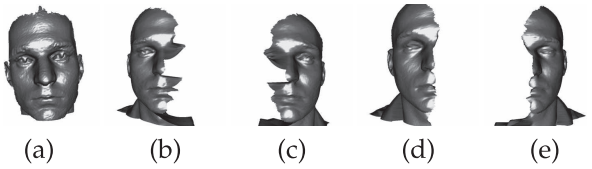
The D_{NPSS} distance metric is used in the presented landmark detection method, which will henceforth be called "METHOD SISI-NPSS."

In Fig. 9 (top), gray boxes represent landmark sets consistent with FLM5L and FLM5R, while white boxes are landmark sets consistent with FLM8. Note that the FLM8 consistent landmark set is not always the best solution; FLM5L and FLM5R are better solutions for semiprofile and profile facial datasets (Fig. 9 (bottom)).

5 Landmark Localization Results

5.1 Test Databases

To evaluate the performance of the presented landmark detector, we used two of the largest publicly available 3D face databases [8]. For frontal facial datasets, we used the *FRGC v2* database [6], [7]. The FRGC v2 database contains a total of 4,007 range images of 466 individuals. Subjects have almost frontal poses and various facial expressions (e.g., happiness and surprise). Hence, FRGC v2 is more challenging than FRGC v1. For the purposes of this evaluation we manually annotated 975 frontal facial datasets obtained from 149 different subjects, selected from the FRGC v2 database subset II (Fig. 11), including several subjects with various facial expressions. This database will henceforth be referred as *DB00F* (Fig. 10a). To quantitatively assess the performance of our 3D landmark detector on facial datasets with varying degrees of expressions, we manually classified the *DB00F* datasets into three subclasses according to expression intensities: "*neutral*," "*mild*," and "*extreme*."



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

Fig. 10. Depiction of frontal view of scans from the database used: (a) frontal (*DB00F*), (b) 45 degrees right (*DB45R*), (c) 45 degrees left (*DB45L*), (d) 60 degrees right (*DB60R*), and (e) 60 degrees left (*DB60L*). Note the extensive missing data in (b)-(e).

For semiprofile and profile facial datasets, we used the *Ear Database* from the University of Notre Dame (UND), collections F and G [8]. This database (which was created for ear recognition purposes) contains 119 side scans of 119 subjects at ± 45 degrees and 88 side scans of 88 subjects at ± 60 degrees. Note that though the creators of the database marked these side scans as 45 and 60 degrees, the computed maximum angle of yaw rotation is 69 and 82 degrees, respectively (Table 5). For the purposes of this evaluation, we manually annotated 118 right and 118 left, 45 degrees side datasets, obtained from 118 different subjects. These databases will be referred to as *DB45L* and *DB45R*, respectively (Figs. 10b and 10c), and their union is *DB45RL*. We also annotated 87 right and 87 left 60 degrees side datasets, obtained from 87 different subjects. These databases will be referred to as *DB60L* and *DB60R*, respectively (Figs. 10d and 10e), and their union is *DB60RL*. Finally, we composed a database with datasets of 39 common subjects found in *DB00F*, *DB45R* and *DB45L*. This database consists of 117 (3×39) scans in three poses, frontal and 45 degrees left and right, and will henceforth be referred to as *DB00F45RL*.

In the evaluation databases, only facial datasets with all landmark points visible were included (eight for frontal scans and five for side scans). The exact datasets that were used from the source databases for training and testing can be found from the landmark annotation files available through our website [37].

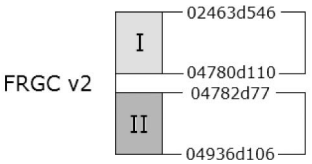


Fig. 11. FRGC v2 partitioning: (I) 300 facial scans for training FLMs, shape index target values, and spin image templates, and (II) 975 facial scans for testing.

