A Pose-Independent Method for 3D Face Landmark Formalization

Enrico Vezzetti[1], *Sandro Moos, Federica Marcolin, Vincenzo Stola*
April 2012
Dipartimento di Ingegneria Gestionale e della Produzione
Politecnico di Torino, Italy

Abstract

Recently, 3D landmark extraction has been widely researched and experimented in medical field, for both corrective and aesthetic purposes. Automation of these procedures on three-dimensional face renderings is something desirable for the specialists who work in this field. In this work we propose a new method for accurate landmark localization on facial scans. The method relies on geometrical descriptors, such as curvatures and Shape Index, for computing candidate and initial points, and on a statistical model based on Procrustes Analysis and Principal Component Analysis, which is fitted to candidate points, for extracting the final landmarks. The elaborated method is independent on face pose.

Keywords

Landmark extraction, 3D face, Differential Geometry, PCA, Procrustes Analysis.

1. Introduction

Landmarks are body points with a particular biological meaning. In this paper we only deal with facial landmarks that lie on the skin, meaning soft-tissue landmarks, which have been widely employed in many activities involving various fields such as medical and, above all, maxillo-facial surgery. The positioning of these points helps surgeons to study their patients in pre-surgery phases both for plastic or corrective purposes, then to decide how to intervene. The

[1] Corresponding Author.
E-mail address: enrico.vezzetti@polito.it (E. Vezzetti).

landmarking was done firstly by hand, then with tools and algorithms supporting the activity of the specialists.

Experimental studies have been carried out on manual location of lower limb landmarks on dry bone models with a group of surgeons and concluded that the variations are in the range between 6 and 25 millimetres. It has also been possible to find a manual identification procedure for a scanned surface of foot model, guided by curvature values, for evaluating tibial torsion. Reproducibility of the results appears to be dependent on user's knowledge of landmarks (Della Croce *et al.*, 1999). Van Sint Jan (Van Sint Jan, 2007) listed a comprehensive set of bony landmarks and exposed a localization procedure description on a patient by palpation. Other researchers (Yang *et al.*, 2001) diagnosed facial growth abnormalities prior to treatment by studying the relationships between bony- and soft-tissue landmarks using cephalometric radiographs. Similarly, working with a manual procedure (Maudgil *et al.*, 1999), a new methodology has been proposed for extracting anatomical landmarks on a three-dimensional model reconstructed from Magnetic Resonance Imaging (MRI) images for morphometric analysis. They developed method that classifies each point on pre- and post-operative facial surfaces into one of eight surface patch types. This classification is based on the Mean and Gaussian curvatures: adjacent points with the same class are grouped into the same surface patch.

Recently, many automatic landmark extraction algorithms have been implemented on 3D medical images. 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 (Frantz et al., 1998; Frantz et al., 1999; Frantz et al., 2000; Alker et al., 2001; Frantz, et al., 2005; Wörz and Rohr, 2005). Other studies have been carried out on interactive landmarking on dry bone models, medical images (Griffin, 2000) or laser digitized data (Liu, 2004; Yahara, 2005).

Furthermore, in recent years, face study and landmark localization aimed at different purposes ware performed using both Principal Component Analysis (PCA) and Procrustes Analysis. Dalal and Phadke (Dalal and Phadke, 2007) used geometric morphometrics to analyze face variations of normal individuals and of people with dysmorphic syndromes. They used the Morphologika program for Procrustes Analysis and Principal Component Analysis; in

