CGL Technical Report, No. TP-2010-01, January 2010

3D Facial Landmark Detection & Face Registration
A 3D Facial Landmark Model & 3D Local Shape Descriptors Approach

Panagiotis Perakis[1,2], **Georgios Passalis**[1,2], **Theoharis Theoharis**[1,2] **and Ioannis A. Kakadiaris**[2]
[1] Computer Graphics Laboratory
Department of Informatics and Telecommunications
University of Athens, Ilisia 15784, GREECE
[2] Computational Biomedicine Lab
Department of Computer Science
University of Houston, Texas 77204, USA

Abstract In this Technical Report a novel method for 3D landmark detection and pose estimation suitable for both frontal and side 3D facial scans is presented. It utilizes 3D information by using 3D local shape descriptors to extract candidate interest points that are subsequently identified and labeled as anatomical landmarks. The shape descriptors include the *shape index*, a continuous map of principal curvature values of 3D objects, the *extrusion map*, a measure of the extruded areas of a 3D object and the *spin images*, local descriptors of the object's 3D point distribution. However, feature detection methods which use general purpose shape descriptors cannot identify and label the detected candidate landmarks. Therefore, the topological properties of the human face need to be taken into consideration. To this end, we use a *Facial Landmark Model* (FLM) of facial anatomical landmarks. Candidate landmarks, irrespectively of the way they are generated, can be identified and labeled by matching them with the corresponding FLM. The proposed method is evaluated using an extensive 3D facial database, and achieves high accuracy even in challenging scenarios.

Keywords Shape Models, Shape Index, Extruded Points, Spin Images, 3D Feature Extraction, Landmark Detection, Pose Estimation.

Version 1.0 Contact info: P. Perakis (`perakis@di.uoa.gr`.)

Computer Graphics Laboratory
Department of Informatics and Telecommunications
University of Athens
15784 Ilisia
GREECE
`http://graphics.di.uoa.gr`

Contents

1 Introduction

In a wide variety of disciplines it is of great practical importance to measure, describe and compare the shapes of objects. In computer graphics, computer vision and biometric applications, the class of objects is often the human face. Registration of facial scan data with a face model is important in face recognition, facial shape analysis, segmentation and labeling of facial parts, facial region retrieval, partial face matching, face mesh reconstruction, face texturing and relighting, face synthesis, and face motion capture and animation.

In recent years, as scanning methods have become more accessible due to lower cost and greater flexibility, 3D facial datasets are more easily available. In almost any application, requiring processing of 3D facial data, an initial registration step is necessary. Therefore, registration based on feature points (landmarks) correspondence is the most crucial step in order to make a system fully automatic. At the same time, the landmark detection algorithm must be pose invariant in order to allow the registration of both frontal and side facial scans.

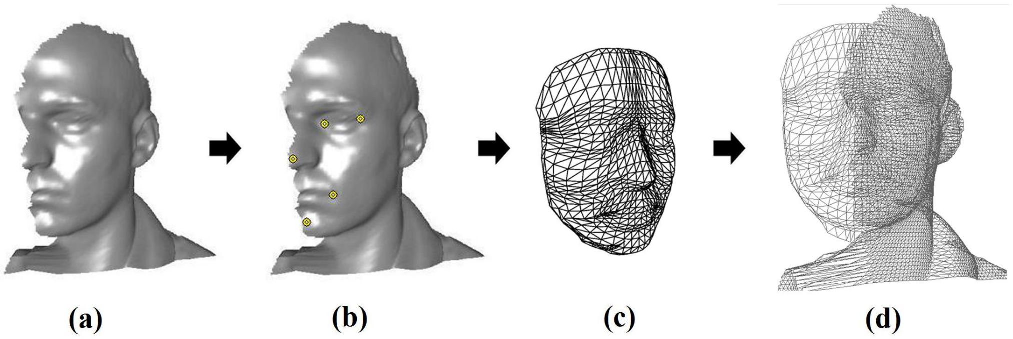


Figure 1: Face registration based on detected landmarks using the proposed method: (a) facial scan with extensive missing data; (b) extracted landmarks; (c) generic Reference Face Model (RFM); and (d) registered facial scan with RFM.

Existing 3D feature detection and localization methods, although they claim pose invariance, fail to address large pose variations and to confront the problem of missing facial areas in an holistic way (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 face meshes. *Side scans usually have large missing areas, due to self-occlusion, that depend on pose variations.* These scans are very common in realistic scenarios such as uncooperative subjects or uncontrolled environments.

The goal of the proposed method is to automatically and pose-invariantly detect landmarks (eye and mouth corners, nose and chin tips) in any 3D facial scan, and hence consistently register any pair of facial datasets. The main contribution of our proposed method is its applicability to large pose variations (up to 80° of yaw rotation), that often result in missing facial data, in an holistic way with high success rates.

At the training phase, our method creates a Facial Landmark Model (FLM) by first aligning the training landmark sets and calculating a mean landmark shape using Procrustes Analysis, and then applying Principal Component Analysis (PCA) to capture the shape variations. At the detection phase, the algorithm first detects candidate landmarks on the queried facial datasets exploiting the 3D geometry-based information. The extracted candidate landmarks are then filtered out and labeled by matching them with the FLM. Registration is then performed, based on resulting landmarks, with a generic Reference Face Model (RFM) (Fig. 1).

Evaluation of the proposed method is performed by computing the distance between manually annotated landmarks (ground truth) and the automatically detected landmarks. The experiments have been carried out on a combination of the largest publicly available databases: FRGC v2 [PFS*05] and UND Ear Database [UND08]. The first is a database for 3D face recognition that contains frontal facial scans, while the second is a database for 3D ear recognition that contains up to 80° side facial scans (both left and right).

In previous work, we have presented methods for detecting landmarks on 3D facial scans. In [PTPK09] shape index and spin images were introduced to locate landmarks in a manner that allows consistent retrieval of facial regions from 3D facial datasets. In [PPT*09] shape index and extrusion maps were introduced to locate landmarks for registering partial facial datasets in a face recognition system.

This report describes in detail and extends our previous methods by utilizing statistically trained spin image templates and alternative similarity distance measures. It achieves significantly higher landmark detection rates, which result in a far more robust face registration. It also contains exhaustive comparative analytical results for landmark localization success rates for all of our methods and other existing methods (Tables 8 and 9).

The rest of this report is organized as follows: Section 2 describes related work in the field, Sections 3 and 4 present the proposed method in detail, Section 5 presents our results, while Section 6 summarizes our method and proposes future directions.

2 Related Work

Facial feature detectors can be distinguished into two main categories: detection of feature points (landmarks) from the geometric characteristics of 2D intensity or color images and detection of feature points (landmarks) from the geometric information of 3D objects or 2.5D scans. Facial feature detectors can also be classified as those that are solely dependent on geometric information or those that are supported by trained statistical feature models. Three-D facial feature extraction has aroused interest with the increasing development of 3D modeling and digitizing techniques and is reported in a number of publications.

Lu, Colbry, Stockman and Jain [LJ05, CSJ05, LJ06, LJC06, Col06], in a series of publications, presented methods to locate the positions of eye and mouth corners, and nose and chin tips, based on a fusion scheme of shape index [DJ97] on range maps and the “cornerness” response [HS88] 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. The 3D feature extraction method presented in [CSJ05] addresses the problem of pose variations in a unified manner, and is tested against a composite database consisting of 953 scans from the FRGC database and 160 scan from a proprietary database with frontal scans extended with variations of pose, expressions, occlusions and noise. Their multimodal algorithm [LJ05] uses 3D+2D information and is applicable to almost-frontal scans ($< 5^\circ$ yaw rotation). It is tested against the FRGC database with 946 near frontal scans. The 3D feature extraction method presented in [LJ06] also addresses the problem of pose variations, and is tested against the FRGC database with 953 near frontal scans along with their proprietary MSU database consisting of 300 multiview scans ($0^\circ, \pm 45^\circ$) from 100 subjects. Results of the methods [LJ05, LJ06, Col06] are presented in Table 8, and of the method [LJ06] in Table 9, for comparison.

Conde *et al.* [CCRA*05] introduced a global face registration method by combining clustering techniques over discrete curvature and spin images for the detection of eye inner corners and nose tip. The method was tested on a proprietary database of 51 subjects with 14 captures each (714 scans). Their database consists of scans with small pose variations ($< 15^\circ$ yaw rotation). Although they presented a feature localization success rate of 99.66% on frontal scans and 96.08% on side scans, they do not define what a successful localization is.

Xu *et al.* [XTWQ06] presented a feature extraction hierarchical scheme to detect the positions of nose tip and nose ridge. They introduced the “effective energy” notion to describe the local distribution of neighboring points and detect the candidate nose tips. Finally, an SVM classifier is used to select the correct nose tips. Although it was tested against various databases, no exact localization results were provided.

Lin *et al.* [LSCH06] 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 used in an automatic 3D face authentication system, but was tested on only 27 human faces with various poses and expressions.

Segundo *et al.* [SQBS07] introduced a face and facial feature detection method by combining a method for 2D face segmentation on depth images with surface curvature information, in order to detect the eye corners, nose tip, nose base, and nose corners. The method was tested on the FRGC v2 database. Although they claim over 99.7% correct detections, they do not define a correct detection. Additionally, nose and eye corner detection presented problems when the face had a significant pose variation ($> 15^\circ$ yaw and roll).

Wei *et al.* [WLY07] introduced a nose tip and nose bridge localization method to determine facial pose. The method was based on a Surface Normal Difference algorithm and shape index estimation, and was used as a preprocessing step in pose-variant systems to determine the pose of the face. They reported an angular error of the nose tip - nose bridge segment less than 15° in 98% of the 2500 datasets of BU-3DFE facial database, which contains complete frontal facial datasets with capture range $\pm 45^\circ$.

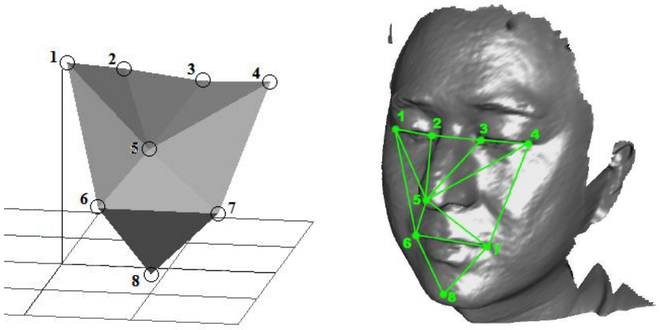
Mian *et al.* [MBO07] introduced a heuristic method for nose tip detection. The method is based on a geometric analysis of the nose ridge contour projected on the $x - y$ plane. It is used as a preprocessing step to cut out and pose correct the facial data in a face recognition system. However, no clear localization error results were presented. Additionally, their nose tip detection algorithm has limited applicability to near frontal scans ($< 15°$ yaw and pitch).

Faltemier *et al.* [FBF08a] introduced a heuristic method for nose tip detection. The method is a fusion of curvature and shape index analysis and a template matching algorithm using ICP. The nose tip detector had a localization error less than 10 *mm* in 98.2% of the 4007 facial datasets of FRGC v2 where it was tested. However, no exact localization distance error results were presented. They also introduced a method called "Rotated Profile Signatures" [FBF08b], based on profile analysis, to robustly locate the nose tip in the presence of pose, expression and occlusion variations. Their method was tested against NDOff2007 database which contains 7,317 facial scans, 406 frontal and 6,911 in various yaw and pitch angles. They reported a 96% to 100% success rate, with distance error threshold 10 *mm*, under significant yaw and pitch variations. Although their method achieved high success rate scores, it is a 2D-assisted 3D method since it uses skin segmentation to eliminate outliers, and is limited to the detection of the nose tip only. Finally, no exact localization distance error results were presented.

Dibeklioğlu, Salah and Akarun [Dib08, DSA08] 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 taken into consideration. They also introduced a nose tip localization and segmentation method using curvature-based heuristic analysis. However, the proposed system shows limited capabilities on 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. Finally, no exact localization distance error results were presented.

Yu and Moon [YM08] presented a nose tip and eye inner corners detection method on 3D range maps. The landmark detector is trained from example facial data using a genetic algorithm. The method was applied on 200 almost-frontal scans from FRGC v1 database. However, a limitation of the proposed system is that it is not applicable to facial datasets with large yaw rotations since it always uses the three aforementioned control points. Results of the method are presented in Table 8 for comparison reasons.

Romero-Huertas and Pears [RHP08] 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. Finally, after the graph matching algorithm has eliminated false candidates, the best combination of landmark points is selected from the minimum Mahalanobis distance to the trained landmark graph model. The method was tested against FRGC v1 (509 scans) and FRGC v2 (3271 scans) databases. They reported a success rate of 90% with thresholds for the nose tip at 15 *mm*, and for the inner eye corners at 12 *mm*.



(a) (b)

Figure 2: Depiction of: (a) landmark model as a 3D object; and (b) landmark model overlaid on a 3D facial dataset.

Nair and Cavallaro [NC09] 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 (nose tip and inner and outer eye corners). A statistical shape model (PDM) of feature points is fitted to the facial dataset by using three control 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 tested against the BU-3DFE facial database, which only contains complete frontal facial datasets with capture range ±45°. Furthermore, their method is not applicable to missing data resulting from pose self-occlusion, since it always uses the aforementioned three control points for model fitting. Results of the method are presented in Table 8 for comparison purposes.

Finally, Perakis *et al.* [PTPK09, PPT*09] presented methods for detecting facial landmarks (eye inner and outer corners, mouth corners, and nose and chin tips) on 2.5D scans. Local shape and curvature analysis utilizing shape index, extrusion maps and spin images were used to locate candidate landmark points. These are identified and labeled by matching them with a statistical facial landmark model. The method addresses the problem of extreme yaw rotations and missing facial areas, and it is tested against FRGC v2 and UND Ear databases.

3 3D Facial Landmark Models

Our method for 3D landmark detection and pose estimation uses 3D information to extract candidate interest points which are identified and labeled as anatomical landmarks by matching them with a Facial Landmark Model (FLM) [PTPK09, PPT*09].

We use a set of 8 anatomical landmarks: right eye outer corner (1), right eye inner corner (2), left eye inner corner (3), left eye outer corner (4), nose tip (5), mouth right corner (6), mouth left corner (7) and chin tip (8) (Fig. 2). Notice that 5 of these points are visible on profile and semi-profile face scans. So the complete set of 8 landmarks can be used for frontal and almost-frontal faces and two reduced sets of 5 landmarks (right and left) for semi-profile and profile faces. The right side landmark set contains the points (1), (2), (5), (6), and (8), and the left side the points (3), (4), (5), (7) and (8).

Each of these sets of landmarks constitute a corresponding 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 main steps to create the FLMs are:

- A statistical mean shape for each landmark set (FLM8, FLM5L and FLM5R) is calculated from a manually annotated training set using Procrustes Analysis. One hundred and fifty frontal face scans with neutral expressions are randomly chosen from the FRGC v2 database as our training examples.
- Variations of each Facial Landmark Model are calculated using Principal Component Analysis (PCA).

3.1 The Landmark Mean Shape

According to Dryden and Mardia [DM98], "a *landmark* is a point of correspondence on each object that matches between and within populations of the same class of objects" and "a *shape* is all the geometrical information that remains when location, scale and rotational effects are filtered out from an object". Shape, in other words, is invariant to Euclidean similarity transformations.

Since, for our purposes, the size of the shape is of great importance, it is not filtered out by scaling shapes to unit size. So, "two objects have the same *size-and-shape* if they are rigid-body transformations of each other" [DM98].

One way to describe a shape is by locating a finite number of landmarks on the outline or other specific points. Dryden and Mardia [DM98] sort landmarks into the following categories:

Anatomical landmarks: *Points assigned by an expert that correspond between organisms in some biologically meaningful way* (e.g., the corner of an eye).

Mathematical landmarks: *Points located on an object according to some mathematical or geometrical property* (e.g., a high curvature or an extremum point).