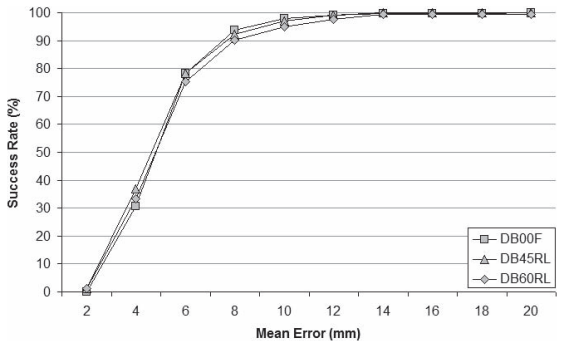


Fig. 12. Mean error cumulative distribution of METHOD SISI-NPSS on *DB00F*, *DB45RL*, and *DB60RL*.

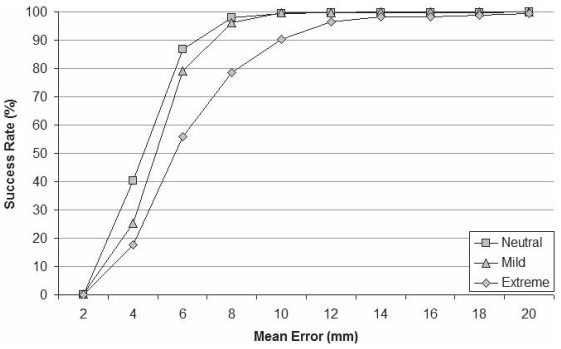


Fig. 13. Mean error cumulative distribution of METHOD SISI-NPSS on *DB00F* "*neutral*," "*mild*," and "*extreme*."

5.2 Performance Evaluation

To evaluate the performance of the presented landmark detection method, we conducted the following two experiments: In *Experiment 1*, we evaluated the performance of *Method SISI-NPSS* against yaw variations, and in *Experiment 2*, we evaluated the tolerance of *Method SISI-NPSS* against expression variations.

The performance evaluation of a landmark detector is generally presented by computing the following values, which represent the localization accuracy of the detected landmarks.

Absolute distance error. The euclidean distance in physical units (e.g., mm) between the position of the detected landmark and the manually annotated landmark, which is considered ground truth.

Detection success rate. The percentage of successful detections of a landmark over a test database. Successful detection is considered as the detection of a landmark with absolute distance error under a certain threshold (e.g., 10 mm).

In our experiments, the *localization error* is represented by the mean and standard deviation of the absolute distance error of the detected landmarks. Also, the overall mean distance error of the eight landmark points for the frontal datasets and of the five landmark points for the side datasets was computed.

The *success rate of landmark localization* with an absolute distance error threshold of 10 mm is reported in the result tables. Note that, as pointed out in [2], our UR3D-S face recognition method can tolerate landmark localization errors up to 10 mm.

The yaw angle of probe faces is computed and its mean value, standard deviation, and minimum and maximum values are presented. The yaw angle results from the rotational transformation of the optimal solution that fits the probe face to the corresponding FLM and thus the probe face is classified as frontal, left side, or right side. Side detection can be crucial in determining follow-up actions in a biometric system. The side detection rate reported in the result tables is the percentage of correct side estimations of the probe faces with respect to their ground-truth side and whose detected landmarks also have an overall mean distance error under 30 mm.

We depict the Cumulative Distribution graph of the mean distance error in Fig. 13 to show the method's tolerance to expression variations and in Fig. 12 to show the method's robustness to yaw rotations. In these graphs, the x-axis represents the mean distance error between the manually annotated landmarks and the automatically detected landmarks in intervals of 2 mm, and the y-axis represents the percentage of face datasets with a mean distance error up to a certain x-value, out of all gallery datasets.

Summary results for METHOD SISI-NPSS on all tested databases are presented in Table 1. The results clearly indicate that our method exhibits high accuracy and robustness both to yaw and expression variations. The mean error is under 6.3 mm, with standard deviation under 2.6 mm on all tested facial scans. Also note that the mean error is under 10 mm for at least 90.4 percent of the tested facial scans and the facial side was correctly estimated on over 98.9 percent of the tested facial scans.

Specifically, the best results were obtained for the frontal facial scans category and the worst for the 60 degrees facial scans. This is due to the fact that, as the yaw angle increases, landmark detection becomes more difficult, mainly due to distortions on their shape index and spin image values caused by the missing data around the nose and chin tip regions (Figs. 10b, 10c, 10d, and 10e). The results that assess the robustness of METHOD SISI-NPSS against yaw variations are presented in Table 5 and Fig. 12.