particular, face coordinates were subjected to Generalized Procrustes Superimposition (GPS) for normalizing for effects of size, rotation, and image position. The obtained Procrustes residuals were then subjected to PCA for data reduction. Soft-tissue landmarks were used as reference points. Mutsvangwa and Douglas applied Procrustes Analysis and PCA to stereo-photogrammetrically obtained landmarks for comparing facial features associated with Fetal Alcohol Syndrome (FAS) in subjects with FAS and normal controls. They state that "application of the Procrustes approach to facial shape analysis is becoming more widespread in syndrome diagnosis. The advantage of using PCA in conjunction with Procrustes Analysis is the ability to give a comprehensive description of the overall facial shape with a small number of landmark measurements that are not conflicting because they are statistically unrelated" (Mutsvangwa and Douglas, 2007: 215). Mena-Chalco *et al.* (Mena-Chalco *et al.*, 2008) described a system for three-dimensional face reconstruction from bi-dimensional photographs relying on a small set of training facial range images. Principal Component Analysis is used to represent facial datasets, thus defining an orthonormal basis of texture and range data. Mahoor and Abdel-Mottaleb (Mahoor and Abdel-Mottaleb, 2009) presented a 3D face recognition approach from frontal range data based on the ridge lines on facial surface. For the initial alignment of ridge points, they utilized the similarity transformation between a set of labelled facial feature points on the probe and gallery images. The parameters of this similarity transformation, namely scale, rotation, and translation, were estimated using Procrustes Analysis. Nair and Cavallaro presented an interesting approach to detect and segment three-dimensional faces, extract landmarks, and achieve fine registration of face meshes. Landmark localization is performed by finding the model fit which minimizes the model deviation from the mean shape. They used Procrustes Analysis to align the training shapes to their mutual mean in a least-squares sense via similarity transformations. Then, PCA was employed to estimate the variations of the shape cloud, providing "an efficient parameterization of the shape model through dimensionality reduction" (Nair and Cavallaro, 2009: 614). In their various works, Perakis *et al.* proposed a unified method addressing partial matching problem for face recognition. In particular, a new 3D landmark detector and a deformable model framework supporting symmetric fitting for detecting face pose are presented (Perakis *et al.*, 2009 [a]). Then, they proposed the first three-dimensional landmark detection method working in datasets with pose rotations of up to 80° around the *y*-axis (Perakis *et al.*, 2009 [b]). In both works, firstly the method creates an Active Landmark Model (ALM) by aligning the training

landmark sets and calculating a mean landmark shape using Procrustes Analysis; then, variations of each facial landmark model are computed using PCA.

Our previous study (Vezzetti and Marcolin) was a geometrically-based formalization of soft-tissue landmarks, in which derivatives, Shape and Curvedness Indexes, mean and Gaussian curvatures, and geometrical descriptors such as coefficients of the First and Second Fundamental Forms e, f, g, E, F, G were employed for performing an efficient landmark localization only through Differential Geometry, or at least a precise identification of the zone-of-interest which the landmark lies in. In this work we present an improvement to our previous algorithm, making use of more accurate conditions on geometrical descriptors and of a Procrustes- and PCA-based statistical model for extracting landmarks, which is independent on the reference system used. Section 2 presents the new method, in particular the statistical model and the face pose estimation. In Section 3 results are discussed.

2. The proposed method

A facial landmark is a point which all faces share and has a particular biological meaning. In particular, we may distinguish between two landmark types:

1. **hard-tissue** landmarks, which lie on the skeletal and may be identified only through lateral cephalometric radiographs;
2. **soft-tissue** landmarks, which are on the skin and can be identified on the point clouds generated by the scanning.

Since a radiograph is more invasive (and harmful) than a photogrammetric acquisition system, in this paper we considered only soft-tissue landmarks. Although soft-tissue landmarks are nearly fifty-nine, in this paper we take into consideration nine identifiable ones (*pronasal*, *nasion*, *subnasal*, *alae*, *endocanthions*, *exocanthions*), as shown in Figure 1. The landmarks close to the mouth are not taken into consideration due to their pose-dependency, while the ones near the face boundaries have been ignored because in those zones the scan is not accurate.

