Eurographics Workshop on 3D Object Retrieval (2009)
I. Pratikakis, M. Spagnuolo, T. Theoharis, and R. Veltkamp (Editors)

Automatic 3D Facial Region Retrieval from Multi-pose Facial Datasets

P. Perakis[1,2] and T. Theoharis[1,2] and G. Passalis[1,2] and I.A. Kakadiaris[2]

[1] Computer Graphics Laboratory, Department of Informatics & Telecommunications, University of Athens, Greece
[2] Computational Biomedicine Laboratory, Department of Computer Science, University of Houston, Texas USA

Abstract

The availability of 3D facial datasets is rapidly growing, mainly as a result of medical and biometric applications. These applications often require the retrieval of specific facial areas (such as the nasal region). The most crucial step in facial region retrieval is the detection of key 3D facial landmarks (e.g., the nose tip). A key advantage of 3D facial data over 2D facial data is their pose invariance. Any landmark detection method must therefore also be pose invariant. In this paper, we present the first 3D facial landmark detection method that works in datasets with pose rotations of up to 80° *around the y-axis. It is tested on the largest publicly available 3D facial datasets, for which we have created a ground truth by manually annotating the 3D landmarks. Landmarks automatically detected by our method are then used to robustly retrieve facial regions from 3D facial datasets.*

Categories and Subject Descriptors (according to ACM CCS): H.3.3 [Information Storage & Retrieval]: Information Search & Retrieval—Selection Process; I.4.6 [Image Processing & Computer Vision]: Segmentation—Edge & Feature Detection

1. Introduction

In a wide variety of disciplines it is of great practical importance to measure, describe and compare the shapes of objects. In biometric and medical applications the class of objects is often the human face. As scanning methods become more accessible due to lower cost and greater flexibility, the number of 3D facial datasets is rapidly growing. It is often the case that, within such 3D facial datasets, one needs to automatically retrieve certain facial regions (e.g., the nasal region).

In a 3D data retrieval system, alignment (registration) between the query and the stored datasets is necessary in order to make the probe and the template comparable. Registration based on feature points' (landmarks) correspondence is the most crucial step in order to make the retrieval system fully automatic. In addition, a landmark detection algorithm must be pose invariant in order to maintain the pose invariance of 3D facial data.

All previously proposed 3D feature detection and localization methods, although they claim pose invariance, fail to address large pose variations and to confront the problem 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" datasets.

In this paper, we present a new way to automatically and pose-invariantly detect landmarks (eye and mouth corners, nose and chin tips) on 3D facial objects, and hence consistently retrieve prescribed facial regions from large 3D facial databases. The main contribution of our proposed method is its applicability to large pose variations (up to 80° around y-axis), in an holistic way with remarkable success rates.

At the training phase, the method creates an Active Landmark Model (ALM) 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 retrieval phase, the algorithm first detects candidate landmarks in the queried facial datasets by identifying Shape Index extrema. The extracted candidate landmarks are then filtered out and classified by calculating their similarity to Spin Image templates. Then, the ALM is used to select the final optimal subset of landmark locations, and surface regions are extracted around the selected landmarks by using a clipping volume. Evaluation is performed by computing the distance between manually annotated landmarks (ground truth) and the automatically detected landmarks.

The rest of this paper is organized as follows: Section 2 describes related work in the field, Section 3 presents in detail the proposed method, Section 4 presents our results, while Section 5 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 characteristics of 2D intensity or depth images, and detection of feature points (landmarks) from the geometric information of 3D objects. Facial feature detectors can also be distinguished into those that are solely dependent on geometric information and those that are supported by a trained statistical feature model.

Three-D facial feature extraction has gained interest with the increasing development of 3D modeling and digitizing techniques.