Pseudo-landmarks: *Constructed points on an object either on the outline or between anatomical or mathematical landmarks.*

Labeled landmarks: *Landmarks that are associated with a label (name or number), which is used to identify the corresponding landmark.*

Synonyms for landmarks include homologous points, interest points, nodes, vertices, anchor points, fiducial markers, model points, markers, key points, etc.

A mathematical representation of an n-point shape in d dimensions can be defined by concatenating all point coordinates into a $k = n \times d$ vector and establishing a *Shape Space* [DM98, SG02, CT01]. The *vector representation* for 3D shapes (i.e., $d = 3$) would then be:

$$\mathbf{x} = [x_1, x_2, ..., x_n, y_1, y_2, ..., y_n, z_1, z_2, ..., z_n]^T \tag{1}$$

where (x_i, y_i, z_i) represent the n landmark points.

To obtain a true representation of landmark shapes, location and rotational effects need to be filtered out. This is carried out by establishing a common coordinate reference to which all shapes are aligned.

Alignment is performed by minimizing the *Procrustes distance*

$$D_P^2 = |\mathbf{x_i} - \mathbf{x_m}|^2 = \sum_{j=1}^{k} (x_{ij} - x_{mj})^2 \tag{2}$$

of each shape $\mathbf{x_i}$ to the mean shape $\mathbf{x_m}$.

The alignment procedure is commonly known as *Procrustes Analysis* [DM98, SG02, CT01] and is used to calculate the mean shape of landmark shapes. Although there are analytic solutions, a typical iterative approach,
adapted from [CT01], is the following:

Algorithm 1: Procrustes Analysis

```
• Compute the centroid of each example shape.
• Translate each example shape so that its centroid is at the origin (0,0,0).
• Scale each example shape so that its size is 1.
• Assign the first example shape to the mean shape $\mathbf{x_m}$.
• REPEAT
    – Assign the mean shape $\mathbf{x_m}$ to a reference mean shape $\mathbf{x_0}$.
    – Align all example shapes to the reference mean shape $\mathbf{x_0}$ by an optimal
      rotation.
    – Recalculate the mean shape $\mathbf{x_m}$.
    – Translate the mean shape so that its centroid is at the origin (0,0,0).
    – Scale the mean shape so that its its size is 1.
    – Align the mean shape $\mathbf{x_m}$ to the reference mean shape $\mathbf{x_0}$ by an optimal
      rotation.
    – Compute the Procrustes distance of the mean shape $\mathbf{x_m}$ to the reference
      mean shape $\mathbf{x_0}$:
      $|\mathbf{x_0} - \mathbf{x_m}|$.
• UNTIL Convergence: $|\mathbf{x_0} - \mathbf{x_m}| < \varepsilon$.
```

In our case, where the size of the facial landmark shape is of great importance, scaling shapes to unit size is omitted. In these cases, shapes are considered rigid shapes and are aligned by performing only the translational and rotational transformations.

Thus the mean shape of landmark shapes (Fig. 3) is created and example shapes are aligned to the mean shape.

The mean shape $\mathbf{x_m}$ is the *Procrustes mean*

$$\mathbf{x_m} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x_i} \tag{3}$$

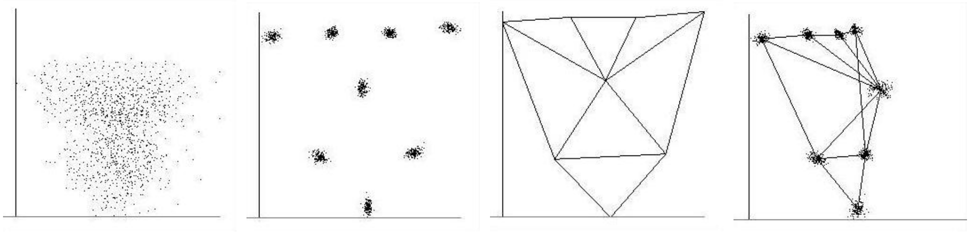

(a) (b) (c) (d)

Figure 3: Depiction of landmarks mean shape estimation: (a) unaligned landmarks; (b) aligned landmarks; (c) landmarks mean shape; and (d) landmark cloud and mean shape at 60^o.

of all N example shapes $\mathbf{x_i}$.

3.2 Shape Alignment Transformations

As we have previously mentioned, to obtain a true representation of landmark shapes, location, scale and rotational effects need to be filtered out by bringing shapes to a common frame of reference. This is carried out by performing translational, scaling and rotational transformations. Notice that different approaches to alignment can produce different distributions of the aligned shapes.

Translation to the centroid is performed by applying to the n landmark points $\mathbf{r_j}$ the following transformation in 3D original space:

$$\mathbf{r_j'} = \mathbf{r_j} - \mathbf{r_c} \tag{4}$$

where $\mathbf{r_c}$ the centroid and $j \in \{1, ..., n\}$.

The *centroid* of a shape is the center of mass (CM) of the physical system consisting of unit masses at each landmark. This is easily calculated as:

$$\mathbf{r_c} = \left[\frac{1}{n} \sum_{j=1}^{n} x_j, \frac{1}{n} \sum_{j=1}^{n} y_j, \frac{1}{n} \sum_{j=1}^{n} z_j \right]^T \tag{5}$$

in a 3D original space (i.e., $d = 3$).

Scaling to unit size is performed by applying to the landmark points $\mathbf{r_j}$ the following transformation in 3D original space:

$$\mathbf{r_j'} = \alpha \mathbf{r_j} \tag{6}$$

where $\alpha = 1/S(\mathbf{x})$ is the scaling factor, $S(\mathbf{x})$ is the shape's size, and $j \in \{1, ..., n\}$.

The shape's size is the square root of the sum of squared Euclidean distances from each landmark $\mathbf{r_j}$ to the centroid $\mathbf{r_c}$:

$$S(\mathbf{x})^2 = \sum_{j=1}^{n} |\mathbf{r_j} - \mathbf{r_c}|^2 \tag{7}$$

in the original 3D space.

Rotation in the original 3D space is slightly more complicated. We must calculate a rotational transformation $R(\mathbf{x})$ so as to minimize the Procrustes distance $|R(\mathbf{x}) - \mathbf{x_0}|$ of the transformed shape $R(\mathbf{x})$ to a reference shape $\mathbf{x_0}$. The rotational transformation $\mathbf{R}$ can be expressed as a product of three rotations around the three principal axes:

$$\mathbf{R} = \mathbf{R}_{x,\theta} \cdot \mathbf{R}_{y,\phi} \cdot \mathbf{R}_{z,\psi} \tag{8}$$

These can be expressed in a matrix form:

$$\mathbf{R}_{x,\theta} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \tag{9}$$

$$\mathbf{R}_{y,\phi} = \begin{bmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{bmatrix} \tag{10}$$

$$\mathbf{R}_{z,\psi} = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{11}$$

After setting partial derivatives of $|R(\mathbf{x}) - \mathbf{x_0}|^2$ w.r.t each parameter to zero and some formal calculations, we have:

$$\theta = \tan^{-1}\left(\frac{S_{z0,y} - S_{y0,z}}{S_{y0,y} + S_{z0,z}}\right) \tag{12}$$

$$\phi = \tan^{-1}\left(\frac{S_{x0,z} - S_{z0,x}}{S_{z0,z} + S_{x0,x}}\right) \tag{13}$$

$$\psi = \tan^{-1}\left(\frac{S_{y0,x} - S_{x0,y}}{S_{x0,x} + S_{y0,y}}\right) \tag{14}$$

where:

$$S_{x0,x} = \textstyle\sum_{j=1}^{n} x_{0j}x_j, \quad S_{x0,y} = \sum_{j=1}^{n} x_{0j}y_j, \quad S_{x0,z} = \sum_{j=1}^{n} x_{0j}z_j,$$

$$S_{y0,x} = \textstyle\sum_{j=1}^{n} y_{0j}x_j, \quad S_{y0,y} = \sum_{j=1}^{n} y_{0j}y_j, \quad S_{y0,z} = \sum_{j=1}^{n} y_{0j}z_j,$$

$$S_{z0,x} = \textstyle\sum_{j=1}^{n} z_{0j}x_j, \quad S_{z0,y} = \sum_{j=1}^{n} z_{0j}y_j, \quad S_{z0,z} = \sum_{j=1}^{n} z_{0j}z_j.$$

So, the rotational transformation of every landmark point $\mathbf{r_j}$ in the original 3D space gives:

$$\mathbf{r_j}' = R(\mathbf{r_j}) = R_{x,\theta}\left(R_{y,\phi}\left(R_{z,\psi}(\mathbf{r_j})\right)\right) \tag{15}$$

Alignment of a shape $\mathbf{x}$ to a reference shape $\mathbf{x_0}$ is done by minimizing the Procrustes distance in an iterative way, as described below:

Algorithm 2: Shape Alignment

```
• Translate $\mathbf{x_0}$ so that its centroid is at the origin (0,0,0).
• Scale $\mathbf{x_0}$ so that its size is 1.
• Translate $\mathbf{x}$ so that its centroid is at the origin (0,0,0).
• Scale $\mathbf{x}$ so that its size is 1.
• Set $\mathbf{R} \leftarrow \mathbf{I}$.
• REPEAT
    – Calculate $\mathbf{R}_{x,\theta}$.
    – Apply $\mathbf{R}_{x,\theta}$ to $\mathbf{x}$ shape points.
    – Set $\mathbf{R} \leftarrow \mathbf{R}_{x,\theta} \cdot \mathbf{R}$.
    – Calculate $\mathbf{R}_{y,\phi}$.
    – Apply $\mathbf{R}_{y,\phi}$ to $\mathbf{x}$ shape points.
    – Set $\mathbf{R} \leftarrow \mathbf{R}_{y,\phi} \cdot \mathbf{R}$.
    – Calculate $\mathbf{R}_{z,\psi}$.
    – Apply $\mathbf{R}_{z,\psi}$ to $\mathbf{x}$ shape points.
    – Set $\mathbf{R} \leftarrow \mathbf{R}_{z,\psi} \cdot \mathbf{R}$.
    – Compute the Procrustes distance of the transformed shape $\mathbf{x}$ to the
      reference shape $\mathbf{x_0}$:
      $|\mathbf{x_0} - \mathbf{x}|$.
• UNTIL Convergence: $|\mathbf{x} - \mathbf{x_0}| < \varepsilon$.
• Get $\mathbf{R}$.
```

Note that, in our case, where the size of the facial landmark shape is of great importance, scaling shapes to unit size is omitted. Also note that the proposed "Shape Alignment" algorithm leaves us the discretion to permit certain rotations (e.g., only around the y-axis).

3.3 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 for describing the shape variations. This shape decomposition is performed by applying Principal Component Analysis (PCA) to the aligned shapes.

Due to size normalization of Procrustes analysis, all shape vectors live in a hyper sphere manifold in shape space, which introduces non-linearities if large shape scalings occur. Since PCA is a linear procedure, all aligned shapes are at first projected to the tangent space of the mean shape. This way, shape vectors lie in a hyper plane instead of a hyper sphere, and non-linearities are filtered out. The tangent space projection linearizes shapes by scaling them with a factor α:

$$\mathbf{x_t} = \alpha \mathbf{x} = \frac{|\mathbf{x_m}|^2}{\mathbf{x_m} \cdot \mathbf{x}} \mathbf{x} \tag{16}$$

where $\mathbf{x_t}$ is the tangent space projection of shape $\mathbf{x}$ and $\mathbf{x_m}$ is the mean shape.

If no size normalization is applied, then tangent space projection can be omitted.

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. If landmark points are not representing a certain class of shapes, then they will be totally uncorrelated (i.e., purely random). On the other hand, if landmark points represent a certain class of shapes, then they will be correlated to some degree. This fact will be exploited by applying PCA to reduce dimensionality and obtain the correlation as deformations.

If landmark points have a specific distribution, we can model this distribution by estimating a vector $\mathbf{b}$ of parameters that describes a shape's deformations [CTCG95, CT01, CTKP05, SG02]. The approach is as follows:

Algorithm 3: Principal Component Analysis

- Determine the mean shape.
- Determine the covariance matrix of the shape vectors.
- Compute the eigenvectors $\mathbf{A}_i$ and corresponding eigenvalues λ_i of the covariance matrix, sorted in descending order.

After applying Procrustes analysis, the mean shape is determined and example shapes are aligned and projected to the mean shape's tangent space. Typically, one would apply PCA on variables with zero mean.

The covariance matrix of N example shapes is calculated according to

$$\mathbf{C_x} = \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{x_i} - \mathbf{x_m})(\mathbf{x_i} - \mathbf{x_m})^T \tag{17}$$

If $\mathbf{A}$ contains (in columns) the $k = nd$ eigenvectors $\mathbf{A}_i$ of $\mathbf{C_x}$, by projecting aligned original example shapes to the eigenspace we uncorrelate them as

$$\mathbf{y} = \mathbf{A}^T \cdot (\mathbf{x} - \mathbf{x_m}) \tag{18}$$

and the covariance matrix of projected example shapes

$$\mathbf{C_y} = \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{y_i} - \mathbf{y_m})(\mathbf{y_i} - \mathbf{y_m})^T \tag{19}$$

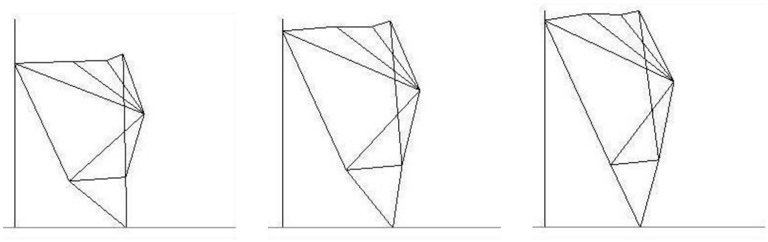
becomes a diagonal matrix of the eigenvalues λ_i, so as to have

$$\mathbf{C_x} \cdot \mathbf{A} = \mathbf{A} \cdot \mathbf{C_y} \quad , \quad \mathbf{C_y} = \mathbf{A}^T \cdot \mathbf{C_x} \cdot \mathbf{A} \tag{20}$$

The resulting transform is known as the *Karhunen-Loéve transform*, and achieves our original goal of creating mutually uncorrelated features.

To back-project uncorrelated shape vectors into the original shape space, we can use

$$\mathbf{x} = \mathbf{x_m} + \mathbf{A} \cdot \mathbf{y} \tag{21}$$



(a) $b_1 = -3\sqrt{\lambda_1}$ (b) $b_1 = 0$ (c) $b_1 = +3\sqrt{\lambda_1}$

Figure 4: First mode of mean shape deformations (viewed at 60°).

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

$$\mathbf{x}' \approx \mathbf{x_m} + \mathbf{A} \cdot \mathbf{b} \tag{22}$$

where $\mathbf{b}$ is a p-dimensional vector given by

$$\mathbf{b} = \mathbf{A}^T \cdot (\mathbf{x} - \mathbf{x_m}) \tag{23}$$

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. The number p of most significant 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, that is the sum V_t of all the eigenvalues

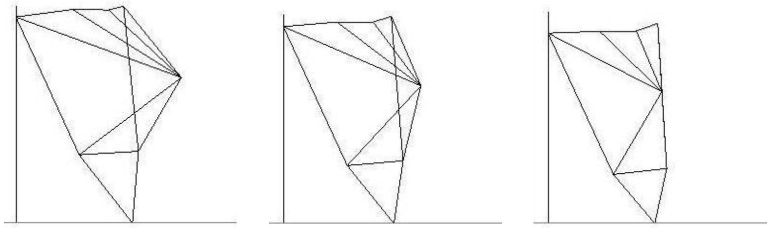
$$\sum_{i=1}^{p} \lambda_i \geq f \cdot V_t \tag{24}$$

where factor f represents the percentage of total variance incorporated into FLM. Least significant eigenvalues that are not incorporated are considered to represent noise [CT01, TK06]. Thus, the *Facial Landmark Model* (FLM) is created [PTPK09, PPT*09].

By applying PCA, we decompose shape variations by projecting to the eigenspace having an ordered basis of eigenvectors, where each shape component is ranked after the corresponding eigenvalue. This gives the components an order of significance. Each eigenvalue represents the variance in eigenspace axes which are orthogonal. Notice that the correlation matrix of shape vectors in the eigenspace has only diagonal elements: the eigenvalues.

