3D Human Face Modelling: Soft Tissues Landmarking

A. Zompi (a), S., Tornincasa (b), E., Vezzetti (b), S., Moos (a), M. G., Violante (a), F., Marcolin (a), D., Speranza (b)

(a) Politecnico di Torino, Italy
(b) Università di Cassino, Italy

Article Information

Keywords:
3D face, landmarking, soft-tissue landmark, landmark extraction, differential geometry.

Corresponding author:
Enrico Vezzetti
e-mail: enrico.vezzetti@polito.it
Address: Politecnico di Torino, corso Duca degli Abruzzi 24, 10129 Torino, Italy

Abstract

Purpose:
The study proposes a structured methodology for soft-tissues landmarks formalization in order to provide a methodology for their automatic identification.

Method:
Working in the Differential Geometry domain, through the Coefficients of the Fundamental Forms, the Principal Curvatures, Mean and Gaussian Curvatures and also with the derivatives and the Shape and Curvedness Indexes, a Matlab® algorithm for extracting nine landmarks is built. Each landmark is unique for the behaviour of these geometrical descriptors in correspondence of its zone of interest. This particular behaviour of the descriptor is used to localize each landmark and for extracting it.

Result:
Nine landmarks were extracted for each of the seventy-nine faces of our database. Results were shown to an expert, a maxillofacial surgeon, who confirmed their correctness. So, the Matlab® algorithm elaborated for the landmarks extraction correctly works for all of them.

Discussion & Conclusion:
This research provided a geometric profile of each landmark which has noticeable feature, gives no identification problem and is far from the areas of face that are more subject and influenced by facial expressions, such as the mouth. The fact is that these points were not only accurately describable, but even suitable to a geometrical description.

1 Introduction

Three-dimensional face recognition is a research field born in the last three decades. It gradually developed and grew into an industry, that specialized companies care for. Nowadays the aim is to refine and improve known recognition algorithms, or automate new procedures so that Police and other safety organizations could use it as an official means for the recognition of suspects and public enemies. This is not the only application field: authentication and aesthetic surgery are other two sectors which have to deal with digital 3D information of human face.

Recently, one of the methods used for studying faces is the one involving the extraction of landmarks. A landmark is a point which all the faces share and which has a particular biological meaning. In human face fifty-nine landmarks could be collected, but the most famous ones are nearly twenty. Some facial recognition algorithms identify faces by extracting landmarks, or features, from an image of the subject’s face. For example, an algorithm may analyze the relative position, size, and/or shape of the eyes, nose, cheekbones, and jaw. Then, these features are used to search for other images with matching features.

Much research has been carried out on this topic. In their various publications, Alker et al., Frantz et al., and Wörz et al. proposed multi-step differential procedures for subvoxel localization of 3D point landmarks, addressing the problem of choosing an optimal size for a region-of-interest (ROI) around point landmarks [9][10]. They introduced an approach for the localization of 3D anatomical point landmarks based on deformable models. To model the surface as a landmark, they used quadric surfaces combined with global deformations [11][1]. They proposed a method based on 3D parametric intensity models, which are directly fitted to 3D images, and introduced an analytic intensity model based on the Gaussian error function in conjunction with 3D rigid transformations and deformations in order to efficiently model anatomical structures [21]. Finally these researchers introduced a new multi-step approach to improve detection of 3D anatomical point landmarks in tomographic images [12].

Romero et al. presented a comparison of several approaches that use graph matching and cascade filtering for landmark localization in 3D face data. For the first method, they apply the structural graph matching algorithm relaxation-by-elimination using a simple distance-to-local-plane node property and a Euclidean-distance arc property. After the graph matching process has eliminated unlikely candidates, the most likely triplet is selected, by exhaustive search, as the minimum Mahalanobis distance over a six dimensional space, which corresponds to three node variables and three arc variables. A second method uses state-of-the-art pose-invariant feature descriptors embedded into a cascade filter to localize the nose tip. After that, local graph matching is applied to localize the inner eye corners [16]. Then, it describes and evaluates their pose-invariant pointpair descriptors, which encode 3D shape between a pair of 3D points. Two variants of descriptor are introduced: the first one is the point-pair spin image, which is related to the classical spin image of Johnson and Hebert, and the second one is derived from an implicit...
radial basis function (RBF) model of the facial surface. These descriptors can really encode edges in graph based representations of 3D shapes. Here they show how the descriptors are able to identify the nose-tip and the eye-corner of a human face simultaneously in six promising landmark localisation systems [17].