The most robust facial features are the nose tip and eye inner corners, with a lower mean error and standard deviation across yaw rotations and expression variations. This is due to the fact that they have more distinct geometry, which is more easily captured by the detectors, and there are no substantial changes in their shape index and spin image

TABLE 1
Summary Results for METHOD SISI-NPSS

Database	Mean Error			Side Detection Rate
	mean (mm)	std.dev (mm)	≤ 10 (mm)	
DB00F	5.00	1.85	97.85%	99.90%
DB00F-neutral	4.52	1.51	99.32%	100.00%
DB00F-mild	4.95	1.46	99.72%	100.00%
DB00F-extreme	6.28	2.60	90.40%	99.44%
DB00F45RL	4.97	1.92	97.44%	100.00%
DB45R	5.03	1.92	96.61%	100.00%
DB45L	4.75	1.91	97.46%	100.00%
DB60R	4.95	1.80	96.55%	98.85%
DB60L	5.30	2.49	93.10%	100.00%

TABLE 2
Experiment 2: METHOD SISI-NPSS Tolerance to Expression Variations on *DB00F*

Expression	Neutral			Mild			Extreme			All		
Side Detection Rate	443 / 443		100.00%	355 / 355		100.00%	176 / 177		99.44%	974 / 975		99.90%
Localization Error (mm)	mean	std.dev.	$\leq$ 10 mm	mean	std.dev.	$\leq$ 10 mm	mean	std.dev.	$\leq$ 10 mm	mean	std.dev.	$\leq$ 10 mm
REOC	5.38	3.14	90.97%	5.76	3.42	88.17%	5.71	3.57	84.75%	5.58	3.33	88.82%
REIC	3.95	2.19	97.52%	4.28	2.35	97.46%	4.38	2.68	94.35%	4.15	2.35	96.92%
LEIC	4.37	2.51	98.19%	4.48	2.33	96.90%	4.40	2.74	96.05%	4.41	2.49	97.33%
LEOC	5.66	3.37	89.16%	5.95	3.38	87.61%	6.02	3.59	86.44%	5.83	3.42	88.10%
NT	3.99	2.24	99.10%	3.92	2.06	98.59%	4.67	3.25	97.18%	4.09	2.41	98.56%
MRC	4.25	2.30	99.10%	5.36	3.10	90.14%	9.26	5.88	59.32%	5.56	3.93	88.62%
MLC	4.35	2.40	97.52%	5.21	3.14	91.27%	8.55	5.87	62.15%	5.42	3.84	88.82%
CT	4.21	2.36	98.42%	4.66	2.70	96.34%	7.27	6.45	82.49%	4.92	3.74	94.77%
Mean Error	4.52	1.51	99.32%	4.95	1.46	99.72%	6.28	2.60	90.40%	5.00	1.85	97.85%

TABLE 3
Comparison of METHOD SISI-NPSS against State of the Art on Almost-Frontal Complete Facial Datasets