Modifying one component at a time we obtain the *principal modes of variations.* So, for each selected eigenvalue λ_i, we calculate the deformation parameter b_i within some limits ($\pm 3\sqrt{\lambda_i} = \pm 3\sigma_i$), and we get a corresponding mode of variations, which represents $f_i = \frac{\lambda_i}{V_{tot}}$ of the total shape variations in the dataset.

We can observe that the first mode (Fig. 4), which is created by setting the deformation parameter values to ($b_1 = -3\sqrt{\lambda_1}$, $b_1 = 0$, $b_1 = +3\sqrt{\lambda_1}$), captures the

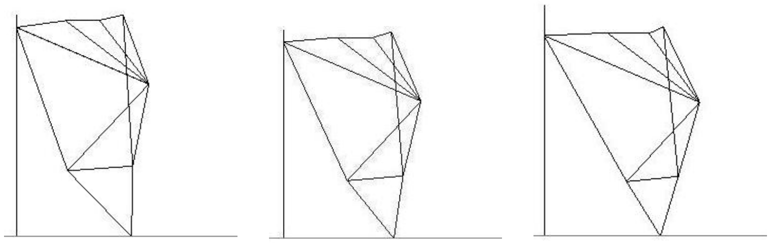


(a) $b_2 = -3\sqrt{\lambda_2}$

(b) $b_2 = 0$

(c) $b_2 = +3\sqrt{\lambda_2}$

Figure 5: Second mode of mean shape deformations (viewed at 60°).



(a) $b_3 = -3\sqrt{\lambda_3}$

(b) $b_3 = 0$

(c) $b_3 = +3\sqrt{\lambda_3}$

Figure 6: Third mode of mean shape deformations (viewed at 60°).

face size and shape (circular vs. oval) and represents 21.9% of total shape variations of FLM8 (Fig. 7).

We can also observe that the second mode (Fig. 5), which is created by setting the deformation parameter values to ($b_2 = -3\sqrt{\lambda_2}$, $b_2 = 0$, $b_2 = +3\sqrt{\lambda_2}$), captures the nose shape (flat vs. peaked) and represents 18.6% of total shape variations of FLM8 (Fig. 7).

Finally, we can observe that the third mode (Fig. 6, which is created by setting the deformation parameter values to ($b_3 = -3\sqrt{\lambda_3}$, $b_3 = 0$, $b_3 = +3\sqrt{\lambda_3}$), captures the chin tip position (extruded vs. intruded) and represents 11.1% of total shape variations of FLM8 (Fig. 7).

The first three principal modes of FLM8 capture 51.6% of the total shape variations. We incorporated 15 eigenvalues (out of the total 24) in FLM8, which represent 99.0% of total shape variations of the complete landmark shapes. We also incorporated 7 eigenvalues (out of the total 15) in FLM5L and FLM5R, which represent 99.0% of total shape variations of the left and right landmark shapes. By selecting the most significant eigenvalues and corresponding eigenvectors, each shape vector in the original shape space is projected to a feature vector in a *feature space* with reduced dimensions.

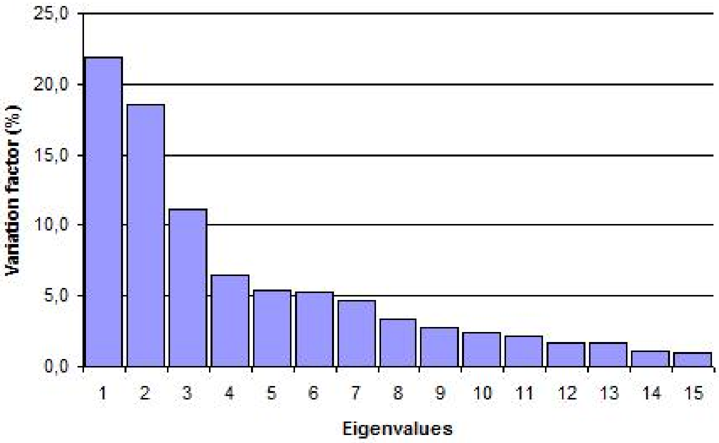


Figure 7: Landmark shape eigenvalues for FLM8 and percentage of total variations they capture.

3.4 Statistical Analysis of Landmarks

Point clouds of aligned landmarks represent landmark "movements" in 3D space. Looking at the correlation matrix we see that these "movements" are highly correlated (Fig. 8). Black squares denote negative correlation values, white, positive correlation values and mean gray, zero correlation. Note that shape vectors are presented in a $(x_1, x_2, ..., x_8, y_1, y_2, ..., y_8, z_1, z_2, ..., z_8)$ manner.

The main diagonal of the covariance matrix contains the variances of each shape vector component:

$$var(x_i) = \frac{1}{N-1} \sum_{k=1}^{N} (x_{k,i} - x_{m,i})^2 \tag{25}$$

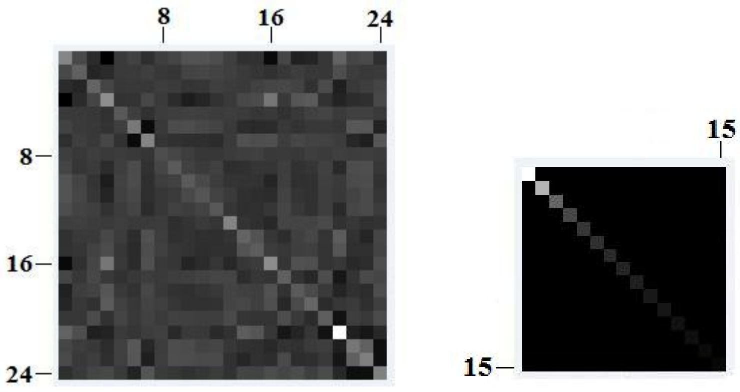
and non diagonal values the covariances between any two components:

$$covar(x_i, x_j) = \frac{1}{N-1} \sum_{k=1}^{N} (x_{k,i} - x_{m,i})(x_{k,j} - x_{m,j}) \tag{26}$$

where $\mathbf{x_m}$ is the mean shape, $\mathbf{x_k}$ any example shape and N the examples number. The covariance matrix is symmetrical about the main diagonal, since $covar(x_i, x_j) = covar(x_j, x_i)$.

The values of the covariance indicates the strength of each relationship, and the sign whether the relationship is positive or negative. If the value is positive, the two components increase together. If it is negative, then if one component increases in one direction the other increases in the opposite direction (decreases). Notice that division is done by $N-1$ since we are using an example dataset, which is a representation of the entire population where division by N can properly been used.

Consider the black square (1,4) in Fig. 8(a); it represents the correlation between (x_1, x_4), which are the x coordinates of left and right eye outer corners. We can conclude that they are negatively correlated; when right eye moves right, left eye moves



(a) (b)

Figure 8: Statistical analysis for FLM8: (a) Correlation Matrix; (b) Eigenvalues.

left and vice versa. This is also indicated in the first mode of variations - circular vs. oval face (Fig. 4). Consider the black square (6,7); it represents the correlation between (x_6, x_7), which are the x coordinates of mouth left and right corners. We can conclude that they are also negatively correlated; when mouth right corner moves right, left corner moves left and vice versa. Black squares (24,22) and (24,23) represent the correlation between (z_8, z_6) and (z_8, z_7), which are the z coordinates of chin tip versus mouth left and right corners. We can see a negative correlation, which means that chin tip and mouth corners "move" in opposite directions on z-axis. This is also indicated in the third mode of variations - extruded vs. intruded chin (Fig. 6). Consider the gray squares of line (5); they represent the correlation of x,y,z coordinates of the nose tip with the other landmarks. We can conclude that the nose is mostly not correlated with any other landmark, because of the same gray color of the corresponding squares. It is the most robust facial landmark point. White square (21,21) represents the variance of (z_5), which is the z coordinate of the nose tip. We can observe that this has the maximum variance, which is also indicated in the second mode of variations - flat vs. peaked nose (Fig. 5).

3.5 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 must be taken into consideration. To address this problem, we use the FLMs. Candidate landmarks, irrespectively of the way they are produced, must 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 landmark points $\mathbf{y}$ to the FLM $\mathbf{x}$ is done by minimizing the Procrustes distance in a simple iterative approach, adapted from [CT01]:

Algorithm 4: Landmark Fitting

```
• Translate $\mathbf{y}$ so that its centroid is at the origin (0,0,0).
• Scale $\mathbf{y}$ shape so that its size is 1.
• REPEAT
    – Align $\mathbf{y}$ to the mean shape $\mathbf{x_m}$ by an optimal rotation.
    – Compute the Procrustes distance of $\mathbf{y}$ to the mean shape $\mathbf{x_m}$: $|\mathbf{y}-\mathbf{x_m}|$.
• UNTIL Convergence: $|\mathbf{y}-\mathbf{x_m}| < \varepsilon$.
• Project $\mathbf{y}$ into tangent space of $\mathbf{x_m}$.
• Determine the model deformation parameters $\mathbf{b}$ that match to $\mathbf{y}$:
  $\mathbf{b} = \mathbf{A}^T \cdot (\mathbf{y}-\mathbf{x_m})$.
• Accept $\mathbf{y}$ as a member of the shape's class if $\mathbf{b}$ satisfies certain constraints.
```

Notice that scaling is not applied when we need to retain shape size.

We consider a landmark shape as plausible if it is consistent with marginal shape deformations. Let us say that certain b_i satisfy the deformation constraint

$$|b_i| \leq 3\sqrt{\lambda_i} \tag{27}$$

then the candidate landmark shape belongs to the shape class with probability $Pr(\mathbf{y})$:

$$Pr(\mathbf{y}) = \frac{\sum \lambda_i}{V_p} \tag{28}$$

where λ_i are the eigenvalues that satisfy the deformation constraints and V_p is the sum of the eigenvalues that correspond to the selected p principal components, and represents the incorporated data variance.

If $Pr(\mathbf{y})$ exceeds a certain threshold limit, the landmark shape is considered plausible, otherwise it is rejected as a member of the class. Other criteria of declaring a shape as plausible can also be applied [CT01, CTKP05].

4 Landmark Detection & Labeling

To detect landmark points, we have used three 3D local shape descriptors that exploit the 3D geometry-based information of facial datasets: shape index, extrusion map 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 w.r.t. a known coordinate system:

$$S(\mathbf{p}) = \{\mathbf{p} \in \mathbb{R}^3 : \mathbf{p} = [x(u,v), y(u,v), z(u,v)]^T, (u,v) \in \mathbb{R}^2\} \tag{29}$$

This global native u, v parameterization of the facial scan allows us to map 3D information into 2D space. Since differential geometry is used for describing local behavior of surfaces in a small neighborhood, such as surface curvature and surface normals, we assume that the surface S can be adequately modeled as being at least piecewise smooth, that is at least be of class C^2 (twice differentiable). Therefore, to eliminate sensor-specific problems, certain preprocessing algorithms (*median cut*, *hole filling*, *smoothing*, and *subsampling*) operate directly on the range data before the conversion to polygonal data [KPT*07].

4.1 Shape Index

The *Shape Index* is extensively used for 3D landmark detection [Col06, LJ06, LJC06, CSJ05, LJ05]. It is a continuous mapping of principal curvature values (k_{max}, k_{min}) of a 3D object point $\mathbf{p}$ into the interval [0,1], according to the formula:

$$SI(\mathbf{p}) = \frac{1}{2} - \frac{1}{\pi} tan^{-1} \frac{k_{max}(\mathbf{p}) + k_{min}(\mathbf{p})}{k_{max}(\mathbf{p}) - k_{min}(\mathbf{p})} \tag{30}$$

We use the Dorai and Jain definition here [DJ97], an extension of Koenderink and van Doorn's original definition [KvD92]. The shape index captures the intuitive notion of "local" shape of a surface. Every distinct surface shape corresponds to a unique value of shape index, except the planar shape. Points on a planar surface have an indeterminate shape index, since $k_{max} = k_{min} = 0$. Five well-known shape types and their locations on the shape index scale are as follows: Cup = 0.0, Rut = 0.25, Saddle = 0.5, Ridge = 0.75, and Cap = 1.0.

After calculating shape index values on a 3D facial dataset, a mapping to 2D space is performed (using the native u, v parameterization of the facial scan) in order to create a *shape index map* (Fig. 9):

$$SI_{map}(u, v) \leftarrow ShapeIndex(x, y, z) \tag{31}$$

Local maxima and minima are identified on the shape index map. Local maxima ($SI_{map}(u, v) \rightarrow 1.0$) are candidate landmarks for nose tips and chin tips and local minima ($SI_{map}(u, v) \rightarrow 0.0$) for eye corners and mouth corners. The shape index's maxima and minima that are located 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 (a maximum of 512 Caps and 512 Cups). In Fig. 15(a) and Fig. 16(a), black boxes represent Caps, and white boxes Cups.

However, experimentation showed that the shape index alone is not sufficiently robust for detecting anatomical landmarks in facial datasets in a variety of poses. Thus, candidate landmarks estimated from shape index values serve as a basis, but must be further classified and filtered according to the following methods.

4.2 Extruded points

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 use the *radial map* (Fig. 10(a)).

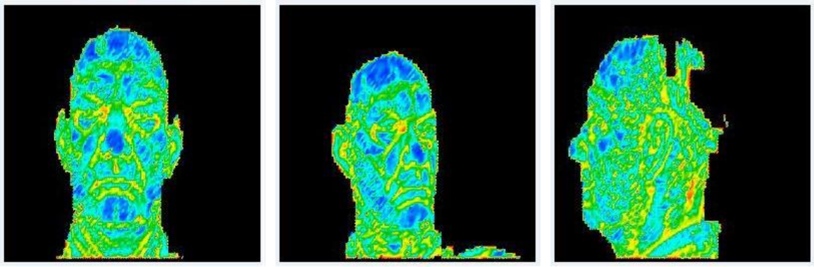


Figure 9: Depiction of shape index maps: (a) frontal face dataset; (b) 45° side face dataset; and (c) 60° side face dataset. (Blue denotes Caps, green Saddle, and red Cups.)

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, normalized to $[0, 1]$:

$$R_{map}(u, v) \leftarrow |\mathbf{r}(x, y, z)| \tag{32}$$

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. 10(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:

$$T_{map}(u, v) \leftarrow cos(\mathbf{r}(x, y, z), \mathbf{n}(x, y, z)) \tag{33}$$

Their product constitutes the *extrusion map* that represents the conjunction of the above two attributes, and is subsequently normalized to $[0, 1]$ (Fig. 10(c)):

$$E_{map}(u, v) = R_{map}(u, v) \times T_{map}(u, v) \tag{34}$$

Since the extrusion map depends only on the position of the centroid, it can be considered pose invariant.

Local maxima of the extrusion map ($E_{map}(u, v) \rightarrow 1.0$) that are also shape index maxima ($SI_{map}(u, v) \rightarrow 1.0$) 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 extruded points is retained (a maximum of 64 extruded points for nose and chin tips).

By using the extrusion map, the number of candidate landmarks for nose and chin tips resulting from shape index's values alone are significantly decreased, and are more robustly localized. We can retain the shape index's minima as candidate landmarks for eye and mouth corners (Fig. 15(a)) and extrusion map maxima as candidate landmarks for the nose and chin tips (Fig. 15(b)). In Fig. 15(b), simple crosses represent extrusion map maxima and circled crosses represent extrusion map maxima that are also shape index's maxima: candidate nose and chin tips.