Ruiz et al. [16] presented an algorithm for automatic localization of landmarks on 3D faces. An Active Shape Model (ASM) is used as a statistical joint location model for configurations of facial features. The ASM is adapted to individual faces through a guided search, whereby landmark specific Shape Index models are matched to local surface patches. Similarly, Sang-Jun et al. [20] applied the Active Shape Models to extract the position of the eyes, the nose and the mouth.

Salah et al. [19] proposed a coarse-to-fine method for facial landmark localization: it relies on unsupervised modelling of landmark features obtained through different Gabor filter channels.

D’Hose et al. [7] presented a method for localization of landmarks on 3D faces: it uses Gabor wavelets to extract the curvature of the 3D faces, which is then used for performing a coarse detection of landmarks.

Although geometry was widely used for general shape analysis not involving faces [15], there are a few works concerning facial animation which deal with this task and use geometry in order to describe face shape. Ansari et al. [2] presented a fully automated algorithm for facial feature extraction and 3D face modelling from a pair of orthogonal frontal and profile view images of a person’s face taken by calibrated cameras. The algorithm starts by automatically extracting corresponding 2D landmark facial features from both view images, then compute their 3D coordinates. Further, they estimated the coordinates of the features hidden in the profile view and this work based on the visible features extracted in the two orthogonal face images. The 3D coordinates of the selected feature points (obtained from the images) are used to align and then to locally deform the corresponding facial vertices of the generic 3D model. Bickel et al. [4] presented a new method for real-time animation of highly-detailed facial expressions. It is based on a multi-scale decomposition of facial geometry into large-scale motion and fine-scale details, such as expression wrinkles. Their hybrid animation is tailored to the specific characteristics of large- and fine-scale facial deformations. This method performs real-time animation of highly-detailed faces with realistic wrinkle formation, and allows both large-scale deformations and fine-scale wrinkles to be intuitively edited. Furthermore, the resulting pose-space representation enables the transfer of facial details to new expressions or other facial models. Blanz et al. [5] introduced a new technique for modelling textured 3D faces. Starting from an example set of 3D face models, they derived a morphable face model by transforming the shape and texture of the examples into a vector space representation. New faces and expressions can be modelled by forming linear combinations of the prototypes. Shape and texture constraints derived from the statistics of the example faces are used to guide manual modelling or automated matching algorithms. They showed 3D face reconstructions from single images and their applications for photo-realistic image manipulations. They also demonstrated face manipulations according to complex parameters such as gender, fullness of a face or its distinctiveness. Berrett et al. [3] addressed the problem of person-independent facial expression recognition using the 3D geometry information extracted from the 3D shape of the face. To this end, a completely automatic approach is proposed: it relies on identifying a set of facial keypoints and on computing SIFT feature descriptors of face depth images around sample points defined starting from the facial keypoints. It is also based on selecting the subset of features with maximum relevance.

Previous works show how a face description is possible with the use of different algorithms, mathematical operators, models and descriptors. What is really missing in this framework is a deep geometrical description of facial features, especially landmarks. This is exactly the aim of this paper: to formalize the geometry of the anatomical landmarks, using derivatives, Shape and Curvedness Indexes, mean and Gaussian curvatures and geometrical descriptors such as e, f, g, E, F and G. The purpose is mainly to show how a quite efficient landmark localization is possible only through geometry, or at least a precise identification of the zone of interest which the landmark lies in. Although it is a heuristic study, this geometrical formalization may be important for supporting many applications, in particular in medical field, face recognition, face detection, authentication, and facial expression recognition.