Mean Localization Error (mm)										
Method		Test DB (scans)	REIC	LEIC	REOC	LEOC	NT	CT	MRC	MLC
Yu et al. [26]	(GA model)	FRGC v1 (200)	4.74	5.59	-	-	2.18	-	-	-
Nair et al. [28]	(w/o PDM)	BU-3DFE (2350)	25.01	26.68	31.84	34.39	14.59	-	-	-
	(w PDM)		12.11	11.89	20.46	19.38	8.83	-	-	-
Lu et al. [11]	(3D)	FRGC v1 (953)	8.30	8.20	9.50	10.30	8.30	-	6.00	6.20
Lu et al. [9]	(3D+2D)	FRGC v1 (946)	6.00	5.70	7.10	7.90	5.00	-	3.60	3.60
Colbry [13]	(w/o CFDM)	FRGC v1 (953)	5.50	6.30	-	-	4.10	11.00	6.90	6.70
	(w CFDM)	+ propr. (160)	5.60	6.00	-	-	4.00	11.70	5.40	5.40
Perakis et al. [3]	(EG-3DOR)	FRGC v2 (975)	7.02	7.46	8.13	9.21	5.23	6.71	8.30	9.83
Passalis et al. [2]		FRGC v2 (975)	5.03	5.48	5.79	5.62	4.91	6.31	5.65	6.47
Perakis et al. (current)	(SISI–NPSS)	FRGC v2 (975)	4.15	4.41	5.58	5.83	4.09	4.92	5.56	5.42
Std. Dev. of Localization Error (mm)										
Method		Test DB (scans)	REIC	LEIC	REOC	LEOC	NT	CT	MRC	MLC
Yu et al. [26]	(GA model)	FRGC v1 (200)	9.76	16.08	-	-	6.83	-	-	-
Nair et al. [28]	(w/o PDM)	BU-3DFE (2350)	-	-	-	-	-	-	-	-
	(w PDM)		-	-	-	-	-	-	-	-
Lu et al. [11]	(3D)	FRGC v1 (953)	17.20	17.20	17.10	18.10	19.40	-	16.90	17.90
Lu et al. [9]	(3D+2D)	FRGC v1 (946)	3.30	3.00	5.90	5.10	2.40	-	3.30	2.90
Colbry [13]	(w/o CFDM)	FRGC v1 (953)	4.90	5.00	-	-	5.10	7.60	8.60	9.30
	(w CFDM)	+ propr. (160)	4.80	4.70	-	-	5.40	7.30	6.80	6.70
Perakis et al. [3]	(EG-3DOR)	FRGC v2 (975)	3.18	3.07	3.79	4.25	3.28	4.32	4.53	4.47
Passalis et al. [2]		FRGC v2 (975)	2.47	2.59	3.45	3.47	2.49	4.43	4.34	4.26
Perakis et al. (current)	(SISI–NPSS)	FRGC v2 (975)	2.35	2.49	3.33	3.42	2.41	3.74	3.93	3.84

TABLE 4
Comparison of METHOD SISI-NPSS against State of the Art on Mixed (Frontal and Profile) Facial Datasets

Mean Localization Error (mm)										
Method		Test DB (scans)	REIC	LEIC	REOC	LEOC	NT	CT	MRC	MLC
Lu et al. [11]	(3D)	MSU (300)	9.00	7.10	13.60	13.30	6.40	-	6.70	5.20
Passalis et al. [2]		FRGC v2 + Ear (117)	5.97	6.87	6.51	6.71	4.60	6.59	5.52	6.10
Perakis et al. (current)	(SISI–NPSS)	FRGC v2 + Ear (117)	4.65	4.90	5.32	6.06	4.41	4.80	5.01	4.91
Std. Dev. of Localization Error (mm)										
Method		Test DB (scans)	REIC	LEIC	REOC	LEOC	NT	CT	MRC	MLC
Lu et al. [11]	(3D)	MSU (300)	13.10	9.20	11.90	10.10	13.40	-	12.90	9.00
Passalis et al. [2]		FRGC v2 + Ear (117)	3.13	2.92	3.68	3.76	3.01	4.16	3.58	4.17
Perakis et al. (current)	(SISI–NPSS)	FRGC v2 + Ear (117)	2.45	2.96	3.71	4.13	2.68	3.52	2.97	2.88

values due to the deformations resulting from facial expressions. The least robust facial feature appears to be the mouth corners, mainly due to the fact that they do not have enough distinct geometry and are also prone to changes in their shape index and spin image values due to the deformations resulting from facial expressions. The results that assess the tolerance of METHOD SISI-NPSS against expression variations are presented in Table 2 and Fig. 13.

5.3 Comparative Results

For comparison of the performance of the presented landmark detection method against other state-of-the-art methods, we present landmark localization errors in Tables 3 and 4. Note that each method uses a different facial database, making direct comparisons difficult. However, these results indicate that METHOD SISI-NPSS outperforms previous methods for the following reasons: 1) It is more accurate since it gives smaller mean localization distance errors for almost all landmarks, and 2) it is more robust since it gives smaller standard deviations for the localization distance error.

Comparative results of landmark localization errors on almost-frontal facial datasets are presented in Table 3. Yu and Moon's method [26] exhibits the minimum mean localization error for the nose tip, but has a large standard deviation. Lu and Jain's method [9] exhibits the minimum mean localization error for the mouth corners, but is not a pure 3D method since it is assisted by 2D intensity data.