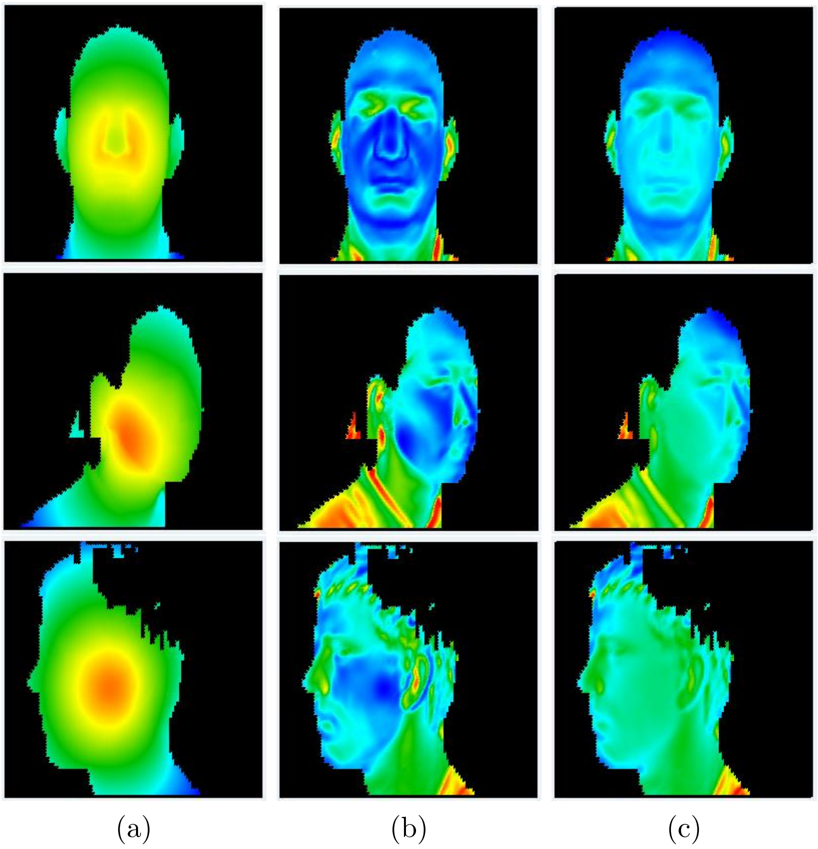


Figure 10: Depiction of extruded points: (a) radial map; (b) tangent map; and (c) extrusion map. (Blue denotes high values, and red low values.)

4.3 Spin Images

A *Spin Image* encodes the coordinates of points on the surface of a 3D object with respect to a local basis, a so-called *oriented point* [Joh97]. An oriented point is the pair $(\mathbf{p}, \mathbf{n})$, where $\mathbf{n}$ is the normal vector at a point $\mathbf{p}$ of a 3D object. A spin image is a local descriptor of the global or local shape of the object, invariant under rigid transformations.

The spin image generation process can be visualized as a grid of bins spinning around the oriented point basis, accumulating points at each bin as it sweeps space. Therefore, 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°.

Locality is expressed with the *Support Distance* parameter, which is:

$$\begin{aligned}(SupportDistance) &= (GridRows) \times (BinSize) \\ &= (GridColumns) \times (BinSize)\end{aligned}$$

A spin image at $(\mathbf{p}, \mathbf{n})$ is a signature of the shape of an object at the neighborhood of $\mathbf{p}$. For our purposes of representing facial features on 3D facial datasets, a 16×16 spin image grid with 2 *mm* bin size was used. This represents the local shape spanned by a cylinder of 3.2 *cm* height and 3.2 *cm* radius.

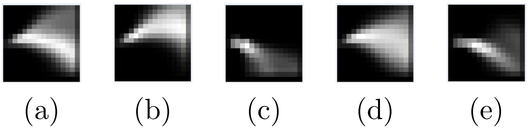


Figure 11: Spin Image templates: (a) eye outer corner; (b) eye inner corner; (c) nose tip; (d) mouth corner; (e) chin tip.

In order to identify interest points on 3D facial datasets by using spin images, we create spin image templates that represent the classes of the landmarks used. Notice that due to the symmetry of the face, landmark points cannot be distinguished according to spin images into left and right. Thus, five classes can be created which represent the eye outer corner, eye inner corner, nose tip, mouth corner and chin tip landmarks.

Spin image templates are statistically generated from 975 manually annotated frontal face scans from FRGC v2 database, and represent the mean spin image grid associated with the five classes of the used landmarks (Fig. 11).

Landmark points can be identified according to the relevance of their spin image grids with the five spin image templates that represent each landmark class.

Relevance is estimated according to a similarity measure between two spin image grids P and Q, which is expressed by the normalized linear correlation coefficient:

$$S(P,Q) = \frac{N\sum p_i q_i - \sum p_i \sum q_i}{\sqrt{\left[N\sum p_i^2 - (\sum p_i)^2\right]\left[N\sum q_i^2 - (\sum q_i)^2\right]}} \tag{35}$$

where p_i, q_i denotes each of the N elements of spin image grids P and Q, respectively [Joh97].

Figure 12 depicts the *spin image similarity maps* of facial datasets for each spin image template (i.e., landmark class). It is a u,v mapping of the $S(P,Q)$ value between the spin image P of every facial dataset point and a spin image template $Q(T)$:

$$SS^T_{map}(u,v) \leftarrow S(P(x,y,z),Q(T)) \tag{36}$$

Notice that areas of red color in Fig. 12 give an insight into the discriminating power of each spin image template. Spin image templates for eye inner corner and nose tip have the most discriminating power, since high similarity areas are located at the proper face regions, although the nose tip template has some similarity with eyebrows and chin regions. Spin image templates for eye outer corner and chin tip have a medium discriminating power, since there is high similarity with other regions of face. The eye outer corner template has similarity with mouth and cheek regions and the chin tip template with nose and eyebrows regions. Finally, the spin image template for mouth corner has the lowest discriminating power, since there is high similarity with large regions of the face, such as cheeks and forehead. These error-prone regions can be filtered out by the use of shape index's 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, local maxima and minima of the shape index map (Caps

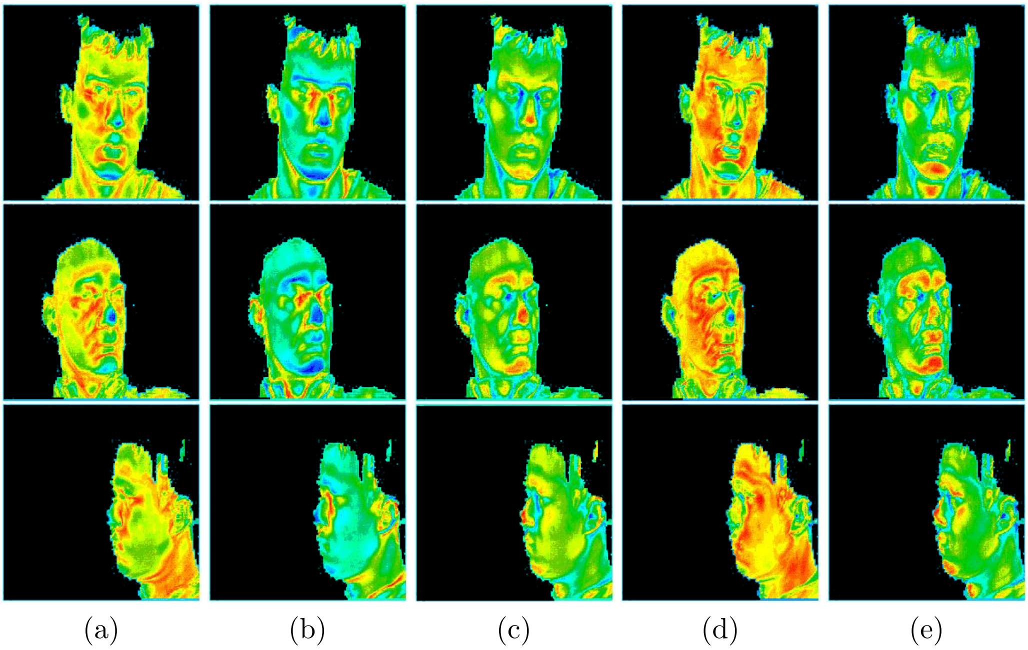


Figure 12: Depiction of spin image similarity maps: (a) eye outer corner; (b) eye inner corner; (c) nose tip; (d) mouth corner; and (e) chin tip. (Blue denotes low similarity values (-1), and red high similarity values ($+1$).)

and Cups) are further classified into five classes (eye outer corner, eye inner corner, nose tip, mouth corner and chin tip) according to the similarity ($SS^T_{map}(u, v) \rightarrow 1.0$) of their spin image grids with the spin image templates that represent each landmark class.

The landmarks of the five landmark classes (eye outer corner, eye inner corner, nose tip, mouth corner and chin tip) are sorted in descending order of significance according to their similarity measure with their corresponding spin image template. The most significant subset for each landmark class is retained (a maximum of 128 eye outer corners, 64 eye inner corners, 64 nose tips, 256 mouth corners and 128 chin tips). By using the spin images, the number of all candidate landmarks resulting from shape index's values are significantly decreased, and are more robustly localized.

In Fig. 16(b), blue boxes represent the eye outer corner, red boxes the eye inner corner, green boxes the nose tip, purple boxes the mouth corner and yellow boxes the chin tip. Notice that some of the classified landmark boxes overlap due to similarity with different templates.

4.4 Locating Landmarks on 2D Maps

In order to locate the most significant landmark points on a 2D map, we use general methods of locating extreme values. First, all 2D maps are normalized by linear stretching to [0,1] so that the problem of locating maximum or minimum is reduced to locating a single value (i.e., 1 or 0). Then, if a 2D map is represented by its values $I(u,v)$ and a target value V is searched within it, we can consider the function $|I(u,v) - V|$ as a transformation of the 2D map and search for its minimum values.

The localization of target values on a 2D map can be implemented by the algorithm below:

Algorithm 5: Landmark Localization

```
• FOR each point $(u,v)$.
    – Calculate $|I(u,v) - V|$.
    – IF $|I(u,v) - V| > Var$ reject point $(u,v)$.
    – IF $|I(u,v) - V|$ is not a minimum in a window of neighbors reject point $(u,v)$.
    – IF $|I(u,v) - V|$ is not a majority value in a window of neighbors reject point $(u,v)$.
    – IF NOT rejected add point in a descending ordered list of points according to $|I(u,v) - V|$.
• END FOR.
• Return list of points.
```

The algorithm calculates the value $|I(u,v) - V|$ and tests if it is within certain accepted variation limits Var in order to reject unwanted values (outliers). Then it tests if $|I(u,v) - V|$ is a local minimum within a window of neighbors by suppressing non minimum candidate points (hill climbing scheme). Finally, it tests wether the target value is a majority value (within some limits $|I(u,v) - V| \leq Var$) in a window of neighbors (voting scheme). Thus a list of points is returned, sorted in descending order of significance, according to the distance from target value $|I(u,v) - V|$.

4.5 Landmark Labeling & Selection

As we have previously mentioned, detected geometric landmarks must be identified and labeled as anatomical landmarks. For this purpose, topological properties of faces must be taken into consideration. Thus, candidate geometric landmarks, irrespectively of the way they are produced, must be consistent with the FLMs. This is done by applying the fitting procedure as described in Section 3.5.

For each facial dataset, the procedure for landmark detection and labeling has the following steps:

1. Extract candidate landmarks from the geometric properties of the facial scans.
2. Create feasible combinations of 5 landmarks from the candidate landmark points.
3. Compute the rigid transformation that best aligns the combinations of five candidate landmarks with the FLM5R and FLM5L.
4. Filter out those combinations that are not consistent with FLM5L or FLM5R, by applying the fitting procedure as previously described.
5. Sort consistent right (FLM5R) and left (FLM5L) 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 landmark sets of 8 landmarks.
7. Compute the rigid transformation that best aligns the combinations of eight landmarks with the FLM8.
8. Discard combinations of landmarks that are not consistent with the FLM8, by applying the fitting procedure as previously described.
9. Sort consistent complete landmark sets in descending order according to a distance metric from the FLM8.
10. Select the best combination of landmarks (consistent with FLM5R, FLM5R or FLM8) based on the distance metric to the corresponding FLM.
11. Obtain the corresponding rigid transformation for registration.

In Fig. 15(c) and Fig. 16(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 better solutions for side facial datasets (Fig. 15(d) and Fig. 16(d)).

The consistent landmark sets determine the pose of the face object under consideration from the alignment transformation with the corresponding FLM. Since our aim is to locate landmark sets on profile, semi-profile and profile faces, we retain the complete landmark solution only if estimated yaw-angle is within certain limits ($\pm 30^\circ$ around y-axis), otherwise the left or right landmark sets are preferred according to pose.

Finally, using the selected best solution, the registration transformation is calculated, the yaw-angle is estimated, and the facial dataset is classified as frontal, left side or right side.

Note that the use of landmark sets of 5 landmarks serves two purposes: (i) it is the potential solution for semi-profile and profile faces, and (ii) it reduces the combinatory 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 the FLMs, we create 5-tuples of landmarks, and check $N^5 + N^5 = 2N^5$ combinations for consistency with FLM5L and FLM5R. We retain 256 landmark sets consistent with FLM5L and 256 landmark sets consistent with FLM5R. By fusing them and checking consistency with FLM8 we have an extra of 256×256 combinations to be checked. Thus, by this approach $2N^5 + 256^2 \ll N^8$ combinations are checked, with $O(N^5) \ll O(N^8)$. For $N = 128$ we have approx. 69×10^9 instead of 72×10^{15} combinations to be checked.

4.5.1 Landmark Detection Methods

We applied two alternative methods for detecting the geometric candidate landmarks:

METHOD 1: Shape Index + Extrusion Map: In this method, shape index's minima are the candidate landmarks for eye and mouth corners and shape

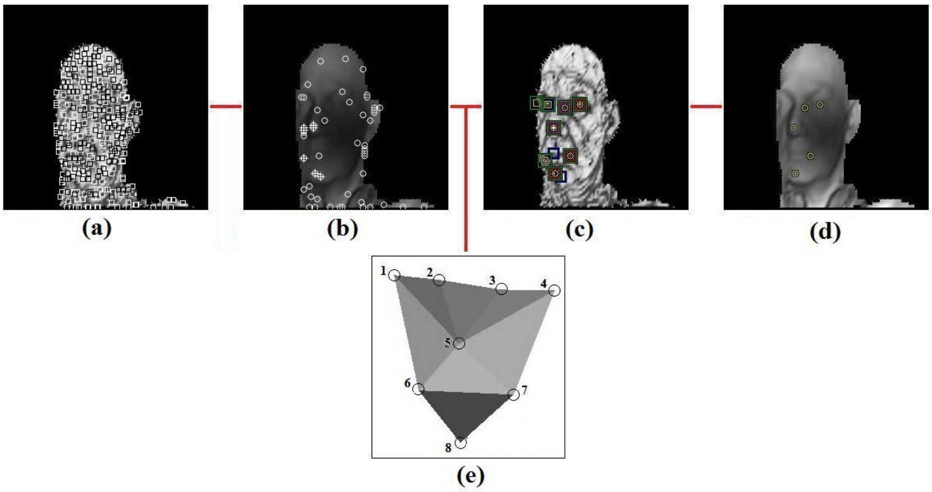


Figure 13: **METHOD 1: Shape Index + Extrusion Map**: Process pipeline: (a) shape index's maxima and minima; (b) extrusion map's candidate nose and chin tips; (c) extracted best landmark sets; (d) resulting landmarks; and (e) Facial Landmark Model (FLM) filtering.

index's maxima that are also Extrusion's map maxima are the candidate landmarks of the nose and chin tips (Fig. 13). To find the best solution, we used the *normalized Procrustes distance* D_{NP} (Eq. 41). This method will be referred as **METHOD SIEM–NP**.

METHOD 2: Shape Index + Spin Images: In this method, shape index's maxima and minima are further classified into five classes by the spin image templates and are the candidate landmarks for eye outer corner, eye inner corner, nose tip, mouth corner and chin tip (Fig. 14). To find the best solution, we used two alternative distance metrics. The *normalized Procrustes distance* D_{NP} (Eq. 41) and the *normalized Procrustes × mean spin similarity distance* D_{NPSS} (Eq. 43). These methods will be referred as **METHOD SISI–NP** and **METHOD SISI–NPSS** respectively.

4.5.2 Landmark Constraints

As previously mentioned, from the classified candidate landmark points we create combinations of 5 landmarks. Since an exhaustive search of all possible combinations of the candidate landmarks is not feasible, simple landmark position constraints from the shape model and its deformations are used to reduce the search space (pruning).