Colbry *et al.* [Col06, LJ06, LJC06, CSJ05, LJ05], in a series of works, have presented methods to locate the positions of eye and mouth corners, and nose and chin tips, based on a fusion scheme of shape index on range maps and the "cornerness" response on intensity maps. They also used 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. Conde *et al.* [CCRA*05] introduced a global face registration method by combining clustering techniques over discrete curvature and Spin Images for facial feature detection. The method was tested on a database of 51 subjects with 14 captures each (714 scans). Although they presented a feature localization success rate of 99.66% on frontal scans, and 96.08% on side scans, their database consisted of scans with small pose variations ($< 15^\circ$ around y-axis). Dibeklioglu [Dib06] introduced a nose tip localization and segmentation method using curvature based heuristic analysis to enable pose correction in a face recognition system that allows identification under significant pose variations. A significant limitation of the proposed system is that it is not applicable to yaw rotations greater than 45° and partially occluded faces. Additionally, even though the Bosphorus database he uses consists of 3,396 facial scans, they belong to only 81 subjects. 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 as the extreme vertex along the normal direction of eye sockets. The method was used in an automatic 3D face authentication system but was tested on only 27 human faces with various poses and expressions. Wei *et al.* [WLY07] introduced a nose tip localization method to determine the facial pose. The method was based on a Surface Normal Difference algorithm and Shape Index estimation, and used as a preprocessing step in pose-variant systems to determine the pose of the face. No claims where made with respect to the invariance in pose variations. Segundo *et al.* [SQBS07] introduced a face and facial feature detection method by combining an adapted method for 2D face segmentation on depth images with the surface curvature information for detecting facial features such as eye corners and nose tip. The method was tested on the FRGC 2.0 data with over 99.7% correct detections. However, nose and eye corner detection presented problems when the face had a significant pose variation ($> 15^\circ$ around the y and z-axes).

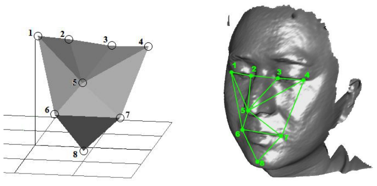
3. 3D Facial Region Localization & Retrieval

Our new method for 3D Facial Region Localization uses 3D information to extract candidate interest points, which are identified and labeled as anatomical landmarks by matching them with an Active Landmark Model (ALM) [CT01]. Once anatomical landmarks are localized, the corresponding facial regions are extracted by clipping them with a bounding volume of proper position and dimensions [TPPP08].

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

Note that five of these landmarks are visible in side scans (right side contains landmarks 1, 2, 5, 6, 8 and left side contains 3, 4, 5, 7, 8). These sets of landmarks constitute an Active Landmark Model (ALM). In the following, the model of the complete set of 8 landmarks will be referred to as ALM8 and of the two reduced sets of 5 landmarks (left and right) as ALM5L and ALM5R respectively. The steps to create the ALMs are:

- A statistical Mean Shape for each landmark set (ALM8, ALM5L and ALM5R) is estimated from a manually annotated training set (150 frontal faces with neutral expressions) using Procrustes Analysis.



(a) (b)

Figure 1: *Landmark models: (a) Landmark model as a 3D object; (b) Landmark model overlaid over a 3D facial dataset.*

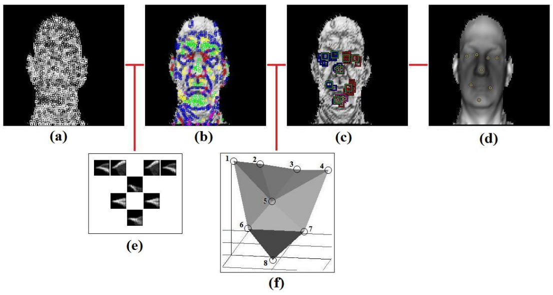


Figure 2: *Landmark detection procedure for facial region localization: (a) Shape Index extrema; (b) Spin Image classification; (c) Consistent landmark sets; (d) Best landmark set; (e) Spin Image templates; (f) Landmark model.*

- Variations of each Active Landmark Model are calculated using PCA.

For each facial dataset the landmark detection and region retrieval procedure has the following steps (Fig. 2):

1. Extract candidate landmarks from the Shape Index map.
2. Classify candidate landmarks by matching them with corresponding Spin Image templates.
3. Compute the rigid transformation that best aligns combinations of 8 or 5 candidate face landmarks with the corresponding Mean Shape.
4. Select the best combination of candidate landmarks (based on the minimum Procrustes distance).
5. Discard combinations of candidate landmark sets that are not consistent with the ALMs.
6. Clip and retrieve the facial regions of interest.