Finally, Colbry's method [13] seems to perform well for all landmarks, comparatively close to our method, but has larger standard deviations. Note that the FRGC v1 database used in Yu and Moon [26], Lu and Jain [11], Lu and Jain [9], and Colbry [13] is considered less challenging than the FRGC v2 used in our experiments since FRGC v1 contains subjects with neutral expressions, while FRGC v2 contains subjects with various facial expressions. Furthermore, the database used by Colbry [13] contains a small portion (≈ 5 percent) of proprietary datasets with pose variations, occlusions, and expressions. The BU-3DFE database [29] used in Nair and Cavallaro [28] contains frontal only 3D facial datasets, which were created by the fusion of facial data acquired at ± 45 degrees yaw, from 100 subjects that perform seven universal expressions.

Comparative results of landmark localization errors on mixed (frontal and profile) facial datasets are presented in Table 4. To the best of our knowledge, Lu and Jain's method [11] is the only method in which localization errors on both frontal and profile facial datasets were presented. These results indicate that METHOD SISI-NPSS outperforms Lu and Jain's method in both accuracy and robustness. The proprietary MSU database used in Lu and Jain [11] contains 300 3D facial scans from 100 subjects, three scans for each subject captured at 0 and ± 45 degrees yaw angles. The *DB00F45RL* database used in our experiments, despite having fewer subjects, is considered more challenging, since yaw angles lie in the range $[-65, +67]$ degrees (Table 5).

The inclusion of facial expressions into the FLMs and the use of separate shape index target values for each individual landmark resulted in an improved accuracy of our landmark detector (by up to 28 percent) and an improved detection rate (by up to 16 percent) compared to our early results that appeared in [2].

5.4 Computational Cost

For the evaluation of the presented method's computational efficiency, a PC with the following specifications was used: Intel Core i5 2.5 GHz with 4 GB RAM. Using this PC, 6.68 s (on average) was required to locate the landmarks for each facial scan. The average time taken for each step of the method is: Data loading 0.04 s, shape index computation and landmark localization 0.26 s, spin image computation and landmark filtering 0.31 s, FLM5L-FLM5R matching and landmark labeling 5.05 s, and FLM8 matching and optimal landmark set selection 1.02 s. The procedures for determining the optimal rotation for the alignment of the landmark shapes to the FLMs require at most eight iterations to converge. Speedups through parallelization are possible and thus the computational efficiency of the presented landmark detector makes it applicable to real-world applications.

6 Conclusion

We have presented an automatic 3D facial landmark detector that offers pose invariance and robustness to large missing (self-occluded) facial areas with respect to large yaw variations. It also offers high tolerance to large expression variations. The presented approach consists of methods for 3D landmark localization that exploit the 3D geometry-based information of faces and the modeling ability of trained landmark models. It has been evaluated using the most challenging 3D facial databases available, which contain scans with yaw variations up to 82 degrees

TABLE 5
Experiment 1:
Performance of METHOD SISI-NPSS against Yaw Variations

DB00F

Side Detection Rate	974 / 975		99.90%
Yaw Estimation	$+0.93^\circ \pm 4.03^\circ$	$[-17.98^\circ \sim +16.98^\circ]$	
Localization Error	mean (mm)	std.dev. (mm)	≤ 10 mm
REOC	5.58	3.33	88.82%
REIC	4.15	2.35	96.92%
LEIC	4.41	2.49	97.33%
LEOC	5.83	3.42	88.10%
NT	4.09	2.41	98.56%
MRC	5.56	3.93	88.62%
MLC	5.42	3.84	88.82%
CT	4.92	3.74	94.77%
Mean Error	5.00	1.85	97.85%

DB00F45RL

Side Detection Rate	117 / 117		100.00%
Yaw Estimation	$+1.15^\circ \pm 41.35^\circ$	$[-65.23^\circ \sim +66.82^\circ]$	
Localization Error	mean (mm)	std.dev. (mm)	≤ 10 mm
REOC	5.32	3.71	88.46%
REIC	4.65	2.45	96.15%
LEIC	4.90	2.96	96.15%
LEOC	6.06	4.13	80.77%
NT	4.41	2.68	98.29%
MRC	5.01	2.97	92.31%
MLC	4.91	2.88	96.15%
CT	4.80	3.52	93.16%
Mean Error	4.97	1.92	97.44%