We use two types of constraints for the candidate landmarks:

Absolute Distance constraint: This constraint expresses 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 facial mean landmark shape and all

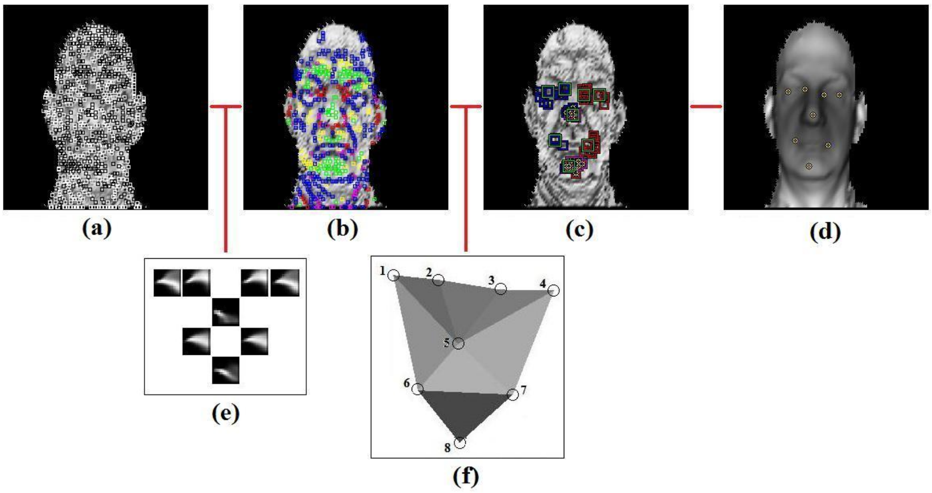


Figure 14: **METHOD 2: Shape Index + Spin Images**: Process pipeline: (a) shape index's maxima and minima; (b) spin image classification; (c) extracted best landmark sets; (d) resulting landmarks; (e) spin image templates filtering; and (f) Facial Landmark Model (FLM) filtering.

mean shape variations for the selected eigenvalues. Actually, for all modes of FLM marginal variations ($b_i = \pm 3\sqrt{\lambda_i}$), we calculate the minimum D_{min} and maximum D_{max} distance of every pair of landmarks $(\mathbf{r_i}, \mathbf{r_j})$. We constrain candidate landmark distances $|\mathbf{r_i} - \mathbf{r_j}|$ within these margins plus a tolerance t:

$$(1-t) \cdot D_{min}(\mathbf{r_i}, \mathbf{r_j}) \leq |\mathbf{r_i} - \mathbf{r_j}| \leq (1+t) \cdot D_{max}(\mathbf{r_i}, \mathbf{r_j}) \tag{37}$$

where $\mathbf{r_i}$, $\mathbf{r_j}$ denote the positions of landmarks, with $i \neq j$.

Relative Position constraint: This constraint expresses the fact that the relative positions of landmark points must be consistent with the face shape.
If we define a counter-clockwise direction **CCDir**, then the vectors from the nose tip to the other landmarks have also a counter-clockwise direction:

$$\mathbf{CCDir} = (\mathbf{r_m} - \mathbf{r_5}) \times (\mathbf{r_n} - \mathbf{r_5}) \tag{38}$$

and

$$[(\mathbf{r_i} - \mathbf{r_5}) \times (\mathbf{r_j} - \mathbf{r_5})] \cdot \mathbf{CCDir} > 0 \tag{39}$$

where $\mathbf{r_m}$, $\mathbf{r_n}$, $\mathbf{r_i}$, $\mathbf{r_j}$ denote the positions of certain landmarks and $\mathbf{r_5}$ the position of nose tip. For FLM5R:
$(m, n) = (2, 1)$ and $(i, j) \in \{(1, 6), (6, 8)\}$,
and for FLM5L:
$(m, n) = (4, 3)$ and $(i, j) \in \{(7, 4), (8, 7)\}$.

The purpose of the above constraints is to speed up the search algorithm by removing outliers and not potential solutions, so care must be taken not to be over-constrained.

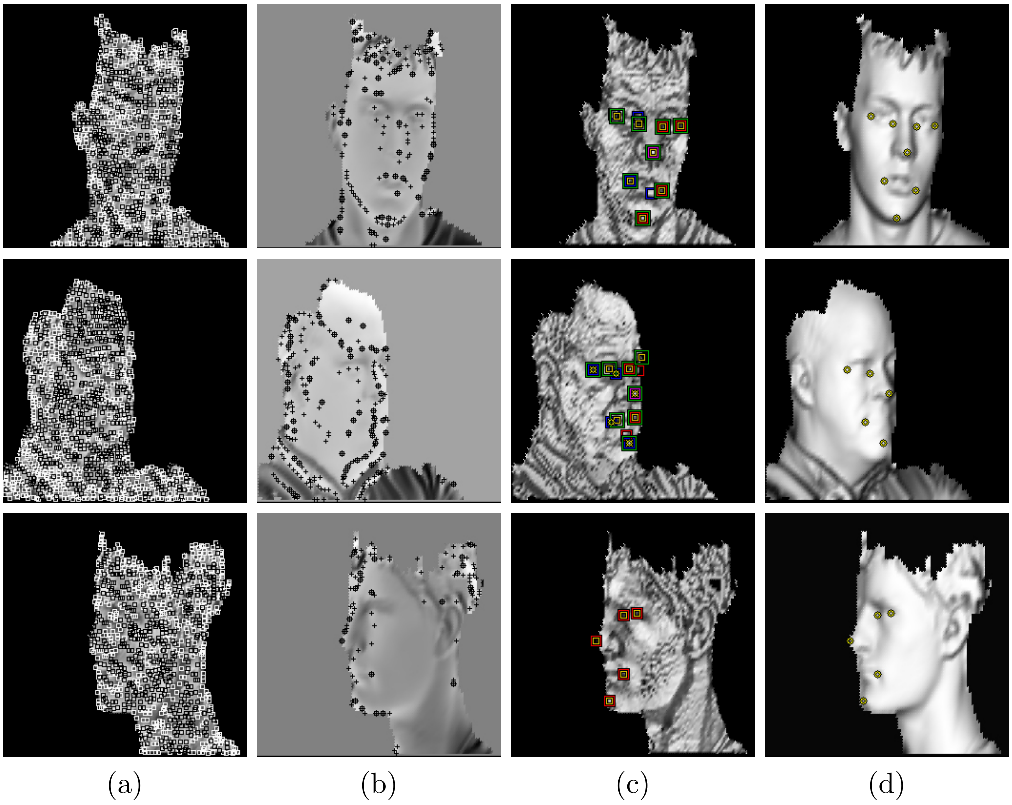


Figure 15: METHOD 1: Shape Index + Extrusion Map: Results of landmark detection and selection process: (a) shape index's maxima and minima; (b) candidate nose and chin tips; (c) extracted best landmark sets; and (d) resulting landmarks.

4.5.3 Distance Metrics

Since FLM5R, FLM5L, FLM8 have different dimensions in shape space, Procrustes distances cannot be used as a distance measure because they are not directly comparable:

$$D_P = \sqrt{\sum_{j=1}^{k} (x_j - y_j)^2} \tag{40}$$

where D_P is the Procrustes distance, $\mathbf{x}$ and $\mathbf{y}$ are the two shape vectors and k is the shape space dimension ($k = 24$ for FLM8 and $k = 15$ for FLM5R and FLM5L).

Thus, we must use alternative measures for the distance between two landmark shapes that can be comparable irrespectively of their dimensions.

We use an intuitive *normalized Procrustes distance*, taking into consideration the shape space dimensions:

$$D_{NP} = \frac{D_P}{k^2} \tag{41}$$

where D_{NP} is the normalized Procrustes distance, D_P the Procrustes distance, and k is the shape space dimension. The division by k^2 instead of k is preferred to give

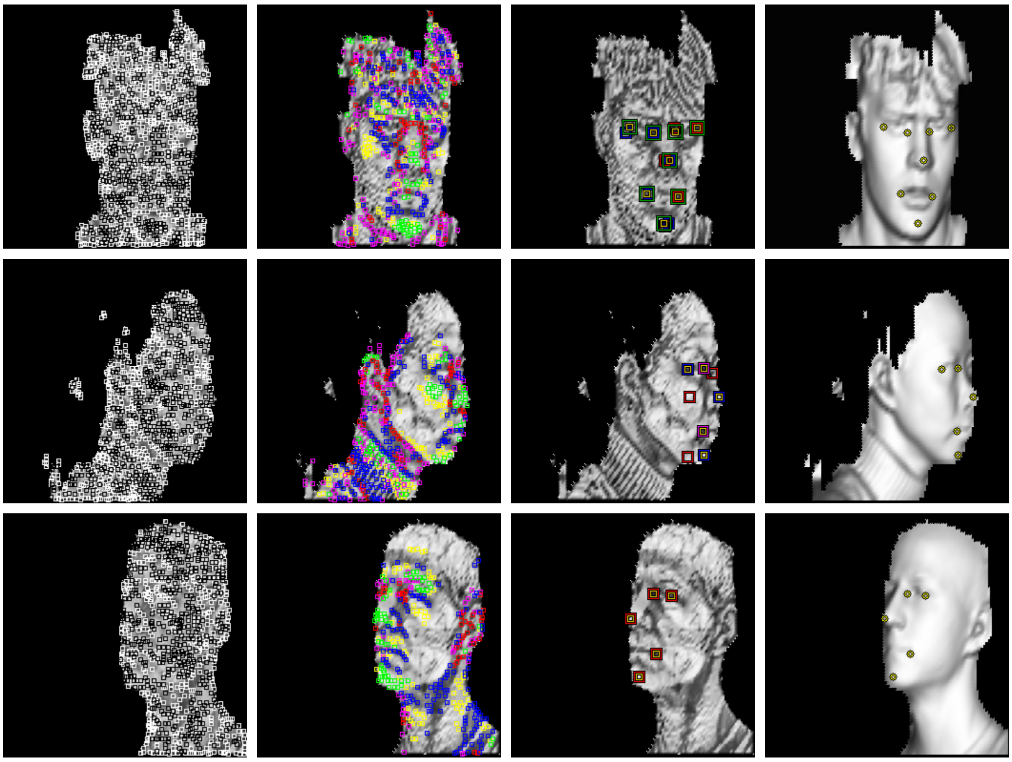


Figure 16: METHOD 2: Shape Index + Spin Images: Results of landmark detection and selection process: (a) shape index's maxima and minima; (b) spin image classification; (c) extracted best landmark sets; and (d) resulting landmarks.

a bias to the complete solution.

Another non-geometric measure of the quality of a landmark shape is its *mean spin image similarity* normalized to [0,1] (0 for high similarity and 1 for low similarity). Here, we take into consideration the spin image similarities between detected landmarks and spin image templates:

$$D_{SS} = \frac{1}{2}\left[1 - \frac{\sum_{i=1}^{n} S(P_i, Q_i)}{n}\right] \tag{42}$$

where D_{SS} is the mean spin similarity distance, $S(P_i, Q_i)$ the similarity measure between the landmark spin image grid P_i and the corresponding template Q_i, and n the number of landmarks ($n = 8$ for FLM8 and $n = 5$ for FLM5R and FLM5L).

Finally, we can use an intuitive *normalized Procrustes* × *mean spin similarity distance*, taking into consideration the geometric distance and the spin image similarities:

$$D_{NPSS} = D_{NP} \cdot D_{SS} \tag{43}$$

where D_{NP} is the normalized Procrustes distance and D_{SS} the mean spin image similarity.

An overall measure which reflects the quality of the landmark detection process is the *mean Euclidian distance* between two landmark shapes in original 3D space:

$$D_{ME} = \frac{\sum_{i=1}^{n} |\mathbf{x}_i - \mathbf{y}_i|}{n} \tag{44}$$

where D_{ME} is the mean Euclidian distance, $|\mathbf{x}_i - \mathbf{y}_i|$ the Euclidian distance between the landmark points $\mathbf{x}_i$ and $\mathbf{y}_i$ of the two shapes and n the number of landmarks ($n = 8$ for FLM8 and $n = 5$ for FLM5R and FLM5L).

Another measure which reflects the quality of the registration process is the *modified directed Hausdorff distance* D_{MDH}, of a face model M to a test face T, which is defined as [Gao03]:

$$D_{MDH}(M, T) = \frac{1}{p} \sum_{m_i, i=1}^{p} \min_{t_j} |\mathbf{m}_i - \mathbf{t}_j| \tag{45}$$

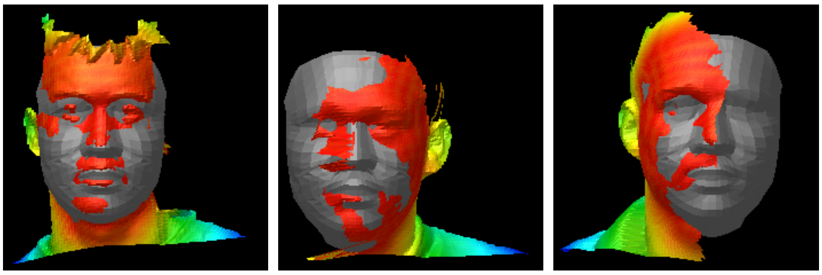
where $|\mathbf{m}_i - \mathbf{t}_j|$ is the Euclidian distance between the face model vertices $\mathbf{m}_i$ and the test face vertices $\mathbf{t}_j$, and p the number of the face model vertices. The $D_{MDH}(M, T)$ expresses the mean value of the minimum Euclidian distances $|\mathbf{m}_i - \mathbf{t}_j|$ of the vertices of the face model M, to which a test face scan T is registered.

We used the "normalized Procrustes" D_{NP} distance metric to select the best landmark set solution in "Method SIEM–NP" and "Method SISI–NP", and the "normalized Procrustes × mean spin similarity" D_{NPSS} distance metric in "Method SISI–NPSS", where spin images are available. Finally, we used the "mean Euclidian distance" D_{ME} to express the mean localization error of the landmarks and the "modified directed Hausdorff distance" D_{MDH} to express the quality of the registration process.

4.5.4 Face Registration & Pose Estimation

In a 3D face recognition system, alignment (registration) between the query and the stored datasets is necessary in order to make the probe and the gallery dataset comparable. Registration can be done against a common frame of reference, i.e. a *Reference Face Model* (RFM) of known coordinates (Fig. 17). Registration of facial datasets to a reference face model can be accomplished, by minimizing the Procrustes distance between a set of landmark points on the facial dataset and the corresponding landmark points on the Reference Face Model. Landmark points $\mathbf{x}$ on the facial datasets have to be detected by applying one of the previously mentioned methods, and landmark points $\mathbf{x_0}$ on the Reference Face Model are manually annotated once at a preprocessing stage.

Alignment of a set of face landmark points $\mathbf{x}$ to the RFM landmark points $\mathbf{x_0}$ is done by minimizing the Procrustes distance in an iterative approach:



(a) (b) (c)

Figure 17: Reference face model (RFM) and test face superposed after alignment: (a) frontal face dataset; (b) 45° left side face dataset; and (c) 60° right side face dataset. (Gray color denotes the face model. Color range – red: near to blue: far – denotes min distances of test face vertices to model.)

Algorithm 6: Face Registration

```
• Calculate $\mathbf{T}$ to translate $\mathbf{x}$ so that its centroid is at the origin (0,0,0).
• Scale $\mathbf{x}$ shape so that its size is 1.
• Calculate $\mathbf{T}_0$ to translate $\mathbf{x_0}$ so that its centroid is at the origin (0,0,0).
• Scale $\mathbf{x_0}$ shape so that its size is 1.
• REPEAT
    – Align $\mathbf{x}$ to the reference shape $\mathbf{x_0}$ by an optimal rotation $\mathbf{R}$.
    – Compute the Procrustes distance of $\mathbf{x}$ to the reference shape $\mathbf{x_0}$.
• UNTIL Convergence: $|\mathbf{x} - \mathbf{x_0}| < \varepsilon$.
• Apply $\mathbf{T}_0^{-1} \cdot \mathbf{R} \cdot \mathbf{T}$ to register face data.
```

Thus the final transformation to register a facial dataset to an RFM is:

$$\mathbf{x}' = \mathbf{T}_0 \cdot \mathbf{R} \cdot \mathbf{T} \cdot \mathbf{x} \tag{46}$$

and pose is estimated from $\mathbf{R}$. Notice that scaling can be omitted when the probe and reference shapes are of the same size.