## 2 Proposed method

As previously said, a facial landmark is a point which all faces share and has a particular biological meaning. Hard-tissue landmarks lie on the skeletal and may be identified only through lateral cephalometric radiographs; soft-tissue landmarks are on the skin and can be identified on the point clouds generated by the scanning. In fact, this study only has to deal with soft-tissue landmarks. Although soft-tissue landmarks are nearly fifty-nine, this research only considered nine identifiable ones. The landmarks close to the mouth are not taken into consideration, because they are more influenced by the pose. The ones near the boundaries of the face have been ignored because in those zones the scan is not accurate. A large set of landmarks and the set used here are shown in Figure 1.

**Figure 1. Anthropometric soft- tissue landmarks:** g-glabella, n-nasion, en-endocanthion, ex-exocanthion, or-orbital, prn-pronasal, sn-subnasal, al-alae, ch-choiliion, pg-pogonion, gn-gnathion, go-gonion, me-menton [6].

Landmarks are used to study the face: through their localization a face description is possible. Considering the morphological features of the face in order to extract the landmarks, it is necessary to employ a refining procedure, that first identifies the region and collects the significant points and, at the end, it extracts the specific landmarks. Considering the different peculiarities of each facial region where the landmarks are located, the different
combination of the first, the second and the mixed derivatives, the Coefficients of the Fundamental Forms \( E, F, G, e, f \) and \( g \), the mean, Gaussian, and principal curvatures \( K, H, k_1, k_2 \) [8-13], and Shape and Curvedness Indexes \( S \) and \( C \) introduced by Koenderink and van Doorn [14] as descriptors have been employed.

Particularly, the identification of the zone of interest, i.e. the localization, was generally performed selecting the interest range of the Shape Index: the critical points need to be selected, if the point at issue is one of them, then is reduced the range of one or more descriptors. During these steps, the neighbourhood of the landmark, namely the zone of interest, gets more and more narrow. The identification of the right Shape Index narrows the zone; then the other chosen descriptors, one at a time, gradually narrows the area even more, so that the final landmark neighbourhood is very little and close around it. All the deductions of these behaviours were done theoretically, then they were tested on some samples belonging to the set of faces, used for the final application of the algorithm. If it worked correctly, the procedure was applied to all the faces employed for this study. The extraction of the landmark was obtained maximizing or minimizing one appropriate descriptor, generally a Coefficient of the Fundamental Forms. This is due to the fact that these Coefficients have “minimum and maximum behaviours” in correspondence of the points of interest. This is a general procedure: some of the algorithms explained below may differ a little from this general scheme, as appropriate. A representation diagram of the general process is here reported.

![Localization Diagram](Image)

In order to obtain a final correct extraction algorithm for every landmark, a deep knowledge of differential geometry was needed. Nevertheless, a portion of experimental inference was necessary. The geometrical behaviour of the descriptors was almost predictable but the choice of the ones to be really used in the landmark extraction was not so easy. In particular, the final choice of only some descriptors as means for the extraction is the result of a long sample testing work.

A landmark is here identified as a point of the grid with a particular biological and geometrical meaning. It is described uniquely by its coordinates in a three-dimensional Cartesian reference system: in the figures presented below usually \( x \)-axis is vertical, \( y \) is horizontal and \( z \) enters the sheet.

A geometrical formalization of each landmark, one at a time, is presented below. The localization and extraction processes used in the algorithm are explained and graphically represented. Similarly to the scheme above, the white squares represent the steps of the localization process, while the grey ones are the extraction of the landmark.

### 2.1 Pronasal

The pronasal (PN) is the point on the nose tip. It is surely the point most easily identifiable by human eye, especially because it is the most salient, when the face is well oriented. In fact, once found, it is possible to verify the reliability of the coordinates through a comparison with the coordinates of the point with the lowest value of \( z \).

- 1) it belongs to the points whose Shape Index lies in the range corresponding to the surface of dome;
- 2) in our reference system it is an absolute minimum, so it is a critical point;
- 3) the coefficient \( e \) has a local (absolute, most of the times) maximum in it;
- 4) the coefficient \( g \) has a local (absolute, most of the times) maximum in it;
- 5) the Gaussian curvature \( K \) has a local (absolute, most of the times) maximum in it;
- 6) the mean curvature \( H \) has a local (absolute, most of the times) minimum in it.