DB45R

Side Detection Rate	118 / 118		100.00%
Yaw Estimation	$+44.20^\circ \pm 8.20^\circ$	$[+16.81^\circ \sim +68.04^\circ]$	
Localization Error	mean (mm)	std.dev. (mm)	≤ 10 mm
REOC	5.63	3.76	85.59%
REIC	4.71	2.69	95.76%
NT	4.87	2.90	95.76%
MRC	4.84	3.50	88.98%
CT	5.12	4.95	91.53%
Mean Error	5.03	1.92	96.61%

DB45L

Side Detection Rate	118 / 118		100.00%
Yaw Estimation	$-45.57^\circ \pm 8.95^\circ$	$[-69.22^\circ \sim -16.19^\circ]$	
Localization Error	mean (mm)	std.dev. (mm)	≤ 10 mm
LEOC	5.42	3.42	88.98%
LEIC	5.05	2.79	94.92%
NT	4.64	2.81	95.76%
MLC	4.21	2.93	96.61%
CT	4.45	3.57	96.61%
Mean Error	4.75	1.91	97.46%

DB60R

Side Detection Rate	86 / 87		98.85%
Yaw Estimation	$+57.47^\circ \pm 7.22^\circ$	$[+30.24^\circ \sim +80.81^\circ]$	
Localization Error	mean (mm)	std.dev. (mm)	≤ 10 mm
REOC	6.01	3.52	80.46%
REIC	4.89	3.11	93.10%
NT	3.94	2.35	96.55%
MRC	4.68	3.36	91.95%
CT	5.23	4.72	89.66%
Mean Error	4.95	1.80	96.55%

DB60L

Side Detection Rate	87 / 87		100.00%
Yaw Estimation	$-58.51^\circ \pm 8.06^\circ$	$[-82.52^\circ \sim -30.79^\circ]$	
Localization Error	mean (mm)	std.dev. (mm)	≤ 10 mm
LEOC	5.14	3.19	91.95%
LEIC	5.01	3.02	95.40%
NT	4.13	1.85	100.00%
MLC	5.37	5.12	85.06%
CT	6.86	6.03	85.06%
Mean Error	5.30	2.49	93.10%

and strong expressions. In these databases, it achieved state-of-the-art accuracy, significantly outperforming (by up to 28 percent) our previously published work [2].

Although it is possible to consider extensions for improving accuracy (e.g., by including the nostrils' base or another anatomical landmark into the FLMs, or by applying heuristic methods of postprocessing for fine-tuning the positions of landmarks), we believe that such improvements will be marginal and at the expense of the method's simplicity and speed. We intend to consider algorithmic and architectural speedup techniques to achieve real-time performance.

ACKNOWLEDGMENTS

This research has been cofinanced in part by: 1) the European Union (European Social Fund—ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF)—Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund, and 2) the University of Houston Eckhard Pfeiffer Endowment Fund. All statements of fact, opinion, or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of the sponsors.

REFERENCES

[1] 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 Trans. Pattern Analysis and Machine Intelligence,* vol. 29, no. 4, pp. 640-649, Apr. 2007.

[2] G. Passalis, P. Perakis, T. Theoharis, and I. Kakadiaris, "Using Facial Symmetry to Handle Pose Variations in Real-World 3D Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 33, no. 10, pp. 1938-1951, Oct. 2011.

[3] P. Perakis, T. Theoharis, G. Passalis, and I. Kakadiaris, "Automatic 3D Facial Region Retrieval from Multi-Pose Facial Data Sets," *Proc. Eurographics Workshop 3D Object Retrieval,* pp. 37-44, Mar./Apr. 2009.

[4] P. Perakis, G. Passalis, T. Theoharis, G. Toderici, and I. Kakadiaris, "Partial Matching of Interpose 3D Facial Data for Face Recognition," *Proc. Third IEEE Int'l Conf. Biometrics: Theory, Applications and Systems,* pp. 439-446, Sept. 2009.

[5] P. Perakis, G. Passalis, T. Theoharis, and I. Kakadiaris, "3D Facial Landmark Detection & Face Registration: A 3D Facial Landmark Model & 3D Local Shape Descriptors Approach," Technical Report TP-2010-01, Computer Graphics Laboratory, Univ. of Athens, Jan. 2010.