Note that the landmark set detected on the probe facial scan (complete, right or left) determines the set of the landmarks (FLM8, FLM5R or FLM5L) used for registration with the Reference Face Model. Fig. 17 depicts the registration of a profile facial dataset to a RFM. Note that a left side 5 landmark set detected on the facial scan has to be aligned with the left 5 landmark subset of the RFM, to have a correct global registration.

As a Reference Face Model the complete Facial Landmark Model (FLM8) (Fig. 3(c)) or an Annotated Face Model can be used (Fig. 18). The *Annotated Face Model* (AFM) [KPT*07] is an anthropometrically correct 3D model of the human face [Far94]. It is constructed only once and is used in the alignment, fitting, and metadata generation for face recognition [KPT*07, PPT*09]. The AFM is annotated into different areas (e.g., mouth, nose, eyes) and can have predefined landmark points. Using

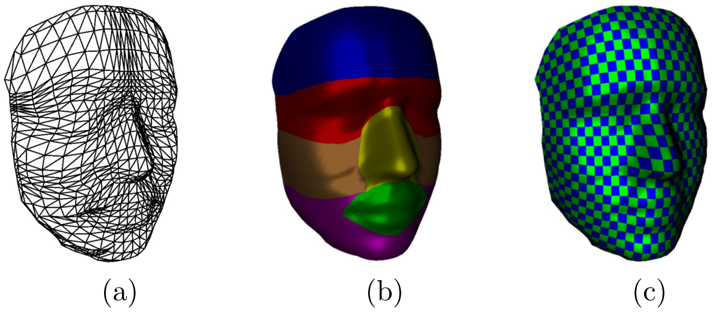


Figure 18: Annotated Face Model [KPT*07]: (a) polygonal mesh; (b) annotated areas; (c) u, v parameterization.

the global u, v parameterization of the AFM, a 2D mapping of the 3D geometric information of the facial dataset can be performed. It can also be used for facial area segmentation and certain 3D facial region retrieval [PTPK09].

5 Landmark Localization Results

5.1 Face Databases

A short description of the databases widely available to the research community is given below:

The **FRGC** [PFS*05] database from the University of Notre Dame (UND) contains 4,950 facial scans and is divided into two completely disjoint subsets: FRGC v1 and FRGC v2. The hardware used to acquire these range data was a Minolta Vivid 900/910 laser range scanner, with a resolution of 640 × 480.

The **FRGC v1** database contains 943 range images of 275 individuals, acquired before Spring 2003 (*FRGC 3D Training Set*). Subjects have neutral expressions and almost frontal pose.

The **FRGC v2** database contains a total of 4,007 range images of 466 individuals, acquired between Fall 2003 and Spring 2004 (*FRGC 3D Validation Set*). Subjects have various facial expressions (e.g., happiness, surprise) and almost frontal pose. FRGC v2 is considered more challenging than FRGC v1.

The **Ear Database** from the University of Notre Dame (UND), collections F and G [UND08]. This database (which was created for ear recognition purposes) contains side scans with a vertical rotation of 45°, 60° and 90°. In the 90° side scans, both sides of the face are occluded from the sensor, therefore these were excluded since they contain no useful information. The UND database contains 119 side scans at ±45° (119 subjects, 119 left and 119 right) and 88 side scans at ±60° (88 subjects, 88 left and 88 right).

The **MSU** [LJ06] database from the Michigan State University contains 300 multiview 3D facial scans from 100 individuals. For each subject, three scans where captured with yaw angles of less than $-45°$, 0° (frontal) and more than $+45°$.

The **BU-3DFE** [YWS*06] database from the University of New York at Binghamton contains 2500 3D facial data of 100 individuals. The system used to acquire

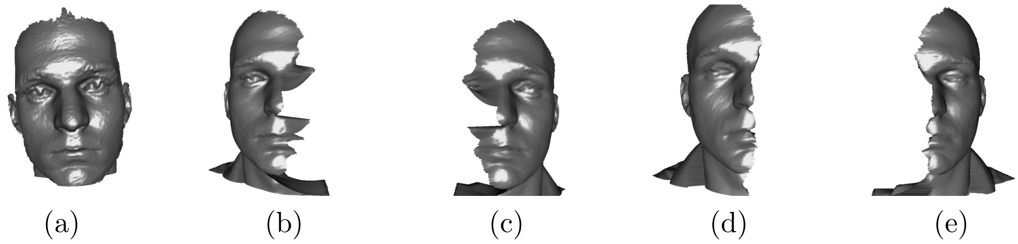


Figure 19: Front view of scans from the used database: (a) frontal (**DB00F**); (b) 45° right (**DB45R**); (c) 45° left (**DB45L**); (d) 60° right (**DB60R**); (e) 60° left (**DB60L**). Notice the extensive missing data in (b-e).

these data consists of six digital cameras and two light pattern projectors evenly positioned at 45° at each side of the subject. The system creates a single complete 3D polygon surface mesh of the face (20,000 - 35,000 polygons), by merging the cameras' viewpoints. Subjects perform seven universal expressions (i.e., *neutral*, *happiness*, *surprise*, *fear*, *sadness*, *disgust* and *anger*).

5.2 Test Databases

For performance evaluation we combined the largest publicly available 3D face and ear databases. In order to evaluate performance for the landmark detection and localization method, we manually annotated the face datasets.

For frontal facial datasets, we used the FRGC v2 database [PFS*05]. For the purposes of this evaluation we manually annotated 975 frontal facial datasets from 149 different subjects, randomly selected from the FRGC v2 database, including several subjects with various facial expressions. This database will be referred as **DB00F** (Fig. 19 (a)).

For side facial datasets, we used the Ear Database from the University of Notre Dame (UND), collections F and G [UND08]. Note that though the creators of the database marked these side scans as 45° and 60°, the measured average angle of rotation is 65° and 80°, respectively. However, when we refer to these datasets we will use the database notation (45° and 60°).

For the purposes of this evaluation, we manually annotated 118 left and 118 right 45° side datasets, which come from 118 different subjects. These databases will be referred as **DB45L** and **DB45R** respectively (Fig. 19 (b-c)). We also annotated 87 left and 87 right 60° side datasets, which come from 87 different subjects. These databases will be referred as **DB60L** and **DB60R**, respectively (Fig. 19 (d-e)). Finally, we composed a database with datasets of 39 common subjects found in **DB00F**, **DB45L** and **DB45R**. This database consists of 117 (3 × 39) scans in three poses, frontal and 45° left and right, and will be referred as **DB00F45RL**.

In the evaluation databases, only facial datasets with all the necessary landmark points visible were included (8 for frontal scans and 5 for side scans). Great care was given to the accuracy of the manual annotation procedure, since the annotated datasets form our ground truth.

5.3 Performance Evaluation

For the performance evaluation of the proposed landmark detection method, we conducted the following three experiments:

Experiment 1: In this experiment we used *Method SIEM–NP*. Thus, shape index's minima are the candidate landmarks for eye and mouth corners and shape index's maxima that are also extrusion's map maxima are the candidate landmarks of the nose and chin tips. To find the best solution, we used the *normalized Procrustes distance* D_{NP}.

Experiment 2: In this experiment we used *Method SISI–NP*. Thus, shape index's maxima and minima are further classified into five classes by the spin image templates and are the candidate landmarks for eye outer corner, eye inner corner, nose tip, mouth corner and chin tip. To find the best solution, we used the *normalized Procrustes distance* D_{NP}.

Experiment 3: In this experiment we used *Method SISI–NPSS*. Thus, shape index's maxima and minima are further classified into five classes by the spin image templates and are the candidate landmarks for eye outer corner, eye inner corner, nose tip, mouth corner and chin tip. To find the best solution, we used the *normalized Procrustes* × *mean spin similarity distance* D_{NPSS}.

The performance evaluation is generally presented by calculating the following values, which represent the localization accuracy of the detected landmarks:

Absolute Distance Error: It is 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.

Success Rate: It is the percentage of successful detections of a landmark over a tested database. Successful detection is considered a detection of a landmark with Absolute Localization Distance Error under a certain threshold (e.g., 10 mm).

In all our experiments, the mean and standard deviation of the absolute distance error between the manually annotated landmarks and the automatically detected landmarks was calculated to represent the *localization error*. Also, the overall mean distance error of the 8 landmark points for the frontal datasets and of the 5 landmark points for the side datasets was computed. This error is expressed with the "mean Euclidian distance" D_{ME} (Eq. 44) between the manually annotated landmarks and the automatically detected landmarks.

Localization error analysis for the mean and standard deviation was carried out only on results where the pose of the probe was correctly estimated, and is presented in Tables: 2, 3, 4, 5, 6 and 7. The pose detection rate is the percentage of correct pose estimations of the probe (frontal, left profile, right profile), according to the known pose of the probe, and which also have a mean distance error under 30.00 mm. A false pose estimation or a mean distance error estimation over 30.00 mm is considered a pose detection failure. Also, the yaw angle is calculated and its mean value,

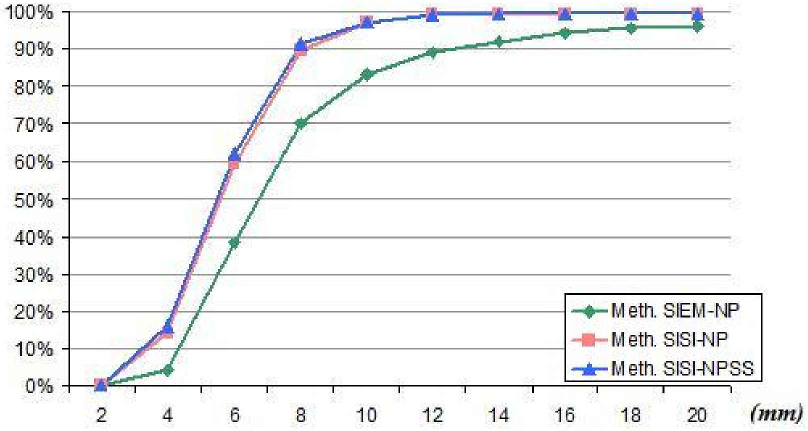


Figure 20: Mean Error (mm) Cumulative Distr. for **DB00F** (x-axis is the mean distance-error in mm between ground truth and automatically detected landmarks. y-axis is the percentage of facial datasets which have errors up to an x-value).

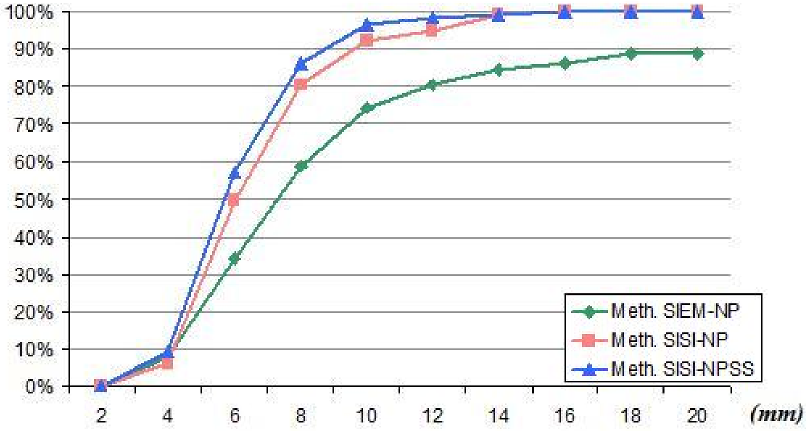


Figure 21: Mean Error (mm) Cumulative Distr. for **DB00F45RL**.

standard deviation, and minimum and maximum values are presented. Finally, we present the *success rate of landmark localization* with a threshold of 10 mm.

We depict the Cumulative Error Distribution graphs for the mean distance-error only in Figs.: 20, 21, 22, 23, 24 and 25. In these graphs the x-axis represents the mean distance-error between the manually annotated landmarks and the automatically detected landmarks on a dataset in intervals of 2 mm, and the y-axis the percentage of face datasets with a mean distance-error up to a certain x-value, out of all gallery datasets (error cumulative distribution).

Additionally, the "modified directed Hausdorff distance" D_{MDH} (Eq. 45), of the face model (RFM) to the test face is computed to express the quality of the face registration process. To get comparative results we used as a model for frontal

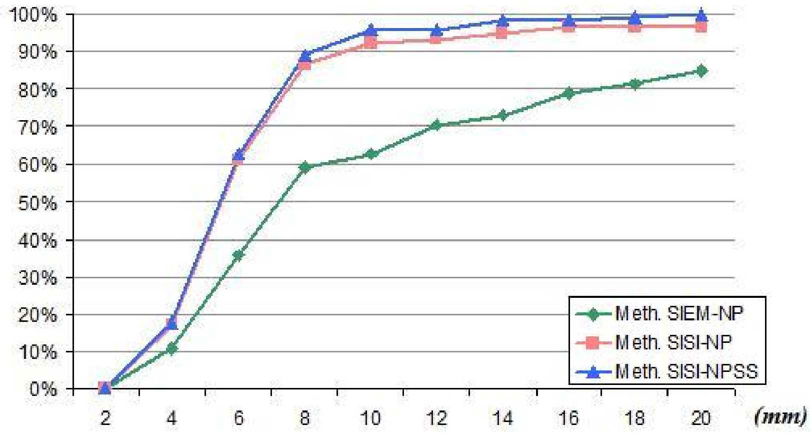


Figure 22: Mean Error (mm) Cumulative Distr. for **DB45R**.

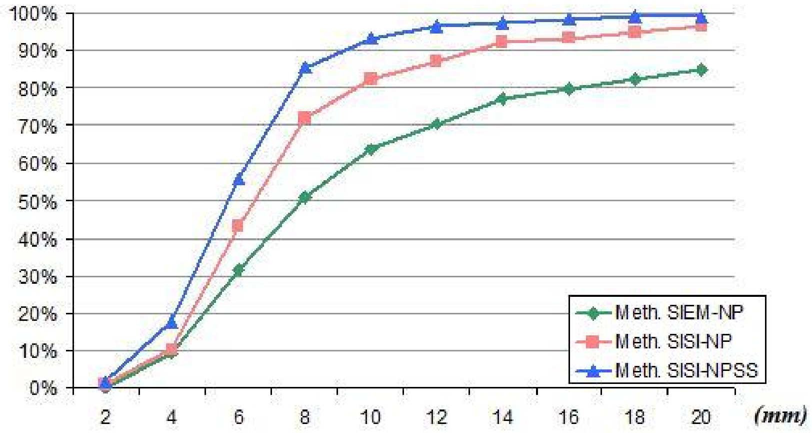


Figure 23: Mean Error (mm) Cumulative Distr. for **DB45L**.

face databases (i.e., DB00F) all the vertices of the complete RFM, for left-side databases (i.e., DB45L and DB60L) the left side vertices of the RFM, and for right-side databases (i.e., DB45R and DB60R) the right side vertices of the RFM.