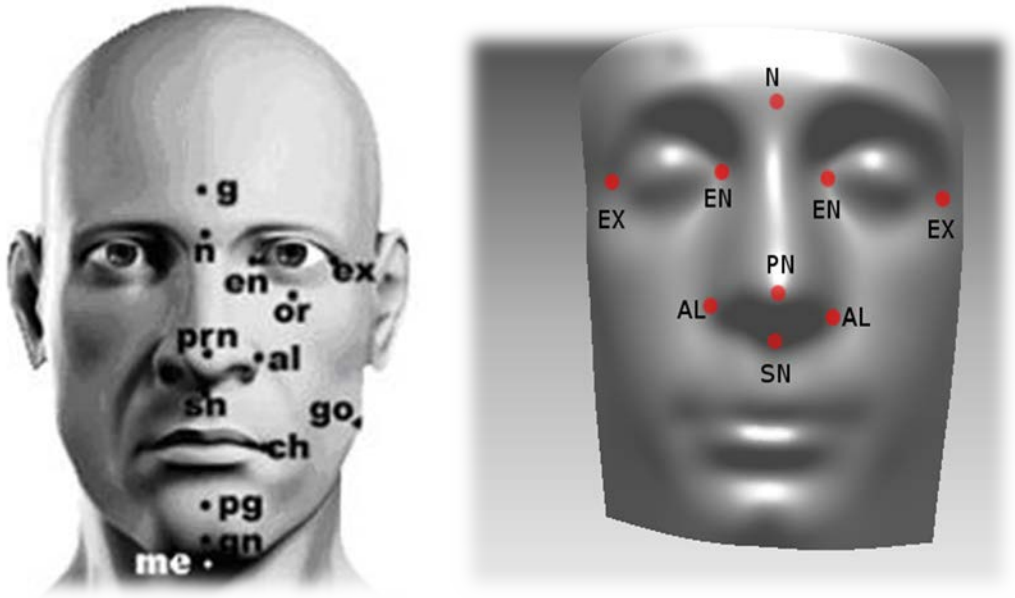


Figure 1. (left) the Anthropometric soft-tissue landmarks (g-*glabella*, n-*nasion*, en-*endocanthion*, ex-*exocanthion*, or-*orbital*, prn-*pronasal*, sn-*subnasal*, al-*alae*, ch-*cheilion*, pg-*pogonion*, gn-*gnathion*, go-*gonion*, me-*menton*) (Calignano, 2009); **(right)** landmarks detected by our proposed method (PN-*pronasal*, SN-*subnasal*, AL-*alae*, N-*nasion*, EN-*endocanthions*, EX-*exocanthions*).

In our previous work (Vezzetti and Marcolin) we made use of some geometric descriptors to detect and extract the landmarks. The employed descriptors are presented and defined in the following lines.

The First and the Second Fundamental Forms are employed to measure the distance on surfaces and are defined by

$$E du^2 + 2F dudv + G dv^2,$$

$$e du^2 + 2f dudv + g dv^2,$$

respectively, where E, F, G, e, f, and g are their coefficients and are calculated by the following formulas:

$$E = \|D_u\|^2,$$

$$F = \langle D_u, D_v \rangle,$$

$$G = \|D_v\|^2,$$

$$e = \langle N, D_{uu} \rangle,$$

$$f = \langle N, D_{uv} \rangle,$$

$$g = \langle N, D_{vv} \rangle,$$

where

$$D_u = \begin{cases} X_u = \frac{\partial X(v,u)}{\partial u} \\ Y_u = \frac{\partial Y(v,u)}{\partial u}, \\ Z_u = \frac{\partial Z(v,u)}{\partial u} \end{cases}$$

$$D_v = \begin{cases} X_v = \frac{\partial X(v,u)}{\partial v} \\ Y_v = \frac{\partial Y(v,u)}{\partial v}, \\ Z_v = \frac{\partial Z(v,u)}{\partial v} \end{cases}$$

$$D_{uu} = \frac{\partial Du}{\partial u}, D_{uv} = \frac{\partial Du}{\partial v}, D_{uv} = \frac{\partial Dv}{\partial v}, N = \frac{D_u \times D_v}{|D_u \times D_v|}.$$

Curvatures are used to measure how a regular surface bends in R^3. If D is the differential and N is the normal plane of a surface, then the determinant of DN will be the product of the Principal Curvatures ($det(DN) = (-k_1)(-k_2) = k_1 k_2$), and the trace of DN will be the negative of the sum of Principal Curvature ($trace(DN) = -(k_1 + k_2)$). In the point P, the determinant of DN_P is the Gaussian Curvature (K) at P, while the negative of half of the DN trace is called the Mean Curvature (H) at P. In terms of the principal curvatures it can be written:

$$K = k_1 k_2,$$

$$H = \frac{k_1 + k_2}{2},$$

where k_1 and k_2 are the Principal Curvatures. Starting from the coefficients of the Fundamental Forms, we may calculate the Gaussian and Mean Curvatures with the following formulas:

$$K = \frac{eg - f^2}{EG - F^2},$$

$$H = \frac{eG - 2fF + gE}{2(EG - F^2)}.$$

Obtained the Gaussian and Mean Curvatures we may calculate the Principal Curvatures in this way:

$$k_1 = H + \sqrt[2]{H^2 - K},$$

$$k_2 = H - \sqrt[2]{H^2 - K}.$$

The most used descriptors are surely the Shape (S) and Curvedness (C) Indexes, introduced by Koenderink and Van Doorn (Koenderink and Van Doorn, 1992):

$$S = -\frac{2}{\pi}\tan^{-1}\frac{k_1+k_2}{k_1-k_2}, \qquad S \in [-1,1], \qquad k_1 \geq k_2,$$

$$C = \sqrt[2]{\frac{{k_1}^2+{k_2}^2}{2}}.$$

The only descriptor we did not use in our previous work is the Tangent Map, an index used by Perakis *et al.* (Perakis *et al.*, 2010) to detect the points which have the normal outward with respect to the centroid of the surface like nose and chin regions. The Tangent Map is calculated by the following formula:

$$T(P) = \langle N(P), R(P) \rangle,$$

where N is the normal of the surface at P and R is the straight line passing through the centroid of the surface and P.

However, one problem affects the previous algorithm: the geometry-based landmark detection is affected by face orientation. In fact, the partial derivatives and the coefficients of the First and Second Fundamental Forms depend on the reference system used. This means that, if the face is not in the standard pose, the local behaviour of the previous descriptors will not be the one described in our previous study. Contrariwise, the curvatures and Shape and Curvedness Indexes are intrinsic proprieties of the surface, thus they are independent on the reference system used.

If the input face is not in a standard pose, the ideal algorithm should rotate the face until the standard pose is identified. Our previous algorithm does not perform any rotation operations. After a brief experimentation phase, we noticed that the old algorithm correctly detects the landmarks if the initial rotations of the face around the single axes is in a range between $-10°$ and $+10°$. For instance, the *subnasal* is the point in the region underlying the *pronasal* that maximizes the Coefficient g; but what happens if we rotate the face around the z-axis by 90°? As shown by comparing Figures 2 and 3, there are two problems:

1. the search region must be to the right of the *pronasal*;
2. the behavior of coefficient g coincides with the behavior of the coefficient e and vice versa.

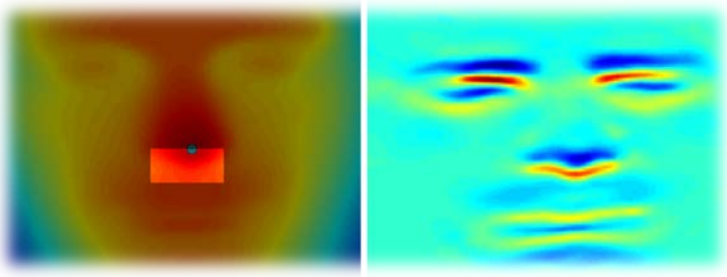


Figure 2. (left) A face well oriented; in blue the *pronasal*, while the brightest region is the search area of the *subnasal*; **(right)** the coefficient g used to extract the *subnasal*.

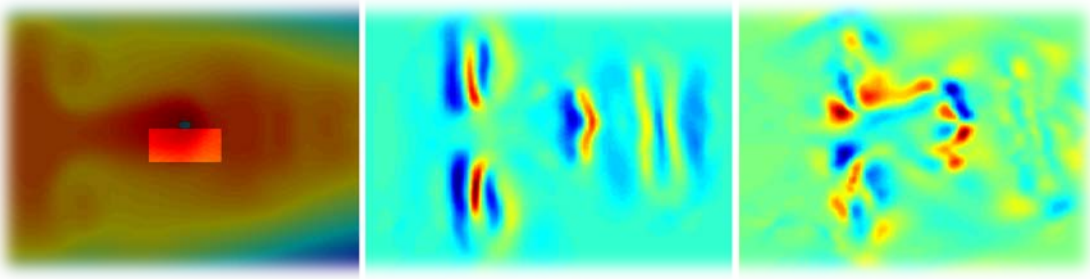
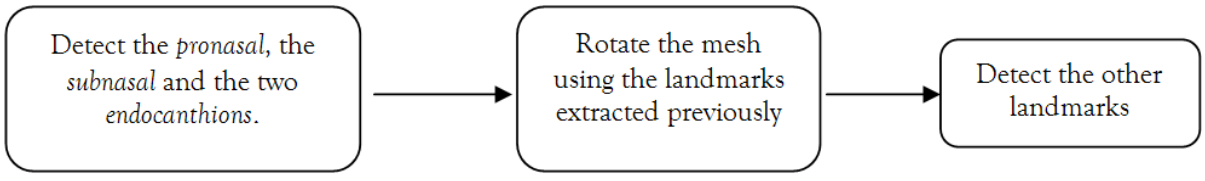