3.1. The Landmark Mean Shape

According to Dryden [DM98], "a *landmark* is a point of correspondence on each object that matches between and within populations of the same class of objects" and "*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. 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 $D^2 = \sum |\mathbf{x_i} - \mathbf{x_m}|^2$ of each shape ($\mathbf{x_i}$) to the mean ($\mathbf{x_m}$). The alignment procedure is commonly known as Procrustes Analysis [DM98, SG02], and is used to calculate the Mean Shape of landmark shapes (Fig. 3). The Mean Shape is the Procrustes mean: $\mathbf{x_m} = \frac{1}{N} \sum \mathbf{x_i}$ for all example shapes $\mathbf{x_i}$ after alignment.

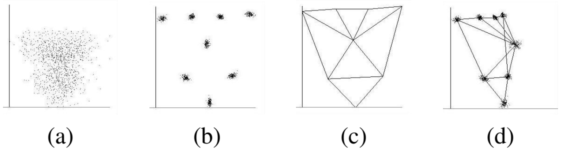


Figure 3: *Landmark Mean Shape estimation: (a) Unaligned Landmarks; (b) Aligned Landmarks; (c) Landmark Mean Shape; (d) Landmark Cloud & Mean Shape rotated* 60° *around the y-axis.*

3.2. 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 to the aligned shapes.

Since PCA is a linear procedure, all aligned shapes are first projected to the tangent space of the Mean Shape. The tangent space projection linearizes shapes by scaling them with a factor α: $\mathbf{x_t} = \alpha \mathbf{x} = \frac{|\mathbf{x_m}|^2}{\mathbf{x_m} \cdot \mathbf{x}} \mathbf{x}$

Aligned shape vectors form a distribution in the *nd* dimensional shape space, where n is the number of landmarks and d the dimension of each landmark. We can model this distribution by estimating a vector $\mathbf{b}$ of parameters that describes a shape's deformations. The approach according to Cootes & Taylor [CT01, CTKP05] is as follows:

1. Determine the mean shape.
2. Determine the covariance matrix of the shape vectors.
3. Compute the eigenvectors $\mathbf{A_i}$, and corresponding eigenvalues λ_i of the covariance matrix, sorted in descending order.

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

The vector $\mathbf{b}$ is the projection of $\mathbf{x}$ onto the subspace spanned by the p most significant eigenvectors of the eigenspace (*principal components*). By selecting the p largest eigenvalues, the mean square error between $\mathbf{x}$ and its approximation $\mathbf{x}'$ is minimum. Thus, the *Active Landmark Model* (ALM) is created [CTCG95, CT01, CTKP05].

By varying $\mathbf{b}$ we can create shape variations. By applying limits to each b_i ($|b_i| \leq 3\sqrt{\lambda_i}$) we can create marginal mean shape deformations. The number p of eigenvectors and eigenvalues to retain (*modes of variations*), can be chosen so that the model represents a given proportion of the total variance of the data, i.e., the sum V_t of all the eigenvalues.

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

The factor f represents the percentage of total variance incorporated into the ALM. For our purposes, a factor of 99% was chosen.

3.3. Fitting Landmarks to the Model

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

Fitting a set of points $\mathbf{y}$ to the ALM $\mathbf{x}$ is done by minimizing the Procrustes distance in a simple iterative approach, adapted from Cootes & Taylor [CT01]:

1. Translate $\mathbf{y}$ so that its centroid is at the origin (0,0,0).
2. Repeat an iterative procedure that aligns $\mathbf{y}$ to the mean shape $\mathbf{x_m}$ until convergence (Procrustes distance doesn't change much).
3. Project $\mathbf{y}$ onto the tangent space of $\mathbf{x_m}$.
4. Determine the model deformation parameters $\mathbf{b}$ that match $\mathbf{x_m}$ to $\mathbf{y}$: $\mathbf{b} = \mathbf{A}^T \cdot (\mathbf{y} - \mathbf{x_m})$
5. Accept $\mathbf{y}$ as a member of the shape's class if $\mathbf{b}$ satisfies certain constraints.