[6] 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," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 947-954, 2005.

[7] P. Phillips, T. Scruggs, A. O'Toole, P. Flynn, K. Bowyer, C. Schott, and M. Sharpe, "FRVT 2006 and ICE 2006 Large-Scale Experimental Results," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 32, no. 5, pp. 831-846, May 2010.

[8] UND, "University of Notre Dame Biometrics Data Sets," http://www.nd.edu/~cvrl/CVRL/Data_Sets.html, 2012.

[9] X. Lu and A. Jain, "Multimodal Facial Feature Extraction for Automatic 3D Face Recognition," Technical Report MSU-CSE-05-22, Michigan State Univ., Oct. 2005.

[10] D. Colbry, G. Stockman, and A. Jain, "Detection of Anchor Points for 3D Face Verification," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* p. 118, June 2005.

[11] X. Lu and A. Jain, "Automatic Feature Extraction for Multiview 3D Face Recognition," *Proc. Seventh Int'l Conf. Automatic Face and Gesture Recognition,* pp. 585-590, 2006.

[12] X. Lu, A. Jain, and D. Colbry, "Matching 2.5D Face Scans to 3D Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 28, no. 1, pp. 31-43, Jan. 2006.

[13] D. Colbry, "Human Face Verification by Robust 3D Surface Alignment," PhD dissertation, Michigan State Univ., 2006.

[14] C. Dorai and A.K. Jain, "COSMOS—A Representation Scheme for 3D Free-Form Objects," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 10, pp. 1115-1130, Oct. 1997.

[15] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," *Proc. Fourth Alvey Vision Conf.,* pp. 147-151, 1988,

[16] C. Conde, R. Cipolla, L.J. Rodrígez-Aragón, A. Serrano, and E. Cabello, "3D Facial Feature Location with Spin Images," *Proc. IAPR Conf. Machine Vision Applications,* pp. 418-421, May 2005.

[17] 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.

[18] T. Lin, W. Shih, W. Chen, and W. Ho, "3D Face Authentication by Mutual Coupled 3D and 2D Feature Extraction," *Proc. 44th ACM Southeast Regional Conf.,* pp. 423-427, Mar. 2006.

[19] M. Segundo, C. Queirolo, O. Bellon, and L. Silva, "Automatic 3D Facial Segmentation and Landmark Detection," *Proc. 14th Int'l Conf. Image Analysis and Processing,* pp. 431-436, Sept. 2007.

[20] X. Wei, P. Longo, and L. Yin, "Automatic Facial Pose Determination of 3D Range Data for Face Model and Expression Identification," *Proc. Int'l Conf. Advances in Biometrics,* pp. 144-153, 2007.

[21] A. Mian, M. Bennamoun, and R. Owens, "An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 29, no. 11, pp. 1927-1943, Nov. 2007.

[22] T. Faltemier, K. Bowyer, and P. Flynn, "A Region Ensemble for 3-D Face Recognition," *IEEE Trans. Information Forensics and Security,* vol. 3, no. 1, pp. 62-73, Mar. 2008.

[23] T. Faltemier, K. Bowyer, and P. Flynn, "Rotated Profile Signatures for Robust 3D Feature Detection," *Proc. IEEE Eighth Int'l Conf. Automatic Face and Gesture Recognition,* pp. 1-7, Sept. 2008.

[24] H. Dibeklioğlu, "Part-Based 3D Face Recognition under Pose and Expression Variations," master's thesis, Boğaziçi Univ., 2008.

[25] H. Dibeklioğlu, A. Salah, and L. Akarun, "3D Facial Landmarking under Expression, Pose, and Occlusion Variations," *Proc. Second IEEE Int'l Conf. Biometrics: Theory, Applications and Systems,* pp. 1-6, Sept./Oct. 2008.

[26] T. Yu and Y. Moon, "A Novel Genetic Algorithm for 3D Facial Landmark Localization," *Proc. Second IEEE Int'l Conf. Biometrics: Theory, Applications, and Systems,* Sept./Oct. 2008.