Figure 3. (left) A face rotated by 90° around z-axis; in blue the *pronasal*, while the brightest region should be the search area of the *subnasal*, according to our old algorithm; but with this orientation it must be to the right of the *pronasal*; **(center)** the coefficient e: it coincides with the graphical representation of coefficient g of the same face non-rotated; **(right)** the coefficient g of this face.

To solve this problem, it is necessary to avoid using the descriptors which depend on the reference system and to impose strong geometric constrains. So, there will be more candidates for each landmark and a new method to choose the landmark will be necessary. In this work we use a statistic model to find the real landmarks among the possible candidates. In particular, the new method is divided into three phases:

- firstly the *pronasal*, the *subnasal*, and the *endocanthions* are detected using our previous algorithm only with the descriptors which are not dependent on the reference system;
- subsequently, the mesh is rotated in a standard pose;
- finally, since the face is in the standard pose, a statistical model, described in the next section, is used to detect the remaining landmarks.



2.1 The statistical model

The statistical model here explained intervenes in the last phase of landmark extraction, in particular after having geometrically identified the candidate points and the initial points for each landmark. It is used twice in this algorithm: firstly during pose estimation, namely for identifying pronasal, subnasal, and endocanthions, then in the phase of extraction of the other landmarks. To create the statistical model were used faces belonging to 5 persons performing different expressions, in order to make the statistical model more accurate and more varied. According to Dryden and Mardia (Dryden and Mardia, 1998), "a landmark is a point of correspondence on each object that matches between and within populations of the same class of objects", while "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. In this context, the landmark set of each face is called "example shape" or "landmark shape".

The setting of the statistical model is performed using the same approach used by Perakis *et al.* (Perakis *et al.*, 2009). The main steps to be built it are listed here:

- a "statistical mean shape" is calculated using **Procrustes Analysis**;
- eligible variations of the mean shape are calculated using **Principal Component Analysis** (PCA).

The Procrustes Analysis is used to analyse the distribution of a set of shapes. To compare the shape of two or more objects, the objects must be firstly optimally aligned. Alignment is performed by minimizing the Procrustes distance:

$$D_p^2 = |x_r - x_i|^2 = \sum_{j=1}^{k} \left(x_{rj} - x_{ij}\right)^2,$$

where x_i is the i^{th} among the example shapes x_i that we want to align, x_r is the reference shape, while k is the number of landmarks considered. The alignment procedure between x_i and x_r is performed by the Procrustes function.

The Procrustes function defines a linear transformation (translation, reflection, orthogonal rotation, and scaling) of the points of two shapes. The "goodness-of-fit" criterion is the sum of squared errors (MathWorks). In this work, we want to discriminate the left side from the right side of the face, thus we do not consider the reflection. Furthermore, the size is

also important since it is used like a parameter to find the proper set of candidate points, even if the scaling is not considered.

Given the shapes X and Y, the transformation is obtained this way:

1. find the two centroids of the shapes X and Y (respectively c_X and c_Y);
2. translate the two shapes so that their centroids are at the origin;
3. sum the squares of each element of the shape X (S_X) and of the shape Y (S_Y), then extract their square root (respectively $norm_X$ and $norm_Y$);
4. scale X with $norm_X$ and scale Y with $norm_Y$, to obtain respectively X_N and Y_N;
5. compute the singular value decomposition of the $X_N^T * Y_N$, to obtain the matrices U, S, and V ($X_N^T * Y_N = U * S * V^T$);
6. compute the rotation matrix $R = V * U^T$;
7. if the determinant of R is equal to -1 (the transformation R includes a reflection), then invert the matrices V and S (undo the reflection) and compute again the rotation matrix with the previous formula;
8. the best linear transformation which aligns the shape Y to the shape X is:

   $Z_i = Y_i * R + c_X$,

   where Y_i is a point of the shape Y.