We consider a landmark shape as plausible if it is consistent with marginal shape deformations [CT01, CTKP05].

3.4. Landmark Detection & Selection

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:

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

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

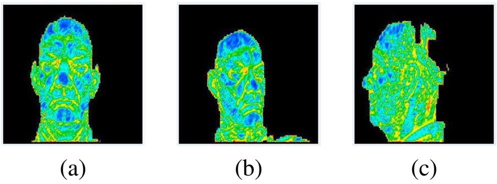

(a) (b) (c)

Figure 4: *Shape Index maps: (a) frontal face dataset; (b) 45° side face dataset; (c) 60° side face dataset. Blue denotes Caps, green Saddle, and red Cups.*

After calculating Shape Index values on a 3D facial dataset, a mapping to 2D space is performed (using the native UV parameterization of the facial scan), in order to create a *Shape Index map* (Fig. 4). Local maxima and minima are identified on the Shape Index map. Local maxima (Cap=1.0) are candidate landmarks for nose tips and chin tips and local minima (Cup=0.0) for eye corners and mouth corners. The Shape Index's 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. In Fig. 6(a), black boxes represent Caps, and white boxes Cups.

However, experimentation showed that the Shape Index alone is not robust enough for detecting anatomical landmarks in facial datasets in a variety of poses. Thus, candidate landmarks estimated from Shape Index values are classified and filtered out according to their relevance with corresponding Spin Image templates.

A *Spin Image* encodes the coordinates of points on the surface of a 3D object with respect to a 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 description 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 local shape spanned by a cylinder of 3.2 *cm* height and 3.2 *cm* diameter.

Thus, local maxima and minima of the Shape Index map (Caps and Cups) are further classified into 5 classes (eye outer corner, eye inner corner, nose tip, mouth corner and chin tip) according to the relevance of their Spin Image grids with 5 Spin Image templates (Fig. 5), that represent each landmark class.

Notice that due to the symmetry of the face, landmark points cannot be further distinguished to left and right. Relevance is estimated according to a similarity measure between two Spin Image grids P and Q, which is expressed by the

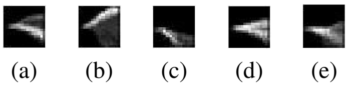


Figure 5: *Spin Image templates: (a) eye outer corner; (b) eye inner corner; (c) nose tip; (d) mouth corner; (e) chin tip.*

normalized linear correlation coefficient:

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

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

Each of the 5 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 anatomical landmark class is retained. In Fig. 6(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 landmarks are overlapped.

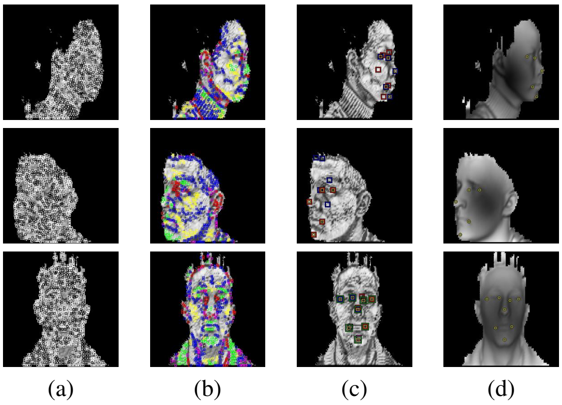


Figure 6: *Results of landmark detection and selection process: (a) Shape Index's Maxima & Minima; (b) Spin Image classification; (c) Extracted Best Landmark Sets; and (d) Resulting Landmarks.*

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 length constraints from the shape model and its deformations are used to reduce the search space (pruning). From all the feasible candidate 5 landmark sets, the ones that do not conform with neither ALM5L nor ALM5R are filtered out. This is done by applying the fitting procedure as described in Section 3.3.