From the theory it is deducible that conditions 1 and 2 are always true, and at least three of the conditions 3, 4, 5 and 6 hold, namely identify the same point as PN. Actually, conditions 5 and 6 are equivalent, i.e. they always give the same result. Particularly, conditions 1 and 2 are used to identify the area, then the maximums of \( e, g \) and \( K \) (or, equivalently, the minimum of \( H \)) are computed to extract the points:

- if the coordinates of PN obtained maximizing \( e \) (condition 3) coincide with the ones obtained maximizing \( g \) or \( K \) both, then the pronasal is the point got from conditions 3, 4, 5 (and 6);
- if conditions 4 and 5 give the same result, then the pronasal is the point obtained from condition 5 (or 6);
- if conditions 4 and 5 give different results, then the pronasal is given by condition 3, namely maximizing \( e \).

The steps of the process are explained in the graph below.

![Localization Diagram](Image)

\[ \text{Select the points whose Shape Index lies in the range corresponding to the surface of dome: } S \in [0.025,1] \]

\[ \text{Select the critical points: the points whose first derivatives with respect to } x \text{ and } y \text{ is approximately equal to zero} \]

\[ \text{If the 2 results coincide: } PN \text{ is the point obtained with all the conditions} \]

\[ \text{If } \max(g) = \max(K) \text{ then PN is the point obtained maximizing } K \]

\[ \text{If } \max(g) \neq \max(K) \text{ then PN is the point obtained maximizing } e \]

### 2.2 Subnasal

The subnasal (SN) is the point which lies exactly below the nose, in that little dimple above the mouth. It has been
studied in a neighborhood of the pronasal. These are its geometrical features:
1) it belongs to the points whose Shape Index lies in the range corresponding to the surfaces of saddle point, saddle ridge, saddle rut and rut;
2) the curvedness index $C$ has a high value in its area; it is a critical point;
3) the coefficient $e$ has a local maximum in it;
4) the coefficient $g$ has a local (absolute, most of the times) minimum in it;
5) the Gaussian curvature $K$ has a local minimum in it;
6) the mean curvature $H$ has a local maximum in it;
7) the coefficient $F$ is close to zero in its area.

2.3 Alae
The alae (AL) are the two points which lie on the left and the right of the widest part of the nose; therefore their distance is exactly the nose width. These are their geometrical features:
1) they belong to the points whose Shape Index lies in the range corresponding to the surface of ridge;
2) the curvedness index $C$ has a high value in their areas;
3) the derivative of $z$ with respect to $x$ $D_x$ is positive on the right ala (from an external point of view) and negative on the left;
4) the second derivative of $z$ with respect to $x$ has a local minimum in them;
5) the coefficient $E$ has a two local maximums in them;
6) the coefficient $f$ is positive on the right ala and negative on the left.

The elaborated algorithm identifies the two areas of interest through conditions 1, 2 and 3, then it extracts the landmarks using condition 5, namely maximizing $E$. Condition 6 is useless. The steps of the process are explained in the scheme below.

2.4 Endocanthions
The endocanthions (EN) are the two points at which the inner ends of the upper and lower eyelid meet. These are their geometrical features:
1) they belong to the points whose Shape Index lies in the range corresponding to the surfaces of cup and rut;
2) in our reference system they are local maximums, so they are critical points;
3) the derivative of $z$ with respect to $x$ $D_x$ is positive on the right endocanthion and negative on the left one;
4) the second derivative of $z$ with respect to $x$ $D_{xx}$ is negative in them;
5) the derivative of $z$ with respect to $y$ $D_y$ is negative in them;
6) the second derivative of $z$ with respect to $y$ $D_{yy}$ is negative in them;
7) the coefficient $e$ has an hollow in them;
8) the coefficient $f$ is negative on the right EN and positive on the left;
9) the coefficient $F$ is negative on the right EN and positive on the left;
10) the mean curvature $H$ is positive in them.

Conditions 1, 2, 3, 4, 8 and 9 were used to localize the areas, while the landmarks were extracted using minimizing $e$, namely using condition 7. The steps of the process are explained in the graph below.