[27] M. Romero-Huertas and N. Pears, "3D Facial Landmark Localization by Matching Simple Descriptors," *Proc. Second IEEE Int'l Conf. Biometrics: Theory, Applications, and Systems,* Sept./Oct. 2008.

[28] P. Nair and A. Cavallaro, "3-D Face Detection, Landmark Localization, and Registration Using a Point Distribution Model," *IEEE Trans. Multimedia,* vol. 11, no. 4, pp. 611-623, June 2009.

[29] L. Yin, X. Wei, Y. Sun, J. Wang, and M. Rosato, "A 3D Facial Expression Database for Facial Behavior Research," *Proc. Seventh Int'l Conf. Automatic Face and Gesture Recognition,* pp. 211-216, Apr. 2006.

[30] I. Dryden and K. Mardia, *Statistical Shape Analysis.* Wiley, 1998.

[31] M. Stegman and D. Gomez, "A Brief Introduction to Statistical Shape Analysis," technical report, Technical Univ. of Denmark, Mar. 2002.

[32] T. Cootes and C. Taylor, "Statistical Models of Appearance for Computer Vision," technical report, Univ. of Manchester, Oct. 2001.

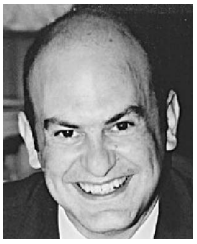
[33] T. Cootes, C. Taylor, H. Kang, and V. Petrovic, "Modeling Facial Shape and Appearance," *Handbook of Face Recognition,* pp. 39-63, Springer, 2005.

[34] S. Theodoridis and K. Koutroumbas, *Pattern Recognition,* third ed. Academic Press, 2006.

[35] J. Koenderink and A. van Doorn, "Surface Shape and Curvature Scales," *Image and Vision Computing,* vol. 10, pp. 557-565, Oct. 1992.

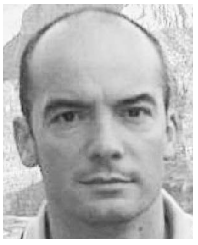
[36] A.E. Johnson, "Spin Images: A Representation for 3-D Surface Matching," PhD dissertation, Robotics Inst., Carnegie Mellon Univ., Pittsburgh, Penn., Aug. 1997.

[37] UH-CBL, "Facial Landmarks Annotation Files," http://www.cbl.uh.edu/URxD/annotations/facial-landmarks.zip, ver. 3, 2012.



Panagiotis Perakis received the BSc degree in physics in 1986 and the MSc degree in ICT in 2008 from the University of Athens, Greece, and is currently working toward the PhD degree in 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, pattern recognition, and physics-based modeling. Since 1993, he has been a co-owner of a Greek software development company. He is a member of the IEEE Computer Society.

Georgios Passalis received the BSc degree from the Department of Informatics and Telecommunications, University of Athens, in 2003, the MSc degree from the Department of Computer Science, University of Houston, in 2004, and the PhD degree from the University of Athens in 2008, with focus on the domains of computer graphics and computer vision. He has previously worked for the video game industry and is now working for a consulting firm.



Theoharis Theoharis received the DPhil degree in computer graphics and parallel processing from the University of Oxford, United Kingdom, in 1988. He subsequently served as a research fellow at the University of Cambridge, a professor at the University of Athens, and NTNU, Norway. His primary research interests include the fields of biometrics, 3D object retrieval, and reconstruction. He is the author of a number of textbooks, including *Graphics and Visualization: Principles and Algorithms*.

Ioannis A. Kakadiaris received the BSc degree in physics from the University of Athens, Greece, the MSc degree in computer science from Northeastern University, and the PhD degree from the University of Pennsylvania. He is the Hugh Roy and Lillie Cranz Cullen University Professor of Computer Science, Electrical and Computer Engineering, and Biomedical Engineering at the University of Houston (UH). He joined UH in August 1997 after a postdoctoral fellowship 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) (the position rotates annually among the institutions). His research interests include biometrics, computer vision, and biomedical image analysis. He is the recipient of a number of awards, including the US National Science Foundation's (NSF) Early Career Development Award, Schlumberger Technical Foundation Award, UH Computer Science Research Excellence Award, UH 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. He is a senior member of the IEEE.