Summary results over the tested databases for **METHOD SISI–NPSS** are presented in Table 1, and the mean error cumulative distribution is depicted in Fig. 26. Note that **METHOD SISI–NPSS** outperforms all other methods, having more correct pose detections, lesser distance errors and standard deviations. It also shows a higher successful detection rate irrespective of pose. The mean error (D_{ME}) for **METHOD SISI–NPSS** is under 6.1 mm on all tested facial scans, with standard deviation under 2.6 mm on frontal and side scans. Also note that pose was correctly estimated on over 97.7% of the tested facial scans, irrespective of pose. Specifically, the best results have been obtained for the frontal facial scans and the worst for the 60° left facial scans. It also shows sufficient robustness across

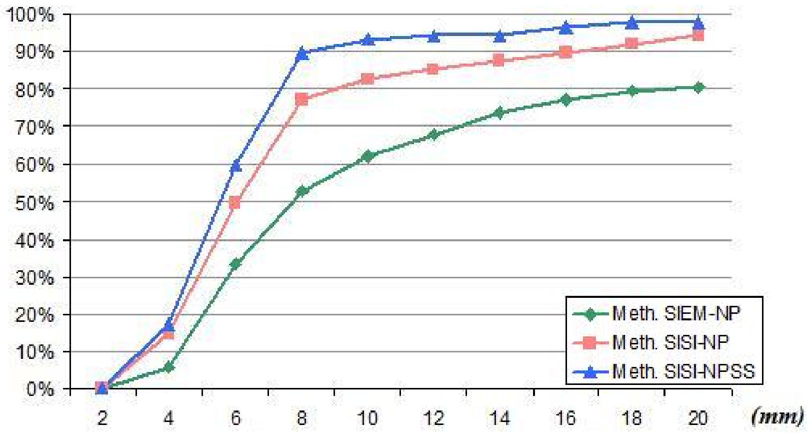


Figure 24: Mean Error (mm) Cumulative Distr. for **DB60R**.

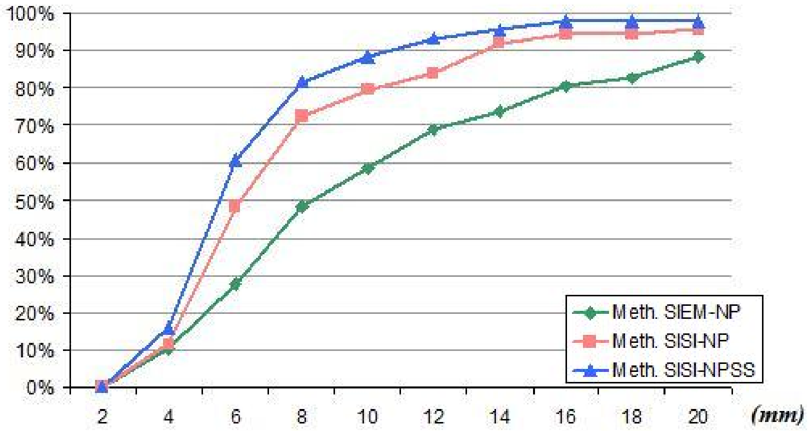


Figure 25: Mean Error (mm) Cumulative Distr. for **DB60L**.

pose variations at the registration process with Hausdorff distances (D_{MDH}) under 4.7 mm and standard deviation under 3.1 mm.

Notice that, we generally observe larger mean errors with larger standard deviations in side scan results than in frontal scans. Also note that our method performs better on 45° facial scans than on 60° facial scans. This is due to the fact that the problem of landmark detection in side scans as yaw angle increases is more difficult and the tested data set was much smaller, so outliers had a more significant effect on results.

The errors that appear in **METHOD SIEM–NP** are mainly due to the fact that the eye and mouth corners are detected from the shape index maps without any other post processing. This fact is alleviated by the spin image methods, **METHOD SISI–NP** and **METHOD SISI–NPSS**, where these landmarks are further filtered out and more robustly detected. Also note that the use of the combined 'normalized

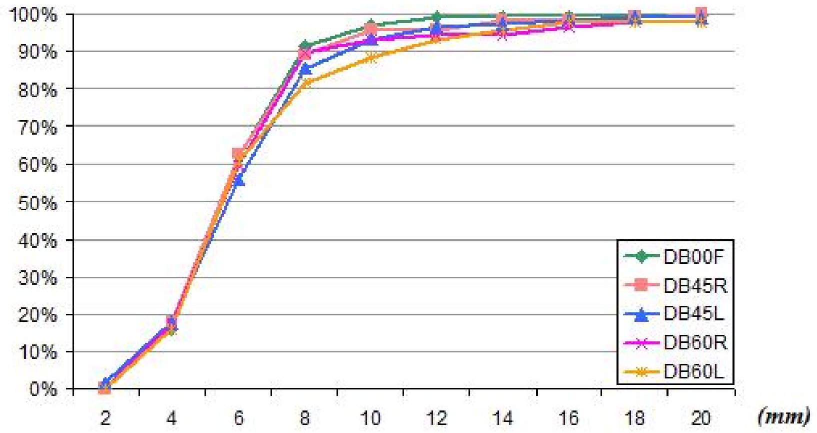


Figure 26: Mean Error (mm) Cumulative Distr. for **SSNPSS** method.

Table 1: Summary results for **METHOD SISI–NPSS**

Database	D_{MDH}		D_{ME}		
	mean (mm)	std.dev (mm)	mean (mm)	std.dev (mm)	≤ 10 (mm)
DB00F (975)	4.72	1.27	5.64	1.74	97.23%
DB45R (118)	3.90	0.95	5.83	2.49	95.76%
DB45L (118)	4.03	1.22	6.02	2.45	93.22%
DB60R (87)	4.37	3.11	5.87	2.47	93.10%
DB60L (87)	4.32	2.41	6.08	2.53	88.51%

Procrustes' and 'mean spin similarity' distance metric D_{NPSS} in **METHOD SISI–NPSS** has improved the overall results on all tested databases, irrespectively of pose.

The most robust facial features are the nose tip and eye inner corners, with a lesser mean error and standard deviation on almost all tested facial scans and for all methods. This is due to the fact that they have more distinct geometry which is more easily captured by the detectors. Contrarily, the least robust facial feature appears to be the mouth corners and chin tips mainly due to the fact that they don't have enough distinct geometry and also expressions change the positions of these landmarks significantly.

Note that the frontal test database **DB00F** contains faces with expressions, which alter facial characteristics mainly at the eyes, eyebrows, mouth and chin. Although the landmark model (FLM) used has not been trained with examples having facial expressions, it has enough tolerance and generality to accept these faces as plausible and label the corresponding landmarks successfully.

5.4 Comparative Results

For comparison of the performance of our proposed landmark detection method and other researchers' methods, we present landmark localization errors in Tables 8 and 9.

Note that each researcher uses a different facial database making direct comparisons extremely difficult. However, the comparative results presented in Tables 8 and 9 indicate that our presented **METHOD SISI–NPSS** outperforms previous methods for the following reasons: (i) it is more accurate, since it gives the minimum mean localization distance error for almost all landmarks, and (ii) it is more robust, since it gives the minimum standard deviation for the localization distance error.

Yu's method [YM08] shows the minimum mean localization error for the nose tip but has a large standard deviation. Lu's method [LJ05] shows the minimum mean localization error for the mouth corners but it is not a pure 3D method, since it is assisted by 2D intensity data. Finally, Colbry's method [Col06] seems to behave well enough for all landmarks, comparatively close to our method, but it gives larger standard deviations. Furthermore, the database used contains a small portion ($\approx 5\%$) of pose variations, occlusions and expressions.

To the best of our knowledge, Lu's method [LJ06] is the only one that presented localization errors on mixed frontal and profile facial datasets. Although it shows small mean localization errors for the mouth corners, it has a large standard deviation. On the other hand, the tested MSU database consists of a larger number of facial datasets.

Notice that for our methods, the same parameters were used in all cases (across frontal, semi-profile and profile scans). If the proposed methods are applied to databases restricted only to frontal scans or only to profile scans, the parameters can be fine-tuned to achieve higher performance.

5.5 Computational Efficiency

For the evaluation of the proposed method's computational efficiency, a typical modern PC was used: Intel Core 2 Duo 2.2GHz, 2GB RAM, NVIDIA GeForce 8600GTS. Using this PC, 9 seconds on average are required for each facial scan to locate the facial landmarks. The procedures of determining the optimal rotation for the alignment of the landmark shapes to the FLMs require at most 8 iterations to converge. The registration step (that takes 8 iterations to converge) requires less than 0.1 sec, depending on dataset size. The computational efficiency of the proposed landmark detection and registration method makes it suitable for real-world applications.

6 Conclusion

We have presented an automatic 3D facial landmark detection method that offers pose invariance and robustness to large missing facial areas, with respect to extreme pose variations. The proposed approach introduced new methods for 3D landmark localization by exploiting 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 that include pose variations up to 80° along the vertical axis. All steps of the method (landmark detection and pose estimation for registration) work robustly, even if half of the face is missing.

Future work will be directed towards increasing the robustness and accuracy of the landmark detector. This could be accomplished in two ways: firstly by increasing the examples datasets for creating the facial landmark model, including expression variations, and secondly by increasing the landmark set, including the nostrils base.

References

[CCRA*05] Conde C., Cipolla R., Rodrígez-Aragón L. J., Serrano A., Cabello E.: 3D facial feature location with spin images. In *Proc. IAPR Conference on Machine Vision Applications* (Tsukuba Science City, Japan, May 16 – 18 2005), pp. 418–421.

[Col06] Colbry D.: *Human Face Verification by Robust 3D Surface Alignment.* PhD thesis, Michigan State University, 2006.

[CSJ05] Colbry D., Stockman G., Jain A.: Detection of anchor points for 3D face verification. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (San Diego, CA, Jun. 20-25 2005), p. 118.

[CT01] Cootes T., Taylor C.: *Statistical Models of Appearance for Computer Vision.* Tech. rep., University of Manchester, Oct. 2001.

[CTCG95] Cootes T., Taylor C., Cooper D., Graham J.: Active shape models - their training and application. *Computer Vision and Image Understanding 61*, 1 (Jan. 1995), 38–59.

[CTKP05] Cootes T., Taylor C., Kang H., Petrovic V.: *Handbook of Face Recognition.* Springer, 2005, ch. Modeling Facial Shape and Appearance, pp. 39–63.

[Dib08] Dibeklioğlu H.: *Part-Based 3D Face Recognition under Pose and Expression Variations.* Master's thesis, Boğaziçi University, 2008.

[DJ97] Dorai C., Jain A. K.: COSMOS - a representation scheme for 3D free-form objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence 19*, 10 (Oct. 1997), 1115–1130.

[DM98] Dryden I., Mardia K.: *Statistical Shape Analysis.* Wiley, 1998.

[DSA08] Dibeklioğlu H., Salah A., Akarun L.: 3D facial landmarking under expression, pose, and occlusion variations. In *Proc. 2^{nd} IEEE International Conference on Biometrics: Theory, Applications and Systems* (Arlington, VA, Sep. 20 - Oct. 1 2008), pp. 1–6.

[Far94] Farkas L.: *Anthropometry of the head and face*, 2^{nd} ed. Raven Press, 1994.

[FBF08a] Faltemier T., Bowyer K., Flynn P.: A region ensemble for 3-D face recognition. *IEEE Transactions on Information Forensics and Security 3*, 1 (Mar. 2008), 62–73.

[FBF08b] Faltemier T., Bowyer K., Flynn P.: Rotated profile signatures for robust 3D feature detection. In *Proc. 8^{th} IEEE International Conference on Automatic Face and Gesture Recognition* (Amsterdam, The Netherlands, Sep. 17-19 2008), pp. 1–7.

[Gao03] Gao Y.: Efficiently comparing face images using a modified Hausdorff distance. In *Proc. IEEE Conference on Vision, Image and Signal Processing* (Dec. 2003), pp. 346–350.

[HS88] Harris C., Stephens M.: A combined corner and edge detector. In *Proc. 4th Alvey Vision Conference* (1988), pp. 147–151.

[Joh97] Johnson A. E.: *Spin Images: A Representation for 3-D Surface Matching.* PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, Aug. 1997.

[KPT*07] Kakadiaris I., Passalis G., Toderici G., Murtuza M., Lu Y., Karampatziakis N., Theoharis T.: Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence 29*, 4 (Apr. 2007), 640–649.

[KvD92] Koenderink J., van Doorn A.: Surface shape and curvature scales. *Image and Vision Computing 10* (Oct. 1992), 557–565.

[LJ05] Lu X., Jain A.: *Multimodal Facial Feature Extraction for Automatic 3D Face Recognition.* Tech. Rep. MSU-CSE-05-22, Michigan State University, Oct. 2005.

[LJ06] Lu X., Jain A.: Automatic feature extraction for multiview 3D face recognition. In *Proc. 7th International Conference on Automatic Face and Gesture Recognition* (Southampton, UK, Apr. 10-12 2006), pp. 585–590.

[LJC06] Lu X., Jain A., Colbry D.: Matching 2.5D face scans to 3D models. *IEEE Transactions on Pattern Analysis and Machine Intelligence 28*, 1 (2006), 31–43.

[LSCH06] Lin T., Shih W., Chen W., Ho W.: 3D face authentication by mutual coupled 3D and 2D feature extraction. In *Proc. 44th ACM Southeast Regional Conference* (Melbourne, FL, Mar. 10 – 12 2006), pp. 423–427.

[MBO07] Mian A., Bennamoun M., Owens R.: An efficient multimodal 2D-3D hybrid approach to automatic face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence 29*, 11 (Nov. 2007), 1927–1943.

[NC09] Nair P., Cavallaro A.: 3-D face detection, landmark localization, and registration using a point distribution model. *IEEE Transactions on Multimedia 11*, 4 (June 2009), 611–623.

[PFS*05] Phillips P., Flynn P., Scruggs T., Bowyer K., Chang J., Hoffman K., Marques J., Min J., Worek W.: Overview of the Face Recognition Grand Challenge. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (San Diego, CA, 2005), pp. 947–954.

[PPT*09] PERAKIS P., PASSALIS G., THEOHARIS T., TODERICI G., KAKADIARIS I.: Partial matching of interpose 3D facial data for face recognition. In *Proc. 3rd IEEE International Conference on Biometrics: Theory, Applications and Systems* (Arlington, VA, Sep. 28-30 2009), pp. 439–446.

[PTPK09] PERAKIS P., THEOHARIS T., PASSALIS G., KAKADIARIS I.: Automatic 3D facial region retrieval from multi-pose facial datasets. In *Proc. Eurographics Workshop on 3D Object Retrieval* (Munich, Germany, Mar. 30 - Apr. 3 2009), pp. 37–44.

[RHP08] ROMERO-HUERTAS M., PEARS N.: 3D facial landmark localization by matching simple descriptors. In *Proc. 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems* (Arlington, VA, Sep. 20 - Oct. 1 2008).

[SG02] STEGMAN M., GOMEZ D.: *A Brief Introduction to Statistical Shape Analysis.* Tech. rep., Technical University of Denmark, Mar. 2002.

[SQBS07] SEGUNDO M., QUEIROLO C., BELLON O., SILVA L.: Automatic 3D facial segmentation and landmark detection. In *Proc. 14th International Conference on Image Analysis and Processing* (Modena, Italy, Sep. 10-14 2007), pp. 431–436.

[TK06] THEODORIDIS S., KOUTROUMBAS K.: *Pattern Recognition*, 3rd ed. Academic Press, 2006.

[UND08] UND: University of Notre Dame biometrics database. http://www.nd.edu/∼cvrl/UNDBiometricsDatabase.html, 2008.

[WLY07] WEI X., LONGO P., YIN L.: *LNCS, Advances in Biometrics.* Springer, 2007, ch. Automatic Facial Pose Determination of 3D Range Data for Face Model and Expression Identification, pp. 144–153.

[XTWQ06] XU C., TAN T., WANG Y., QUAN L.: Combining local features for robust nose location in 3D facial data. *Pattern Recognition Letters 27*, 13 (2006), 62–73.

[YM08] YU T., MOON Y.: A novel genetic algorithm for 3D facial landmark localization. In *Proc. 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems* (Arlington, VA, Sep. 20 - Oct. 1 2008).

[YWS*06] YIN L., WEI X., SUN Y., WANG J., ROSATO M.: A 3D facial expression database for facial behavior research. In *Proc. 7th International Conference on Automatic Face and Gesture Recognition* (Southampton, UK, Apr. 10-12 2006), pp. 211–216.