The final step is to fuse them in complete landmark sets of 8 landmarks that conform with the ALM8. From the three available sets (ALM5R, ALM5L, ALM8), the one that has the minimum Procrustes distance to the corresponding model is considered the final solution. In Fig. 6(c), blue boxes represent landmark sets consistent with the ALM5R, red boxes with the ALM5L, green boxes with the ALM8, and yellow boxes the best landmark set. Notice that some of the consistent landmarks are overlapped. Also note that the ALM8 consistent landmark set is not always the best solution, ALM5L and ALM5R are usually the best solutions for side facial datasets (Fig. 6(d)). Finally, using the best solution, the pose is estimated, and the facial dataset is classified as frontal, left side or right side (based on the rotation angle along the vertical axis).

3.5. Facial Region Retrieval

After landmark localization and pose estimation is performed, the facial region of interest must be automatically retrieved. In its simplest form this involves a 3D clipping procedure against a suitable clipping volume (Fig. 7) [TPPP08].

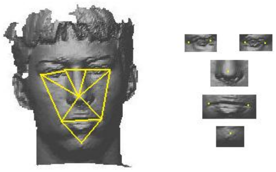


Figure 7: *Face segmentation for region retrieval.*

The position, orientation and size of the clipping volume is determined according to the landmark position within the facial region that must be retrieved. More elaborate methods can also be applied to better segment the facial region of interest; for example a curvature based vertex walk within certain distance limits could produce the nasal region more accurately. Such heuristic techniques must generally be developed on a custom basis per region of interest.

4. Results

4.1. 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]. It contains a total of 4007 range images, acquired between 2003 and 2004. The hardware used to acquire these range data was a Minolta Vivid 900 laser range scanner, with a resolution of 640 × 480. These data were obtained from 466 subjects and contain various facial expressions (e.g., happiness, surprise).

For the purposes of this evaluation we manually annotated 975 frontal facial datasets randomly selected from the FRGC v2 database, containing several subjects with various facial expressions. This database will be referred as **DB00F**.

For side facial datasets, we used 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). Note that even 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 115 left and 115 right 45° side datasets. These databases will be referred as **DB45L** and **DB45R** respectively. We also annotated 80 left and 80 right 60° side datasets. These databases will be referred as **DB60L** and **DB60R** respectively.

In the evaluation dataset, 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.

4.2. Performance Evaluation

In all experiments the distance error between the manually annotated landmarks and the automatically detected landmarks was calculated. Also, the mean error of the 8 landmark points for the frontal datasets and of the 5 landmark points of the side datasets was computed. In this paper we depict the Error Distribution histograms only for the mean error (Figs. 8, 9, 10, 11, 12). In these histograms 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 in a certain interval (error probability distribution).

Distance error analysis was carried out only on results where the pose of the probe was correctly estimated (Tables 1, 2, 3, 4, 5). The pose estimation rate is the percentage of correct pose estimations of the probe (frontal, left profile, right profile), so that the set of landmarks is the same.

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

Table 1: *Results for* ***DB00F***

Correct Pose Estimation	966 / 975	99.08%
Distance Error	mean (mm)	std.dev. (mm)
Right Eye Outer Corner	8.13	3.79
Right Eye Inner Corner	7.02	3.18
Left Eye Inner Corner	7.46	3.07
Left Eye Outer Corner	9.21	4.25
Nose Tip	5.23	3.28
Mouth Right Corner	8.30	4.53
Mouth Left Corner	9.83	4.47
Chin Tip	6.71	4.32
Mean Distance Error	7.74	2.05

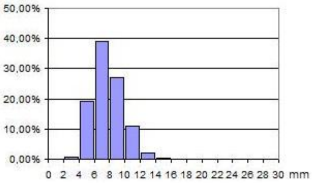


Figure 8: *Mean Error (mm) Distr. for* ***DB00F*** *(x-axis is the mean distance-error in mm between ground truth and automatically detected landmarks. y-axis is the percentage of facial datasets which have errors in each interval).*

Landmark detection and localization results for **DB00F** are shown in Table 1 and Fig. 8. Note that 966 out of 975 tested facial scans were correctly detected as frontal (99.08%). From the Error Distribution histogram we can observe that in 86.1% of the 966 face scans with correct pose detection, landmark positions have a mean error up to 10.0 *mm*. The mean error is (7.74 ± 2.05) *mm*. The best located facial feature is the nose tip with error (5.23 ± 3.28) *mm*, and worst is the mouth left corner with error (9.83 ± 4.47) *mm*.