2.5 Exocanthions
The exocanthions (EX) are the two points at which the outer ends of the upper and lower eyelid meet. These are their geometrical features:
1) they belong to the points whose Shape Index lies in the range corresponding to the surface of ridge;
2) the derivative of $z$ with respect to $x$ $D_x$ is positive on the right exocanthion and negative on the left one;
3) the second derivative of $z$ with respect to $y$ $D_{yy}$ is positive in them;
4) the coefficient $g$ is positive in them;
5) the coefficient $E$ has an increasing behaviour from the inside out of the face;
6) the coefficient $F$ is negative on the right EX and positive on the left one;
7) the Gaussian curvature $K$ is positive or equal to zero in them;
8) the mean curvature $H$ is negative in them.

Conditions 1, 2, 3, 4, 6 and 7 were used to localize the areas. Then the landmarks were extracted maximizing the coefficient $E$, i.e. using condition 5. The steps of the process are explained in the graph below.

### 2.6 Nasion

The **nasion** (N) is a point at the top of the nose, nearly between the eyes. In the horizontal direction, it lies on the high part of the nose bone; in the vertical direction it is in the hollow under the forehead. For these reasons it is a saddle point and a critical point. These are its geometrical features:

1) it belongs to the points whose Shape Index lies in the range corresponding to the surface of **ridge**, **saddle ridge**, **saddle point** or **saddle rut**;
2) in our reference system it is a local maximum in the $x$ direction and a local minimum in the $y$ direction, so it is a saddle point and thus a critical point;
3) the coefficient $g$ has a local minimum in it;
4) the Gaussian curvature $K$ is negative or equal to zero in it.

Conditions 1, 2 and 4 have been used to localize the area, while the landmark was extracted minimizing $g$, namely using condition 3. The steps of the process are explained in the graph below.

### 3 Results

A Matlab® algorithm for landmark extraction was elaborated and implemented. It is totally based on the behaviour of these geometrical descriptors among faces. Seventy-nine scanned faces (obtained by using Cyberware Head and Face Colour Scanner [22]) of twenty persons with different facial expressions were used for the experimentation. They are useful to check if the descriptors and, consequently, the elaborated process are strongly connected to the expression or if they are stable between different persons and poses. The scanned people were all Caucasian, male and female, from 20 to 40 years old. The computing time of the algorithm is 6 seconds for each face shell, namely this is the time required for locating the nine landmarks for each face.

The face data were collected in point clouds through the 3D scanner. These shells were then imported in Matlab® and triangulated. The triangular mesh was converted into a square grid through the known function “gridtrimesh”. On this mesh the algorithm was elaborated and could run.

The resulting nine landmarks of four of the seventy-nine faces are shown in Figure 2.

![Figure 2. The extracted landmarks for four of the seventy-nine faces.](image-url)
arrangement of facial soft-tissues is important for medical evaluations: orthodontists, orthognathic maxillofacial surgeons, and plastic surgeons all require quantitative data about the soft-tissues which complete the evaluation of hard-tissue relations. Whereas skeletal structures can be assessed only with radiographic instruments, the arrangement of soft-tissues can be established by several non-invasive methods. Orthognathic maxillofacial surgeons have used computed anthropometry to assess patients both before and after treatment. At present, several software systems allow clinicians to manipulate digital representations of hard and soft-tissue profile tracings and subsequently to modify the pre-treatment image in order to produce a treatment simulation [2]. Surgeons generally identify these landmarks by touching the patient’s face. But their experience of these points allow them to trace them immediately even on facial images, without any direct contact with the face in exam.

The landmarks in all the faces, although the area of interest endocanthion was not as accurate as that of all the other landmarks, the surgeon identified. The pronasal, the subnasal and the nasion are surely the best positioned landmarks. The values show that the position of the landmarks is im portant for medical evaluations: orthodontists, orthognathic maxillofacial surgeons, and plastic surgeons all require quantitative data about the soft-tissues which complete the evaluation of hard-tissue relations. Whereas skeletal structures can be assessed only with radiographic instruments, the arrangement of soft-tissues can be established by several non-invasive methods. Orthognathic maxillofacial surgeons have used computed anthropometry to assess patients both before and after treatment. At present, several software systems allow clinicians to manipulate digital representations of hard and soft-tissue profile tracings and subsequently to modify the pre-treatment image in order to produce a treatment simulation [2].