A Results Tables

Table 2: Results for **DB00F**

Experiment 1: METHOD SIEM–NP

Correct Pose Detection		947 / 975	97.13%
Yaw Estimation	$-0.94° \pm 5.58°$ $[-14.57° \sim +19.53°]$		
Localization Error	mean (mm)	std.dev. (mm)	$\leq$ 10 mm
Right Eye Outer Corner	6.92	4.47	77.64%
Right Eye Inner Corner	6.18	3.71	86.26%
Left Eye Inner Corner	6.85	3.63	83.69%
Left Eye Outer Corner	7.18	4.84	75.90%
Nose Tip	6.32	4.12	86.87%
Mouth Right Corner	7.42	5.60	71.18%
Mouth Left Corner	7.84	5.22	68.72%
Chin Tip	10.26	6.59	58.67%
Mean Distance Error	7.37	3.20	83.18%

Experiment 2: METHOD SISI–NP

Correct Pose Detection		970 / 975	99.49%
Yaw Estimation	$-0.63° \pm 5.02°$ $[-13.95° \sim +16.47°]$		
Localization Error	mean (mm)	std.dev. (mm)	$\leq$ 10 mm
Right Eye Outer Corner	5.87	3.52	86.56%
Right Eye Inner Corner	5.19	2.54	94.56%
Left Eye Inner Corner	5.59	2.63	94.36%
Left Eye Outer Corner	5.81	3.71	87.79%
Nose Tip	5.28	2.40	96.10%
Mouth Right Corner	5.71	4.30	83.69%
Mouth Left Corner	6.47	4.14	82.15%
Chin Tip	6.30	4.36	88.21%
Mean Distance Error	5.78	1.78	97.23%

Experiment 3: METHOD SISI–NPSS

Correct Pose Detection		970 / 975	99.49%
Yaw Estimation	$-0.63° \pm 5.02°$ $[-13.95° \sim +16.47°]$		
Localization Error	mean (mm)	std.dev. (mm)	$\leq$ 10 mm
Right Eye Outer Corner	5.82	3.43	87.18%
Right Eye Inner Corner	5.10	2.53	94.87%
Left Eye Inner Corner	5.52	2.56	94.56%
Left Eye Outer Corner	5.70	3.52	88.10%
Nose Tip	4.88	2.42	96.72%
Mouth Right Corner	5.64	4.26	84.62%
Mouth Left Corner	6.42	4.17	82.15%
Chin Tip	6.03	4.27	89.54%
Mean Distance Error	5.64	1.74	97.23%

Table 3: Results for **DB00F45RL**

Experiment 1: METHOD SIEM–NP

Correct Pose Detection	105 / 117 89.74%		
Yaw Estimation	$-1.32° \pm 36.50°$ $[-66.09° \sim +79.81°]$		
Localization Error	mean (mm)	std.dev. (mm)	$\leq$ 10 mm
Right Eye Outer Corner	7.17	3.96	68.55%
Right Eye Inner Corner	6.74	3.65	76.03%
Left Eye Inner Corner	7.60	3.69	72.29%
Left Eye Outer Corner	7.91	4.71	64.81%
Nose Tip	6.61	4.92	76.92%
Mouth Right Corner	6.17	5.22	74.79%
Mouth Left Corner	7.29	5.64	68.55%
Chin Tip	8.85	6.02	63.25%
Mean Distance Error	7.49	3.32	74.36%

Experiment 2: METHOD SISI–NP

Correct Pose Detection	117 / 117 100.00%		
Yaw Estimation	$-1.45° \pm 37.05°$ $[-67.39° \sim +60.58°]$		
Localization Error	mean (mm)	std.dev. (mm)	$\leq$ 10 mm
Right Eye Outer Corner	6.91	3.94	71.79%
Right Eye Inner Corner	5.87	3.17	89.74%
Left Eye Inner Corner	6.78	2.97	87.18%
Left Eye Outer Corner	6.82	3.89	78.21%
Nose Tip	5.62	3.70	87.18%
Mouth Right Corner	6.13	4.56	84.62%
Mouth Left Corner	7.05	4.72	78.21%
Chin Tip	6.79	4.50	84.62%
Mean Distance Error	6.55	2.26	92.31%

Experiment 3: METHOD SISI–NPSS

Correct Pose Detection	117 / 117 100.00%		
Yaw Estimation	$-1.45° \pm 37.05°$ $[-67.39° \sim +60.58°]$		
Localization Error	mean (mm)	std.dev. (mm)	$\leq$ 10 mm
Right Eye Outer Corner	6.49	3.93	76.92%
Right Eye Inner Corner	5.91	3.02	92.31%
Left Eye Inner Corner	6.90	2.85	85.90%
Left Eye Outer Corner	6.40	3.74	79.49%
Nose Tip	4.30	2.93	94.02%
Mouth Right Corner	5.70	4.10	87.18%
Mouth Left Corner	6.20	4.30	83.33%
Chin Tip	6.40	4.17	87.18%
Mean Distance Error	5.96	1.92	96.58%

Table 4: Results for **DB45R**

Experiment 1: METHOD SIEM–NP

Correct Pose Detection		100 / 118	84.75%
Yaw Estimation	+44.39° ± 8.83° [+18.09° ∼ +79.81°]		
Localization Error	mean (mm)	std.dev. (mm)	≤ 10 mm
Right Eye Outer Corner	7.60	4.73	61.86%
Right Eye Inner Corner	6.46	3.70	68.64%
Nose Tip	7.69	6.69	66.95%
Mouth Right Corner	7.12	6.60	62.71%
Chin Tip	10.23	7.79	55.08%
Mean Distance Error	7.82	4.21	62.71%

Experiment 2: METHOD SISI–NP

Correct Pose Detection		115 / 118	97.46%
Yaw Estimation	+43.25° ± 8.13° [+20.07° ∼ +71.42°]		
Localization Error	mean (mm)	std.dev. (mm)	≤ 10 mm
Right Eye Outer Corner	6.31	3.90	77.12%
Right Eye Inner Corner	6.17	3.43	84.75%
Nose Tip	5.50	4.39	87.29%
Mouth Right Corner	5.70	4.80	80.51%
Chin Tip	6.00	4.75	87.29%
Mean Distance Error	5.93	2.51	92.37%

Experiment 3: METHOD SISI–NPSS

Correct Pose Detection		118 / 118	100.00%
Yaw Estimation	+43.35° ± 8.26° [+20.07° ∼ +71.42°]		
Localization Error	mean (mm)	std.dev. (mm)	≤ 10 mm
Right Eye Outer Corner	6.62	3.73	80.51%
Right Eye Inner Corner	6.03	3.29	88.14%
Nose Tip	4.39	3.47	94.92%
Mouth Right Corner	5.96	4.98	81.36%
Chin Tip	6.17	5.21	92.37%
Mean Distance Error	5.83	2.49	95.76%

Table 5: Results for **DB45L**

Experiment 1: METHOD SIEM–NP

Correct Pose Detection	105 / 118		88.98%
Yaw Estimation	−47.19° ± 9.19° [−84.14° ∼ −19.43°]		
Localization Error	mean (mm)	std.dev. (mm)	≤ 10 mm
Left Eye Outer Corner	8.41	5.62	60.17%
Left Eye Inner Corner	8.09	4.53	63.56%
Nose Tip	8.21	7.89	66.95%
Mouth Left Corner	9.50	7.48	56.78%
Chin Tip	10.37	7.36	57.63%
Mean Error	8.92	5.17	63.56%

Experiment 2: METHOD SISI–NP

Correct Pose Detection	115 / 118		97.46%
Yaw Estimation	−45.71° ± 8.58° [−68.17° ∼ −13.83°]		
Localization Error	mean (mm)	std.dev. (mm)	≤ 10 mm
Left Eye Outer Corner	7.00	5.08	77.12%
Left Eye Inner Corner	7.02	3.93	81.36%
Nose Tip	6.82	5.00	78.81%
Mouth Left Corner	8.17	5.89	67.80%
Chin Tip	7.18	5.63	78.81%
Mean Distance Error	7.24	3.48	82.20%

Experiment 3: METHOD SISI–NPSS

Correct Pose Detection	117 / 118		99.15%
Yaw Estimation	−45.43° ± 8.76° [−68.17° ∼ −13.83°]		
Localization Error	mean (mm)	std.dev. (mm)	≤ 10 mm
Left Eye Outer Corner	6.39	4.21	77.12%
Left Eye Inner Corner	6.61	3.04	86.44%
Nose Tip	4.02	2.70	95.76%
Mouth Left Corner	6.79	5.28	77.12%
Chin Tip	6.29	4.62	85.59%
Mean Distance Error	6.02	2.45	93.22%

Table 6: Results for **DB60R**

Experiment 1: METHOD SIEM–NP

Correct Pose Detection		72 / 87	82.76%
Yaw Estimation	+58.11° ± 6.70° [+32.10° ∼ +74.89°]		
Localization Error	mean (mm)	std.dev. (mm)	≤ 10 mm
Right Eye Outer Corner	7.30	5.34	63.22%
Right Eye Inner Corner	6.81	4.83	70.11%
Nose Tip	7.51	5.05	70.11%
Mouth Right Corner	8.02	7.48	55.17%
Chin Tip	11.55	8.87	43.68%
Mean Distance Error	8.24	4.43	62.07%

Experiment 2: METHOD SISI–NP

Correct Pose Detection		82 / 87	94.25%
Yaw Estimation	+57.00° ± 7.19° [+31.22° ∼ +74.60°]		
Localization Error	mean (mm)	std.dev. (mm)	≤ 10 mm
Right Eye Outer Corner	7.11	5.26	74,71%
Right Eye Inner Corner	5.90	4.33	83.91%
Nose Tip	5.56	4.45	85.06%
Mouth Right Corner	7.01	5.87	71.26%
Chin Tip	8.04	6.07	71.26%
Mean Distance Error	6.72	3.55	82.76%

Experiment 3: METHOD SISI–NPSS

Correct Pose Detection		85 / 87	97.70%
Yaw Estimation	+56.97° ± 7.38° [+31.22° ∼ +74.60°]		
Localization Error	mean (mm)	std.dev. (mm)	≤ 10 mm
Right Eye Outer Corner	6.88	4.34	77.01%
Right Eye Inner Corner	5.81	3.39	86.21%
Nose Tip	3.98	3.20	95.40%
Mouth Right Corner	6.29	5.08	79.31%
Chin Tip	6.38	4.85	85.06%
Mean Distance Error	5.87	2.47	93.10%

Table 7: Results for **DB60L**

Experiment 1: METHOD SIEM–NP

Correct Pose Detection		78 / 87	89.66%
Yaw Estimation	−60.57° ± 8.71° [−87.87° ∼ −38.61°]		
Localization Error	mean (mm)	std.dev. (mm)	≤ 10 mm
Left Eye Outer Corner	7.41	5.30	67.82%
Left Eye Inner Corner	8.11	4.60	64.37%
Nose Tip	7.25	6.43	71.26%
Mouth Left Corner	10.66	7.46	49.43%
Chin Tip	10.76	7.64	54.02%
Mean Distance Error	8.84	4.52	58.62%

Experiment 2: METHOD SISI–NP

Correct Pose Detection		83 / 87	95.40%
Yaw Estimation	−57.63° ± 7.46° [−74.41° ∼ −32.49°]		
Localization Error	mean (mm)	std.dev. (mm)	≤ 10 mm
Left Eye Outer Corner	5.56	3.73	83.91%
Left Eye Inner Corner	6.66	3.24	79.31%
Nose Tip	5.49	4.63	83.91%
Mouth Left Corner	8.36	6.19	66.67%
Chin Tip	8.20	6.60	72.41%
Mean Distance Error	6.85	3.27	79.31%

Experiment 3: METHOD SISI–NPSS

Correct Pose Detection		85 / 87	97.70%
Yaw Estimation	−57.46° ± 7.45° [−74.41° ∼ −32.49°]		
Localization Error	mean (mm)	std.dev. (mm)	≤ 10 mm
Left Eye Outer Corner	5.66	3.45	87.36%
Left Eye Inner Corner	6.54	3.15	85.06%
Nose Tip	3.67	1.98	95.40%
Mouth Left Corner	7.78	5.93	73.56%
Chin Tip	6.77	6.02	81.61%
Mean Distance Error	6.08	2.53	88.51%

Table 8: Comparison of landmark localization error of different approaches on almost-frontal complete facial datasets

Mean Localization Error (mm)

Method		Test DB (scans)	REIC	LEIC	REOC	LEOC	NT	CT	MRC	MLC
[YM08]	(GA model)	FRGC v1 (200)	4.74	5.59	-	-	2.18	-	-	-
[NC09]	(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	-	-	-
[LJ06]	(3D)	FRGC v1 (953)	8.30	8.20	9.50	10.30	8.30	-	6.00	6.20
[LJ05]	(3D+2D)	FRGC v1 (946)	6.00	5.70	7.10	7.90	5.00	-	3.60	3.60
[Col06]	(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
[PTPK09]	(EG-3DOR)	FRGC v2 (975)	7.02	7.46	8.13	9.21	5.23	6.71	8.30	9.83
Perakis *et al.* (current)	(SIEM–NP)	FRGC v2 (975)	6.18	6.85	6.92	7.18	6.32	10.26	7.42	7.84
	(SISI–NP)		5.19	5.59	5.87	5.81	5.28	6.30	5.71	6.47
	(SISI–NPSS)		5.10	5.52	5.82	5.70	4.88	6.03	5.64	6.42

Std. Dev. of Localization Error (mm)

Method		Test DB (scans)	REIC	LEIC	REOC	LEOC	NT	CT	MRC	MLC
[YM08]	(GA model)	FRGC v1 (200)	9.76	16.08	-	-	6.83	-	-	-
[NC09]	(w/o PDM)	BU-3DFE (2350)	-	-	-	-	-	-	-	-
	(w PDM)		-	-	-	-	-	-	-	-
[LJ06]	(3D)	FRGC v1 (953)	17.20	17.20	17.10	18.10	19.40	-	16.90	17.90
[LJ05]	(3D+2D)	FRGC v1 (946)	3.30	3.00	5.90	5.10	2.40	-	3.30	2.90
[Col06]	(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
[PTPK09]	(EG-3DOR)	FRGC v2 (975)	3.18	3.07	3.79	4.25	3.28	4.32	4.53	4.47
Perakis *et al.* (current)	(SIEM–NP)	FRGC v2 (975)	3.71	3.63	4.47	4.84	4.12	6.59	5.60	5.22
	(SISI–NP)		2.54	2.63	3.52	3.71	2.40	4.36	4.30	4.14
	(SISI–NPSS)		2.53	2.56	3.43	3.52	2.42	4.27	4.26	4.17

Landmarks:
REIC & LEIC: Right and Left Eye Inner Corners
REOC & LEOC: Right and Left Eye Outer Corners
NT & CT: Nose and Chin Tips
MRC & MLC: Mouth Right and Left Corners

Table 9: Comparison of landmark localization error of different approaches on mixed (frontal and profile) facial datasets

Mean Localization Error (mm)

Method		Test DB (scans)	REIC	LEIC	REOC	LEOC	NT	CT	MRC	MLC
[LJ06]	(3D)	MSU (300)	9.00	7.10	13.60	13.30	6.40	-	6.70	5.20
Perakis *et al.* (current)	(SIEM–NP)	FRGC v2	6.74	7.60	7.17	7.91	6.61	8.85	6.17	7.29
	(SISI–NP)	+ Ear (UND)	5.87	6.78	6.91	6.82	5.62	6.79	6.13	7.05
	(SISI–NPSS)	(117)	5.91	6.90	6.49	6.40	4.30	6.40	5.70	6.20

Std. Dev. of Localization Error (mm)

Method		Test DB (scans)	REIC	LEIC	REOC	LEOC	NT	CT	MRC	MLC
[LJ06]	(3D)	MSU (300)	13.10	9.20	11.90	10.10	13.40	-	12.90	9.00
Perakis *et al.* (current)	(SIEM–NP)	FRGC v2	3.65	3.69	3.96	4.71	4.92	6.02	5.22	5.64
	(SISI–NP)	+ Ear (UND)	3.17	2.97	3.94	3.89	3.70	4.50	4.56	4.72
	(SISI–NPSS)	(117)	3.02	2.85	3.93	3.74	2.93	4.17	4.10	4.30

Landmarks:
REIC & LEIC: Right and Left Eye Inner Corners
REOC & LEOC: Right and Left Eye Outer Corners
NT & CT: Nose and Chin Tips
MRC & MLC: Mouth Right and Left Corners