Table 2: *Results for* ***DB45R***

Correct Pose Estimation	114 / 115	99.13%
Distance Error	mean (mm)	std.dev. (mm)
Right Eye Outer Corner	10.83	16.98
Right Eye Inner Corner	10.44	17.22
Nose Tip	9.93	18.11
Mouth Right Corner	11.09	17.79
Chin Tip	10.42	16.84
Mean Distance Error	10.54	16.84

Landmark detection and localization results for **DB45R** are shown in Table 2 and Fig. 9. Note that 114 out of 115 tested facial scans were correctly detected as right scans

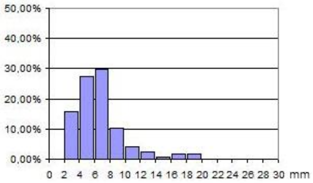


Figure 9: *Mean Error (mm) Distr. for* ***DB45R****.*

(99.13%). From the Error Distribution histogram we can observe that in 83.3% of the 114 face scans with correct pose detection, landmark positions have a mean error up to 10.0 *mm*. The mean error is (10.54 ± 16.84) *mm*. The best located facial feature is the nose tip with error (9.93 ± 18.11) *mm*, and worst is the mouth right corner with error (11.09 ± 17.79) *mm*.

Table 3: *Results for* ***DB45L***

Correct Pose Estimation	115 / 115	100.00%
Distance Error	mean (mm)	std.dev. (mm)
Left Eye Outer Corner	9.73	11.69
Left Eye Inner Corner	9.17	12.67
Nose Tip	8.29	13.18
Mouth Left Corner	11.03	13.83
Chin Tip	10.44	12.14
Mean Distance Error	9.73	12.09

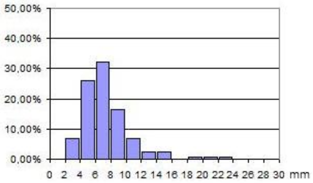


Figure 10: *Mean Error (mm) Distr. for* ***DB45L****.*

Landmark detection and localization results for **DB45L** are shown in Table 3 and Fig. 10. Note that 115 out of 115 tested facial scans were correctly detected as left scans (100.00%). From the Error Distribution histogram we can observe that in 81.8% of the 115 face scans with correct pose detection, landmark positions have a mean error up to 10.0 *mm*. The mean error is (9.73 ± 12.09) *mm*. The best located facial feature is the nose tip with error (8.29 ± 13.18) *mm*, and the worst is the mouth left corner with error (11.03 ± 13.83) *mm*.

Landmark detection and localization results for **DB60R** are shown in Table 4 and Fig. 11. Note that 78 out of 80 tested facial scans were correctly detected as right scans (97.50%). From the Error Distribution histogram we can observe that in 70.5% of the 78 face scans with correct pose detection, landmark positions have a mean error up to 10.0 *mm*. The mean error is (10.16 ± 9.14) *mm*. The best located facial feature is the nose tip with error (8.73 ± 10.35) *mm*, and the worst is the mouth right corner with error (11.51 ± 10.65) *mm*.

Table 4: *Results for* ***DB60R***

Correct Pose Estimation	78 / 80	97.50%
Distance Error	mean (mm)	std.dev. (mm)
Right Eye Outer Corner	10.67	11.06
Right Eye Inner Corner	10.57	10.17
Nose Tip	8.73	10.35
Mouth Right Corner	11.51	10.65
Chin Tip	9.32	8.77
Mean Distance Error	10.16	9.14

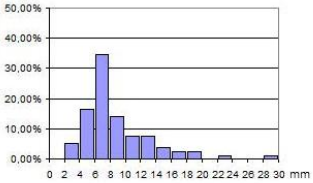


Figure 11: *Mean Error (mm) Distr. for* ***DB60R****.*

Table 5: *Results for* ***DB60L***

Correct Pose Estimation	77 / 80	96.25%
Distance Error	mean (mm)	std.dev. (mm)
Left Eye Outer Corner	11.58	16.43
Left Eye Inner Corner	9.60	18.55
Nose Tip	9.57	18.52
Mouth Left Corner	11.43	18.02
Chin Tip	10.76	16.08
Mean Distance Error	10.59	17.06