The localization seems to be quite accurate on face soft-tissues. The values show that the position of the endocanthions was not as accurate as that of all the other landmarks in all the faces, although the area of interest was correctly identified. The pronasal, the subnasal and the nasion are surely the best positioned landmarks: their sample mean and variance values are approximately equal to zero. The distances between the landmarks are partially equal to zero. The ones with the highest values are recorded for the endocanthions, reaching a maximum of 7.9, although their areas of interest was correctly localized.

To improve the analysis of the results and finally state whether they are acceptable or not, it was asked to the surgeon to define the size of the landmarks neighbourhood that could be considered acceptable. More in detail, two zones of interest were identified, one closer to the point, more precise, and one a bit wider. The surgeon defined these neighbourhoods in millimetres; we re-scaled the dimensions on the points of our meshed shells and used these information for splitting up the distances between obtained and theoretical point into four ranges, so that we were able to define four categories of distances:

- **d=0**: if the distance between the landmark identified by the surgeon and the one obtained by our algorithm was exactly equal to zero, namely the point was perfectly extracted, it was assigned to this distance a **first degree** of correctness;
- **0<d<3**: if the distance was positive and less or equal to 3, a **second degree** of correctness was given to this distance. It means that the point was correctly extracted and that lies in the narrow zone of interest, so it may be considered as well extracted but not as precise as the previous category;
- **3<d<5**: if the distance greater than 3 and less or equal to 5, a **third degree** of correctness was assigned to this distance. The point was well localized and lies in the wider zone of interest, but the extraction was not precise;
- **d>5**: if the distance is greater than 5, a **fourth degree** of correctness was given to the distance. It means that the landmark was neither well extracted nor localized, so its position may be classified as incorrect.

These different degrees of correctness were associated to each of the 711 (= 79 facial shells * 9 landmarks) distances, so that it was then possible to count the number of occurrence of each degree for each landmark and evaluate rates of correctness. The results are graphically represented in Figure 4.

![Graphical representation of values of sample mean and sample variance computed between the 79 faces for the nine landmarks.](image)

**Figure 3.** Graphical representation of values of sample mean and sample variance computed between the 79 faces for the nine landmarks.

![Rates of correctness for each landmark (above) and rates of correctness for all landmarks (below).](image)

**Figure 5.** Rates of correctness for each landmark (above) and rates of correctness for all landmarks (below).

As we can see from the graph of Figure 4, the pronasal, the subnasal, and the nasion are mostly (at...
least 70%) perfectly localized, namely of degree 1, and of the second degree for the rest. Both the alae are mostly (60%) of the first degree of correctness, while only a low percent (maximum 10%) of the distances was of the third degree. The endocanthion and the exocanthion showed lower results, although the fourth degree rates were maximum 22%, obtained by the left exocanthion. Generally speaking, as Figure 4 shows, the great majority as well localized, i.e. about 95%, given by the sum of 50% of first degree category, 31% of degree 2, and 14% of the third degree. Considering that the extraction was only geometrical, the research reached very good results.

4 Conclusion

As described above, facial surface is a particular free-form surface, constituted by intersected and overlapped different patches. Each face is unique but its features are the same for all mankind. This fact allows to use the face as an instrument for recognition and identification.

Differential Geometry gave the research the possibility to describe facial features through landmarks, which are known worldwide by aesthetic plastic surgeons and face recognition companies. This research provided a geometric profile of each landmark which has noticeable feature, gives no identification problem and is far from the areas of face that are more subject and influenced by facial expressions, such as the mouth. The fact is that these points were not only accurately describable, but even suited to a geometrical description. Furthermore, their localization in the face was quite accurate. In fact, the results here obtained were shown to an expert, a maxillofacial surgeon, who confirmed their correctness. So, the Matlab® algorithm elaborated for the landmarks extraction holds for the seventy-nine faces and correctly works for all of them.

References


[22] Cyberware Color 3D Digitizer Product Information. Cyberware Inc. Monterey CA, USA.