Landmark detection and localization results for **DB60L** are shown in Table 5 and Fig. 12. Note that 77 out of 80 tested facial scans were correctly detected as left scans (96.25%). From the Error Distribution histogram we can observe that in 80.5% of the 77 face scans with correct pose detection, landmark positions have a mean error up to 10.0 *mm*. The mean error is (10.59 ± 17.06) *mm*. The best located facial feature is the nose tip with error (9.57 ± 18.52) *mm*, and the worst is the left eye outer corner with error (11.58 ± 16.43) *mm*.

In general, we can observe larger mean errors with larger standard deviations in side scan results than in frontal scans. This is due to the fact that the problem of landmark detection in side scans is more difficult and the tested data set was much smaller, so outliers had a more significant effect on results. The mean error is under 10.6 *mm* on all tested facial scans, although the deviation is much larger on side scans.

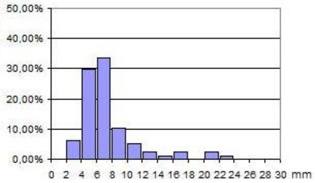


Figure 12: *Mean Error (mm) Distr. for* ***DB60L****.*

The most robust facial feature is the nose tip, with a mean error under 10.0 *mm* on all tested facial scans. Contrarily, the least robust facial feature appears to be the mouth corners, mainly due to the fact that expressions change the positions of these landmarks significantly. Finally note that pose was correctly estimated on over 96% of the tested facial scans, irrespectively of pose. Specifically, the best results have obtained for the frontal facial scans and the worst, for the 60° left facial scans. Extensive testing shows that the facial region retrieval procedure can tolerate the above errors.

5. Conclusion

We have presented an automatic 3D facial region retrieval method that offers pose invariance, with respect to large pose variations. The proposed method introduced a new method for 3D landmark localization and retrieval of facial regions of interest. 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, pose estimation and region of interest retrieval) work robustly, even if half of the face is missing.

Future work will be directed towards increasing the robustness of the landmark detector. This can be accomplished in three ways: firstly by increasing the examples dataset for creating the landmark model, secondly by increasing the landmark set, and thirdly by selecting the Spin Image templates from statistically trained Spin Image grids.

References

[CCRA*05] CONDE C., CIPOLLA R., RODRIGEZ-AROGON L. J., SERRANO A., CABELLO E.: 3D facial feature location with spin images. In *Proc. IAPR Conference on Machine Vision Applications* (2005).

[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* (2005).

[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* (1995).

[CTKP05] COOTES T., TAYLOR C., KANG H., PETROVIC V.: *Handbook of Face Recognition*. Springer, 2005, ch. Modeling Facial Shape and Appearance, pp. 39–63.

[Dib06] DIBEKLIOGLU H.: *Part-Based 3D Face Recognition under Pose and Expression Variations*. Master's thesis, Yeditepe University, 2006.

[DM98] DRYDEN I., MARDIA K.: *Statistical Shape Analysis*. Wiley, 1998.

[Joh97] JOHNSON A. E.: *Spin Images: A Representation for 3D Surface Matching*. PhD thesis, Carnegie Mellon University, 1997.

[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. Int. Conf. on Automatic Face and Gesture Recognition* (2006).

[LJC06] LU X., JAIN A., COLBRY D.: Matching 2.5D face scans to 3D models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2006).

[LSCH06] LIN T., SHIH W., CHEN W., HO W.: 3D face authentication by mutual coupled 3D and 2D feature extraction. *ACM SE* (2006).

[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* (2005).

[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* (2007).

[TPPP08] THEOHARIS T., PAPAIOANNOU G., PLATIS N., PARTIKALAKIS N.: *Graphics & Visualization: Principles and Algorithms*. A K Peters, 2008.

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

[WLY07] WEI X., LONGO P., YIN L.: *Automatic Facial Pose Determination of 3D Range Data for Face Model and Expression Identification*. ICB 2007, LNCS, Springer, 2007.