After the alignment, the Procrustes Analysis, meaning the computation of x_m, could be summarized in the following steps:

1. assign the first example shape to the mean shape x_m;
2. assign the mean shape x_m to the reference shape x_r;
3. align each example shape x_i to the reference shape x_r with the Procrustes function;
4. compute the mean shape as an average of the all example shapes;
5. compute the Procrustes distance between the mean shape x_m and the reference shape x_r;
6. if the Procrustes distance is less than a threshold, then exit; else return at point 2.

Due to size normalization of Procrustes function, all shape vectors live in a hyper-sphere manifold in shape space, which introduces non-linearities if large shape scaling occurs. Since PCA is a linear procedure, all aligned shapes are firstly projected onto the tangent space of the

mean shape. So, shape vectors lie in a hyper-plane instead of a hyper-sphere, and non-linearities are filtered out. A simple bi-dimensional representation is shown in Figure 4.

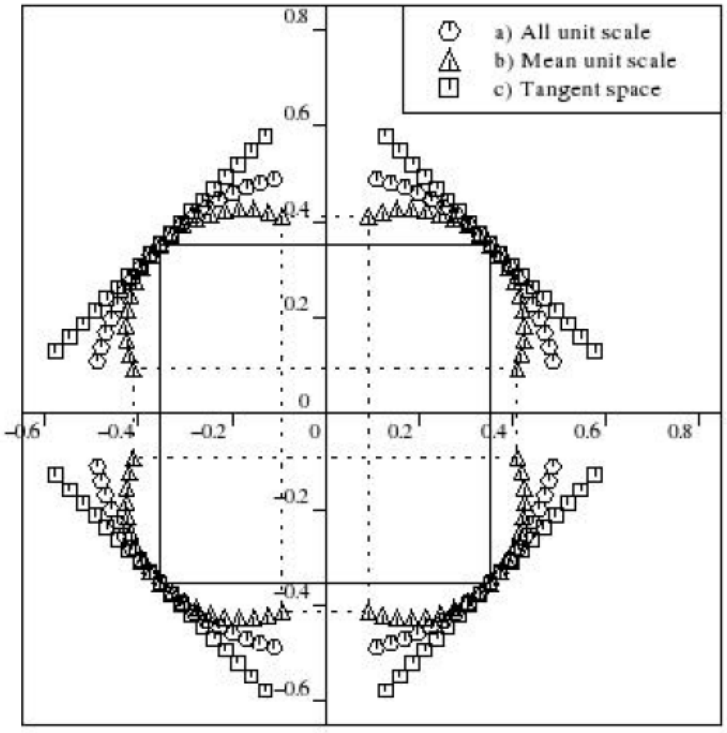


Figure 4. A simple two-dimensional representation of the shape dispositions. In the axes origin there is the mean shape, while the small polygons are the example shapes. In particular, a) represents example shapes which have unit norm, b) represents example shapes which have norm equal to the mean shape norm, c) represents example shapes projected in the tangent plane. We can note that the example shapes which have unit norm lie on a circle. Extending the concept in three-dimensional space, the circle becomes a sphere, while the tangent plane becomes a tangent space.

The tangent space projection linearizes shapes by scaling them with a factor α:

$$x_{it} = \alpha x_i = \frac{|x_m|^2}{x_m \cdot x_i} x_i,$$

where x_{it} is the tangent space projection of shape x_i and x_m is the mean shape. If no size normalization is applied, then tangent space projection can be omitted.

After turning the landmark shapes into a common frame of reference and estimating the landmark mean shape, further analysis can be carried out for describing the shape variations. This shape decomposition is performed by applying PCA to the aligned shapes. To perform PCA, a new representation of shapes may be useful. Each shape is composed by k landmarks in d dimensions, therefore a new representation could be defined by concatenating all point coordinates into a $n = k * d$ vector; in this case, d is equal to 3, thus the representation is:

$$s_i = [x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_k, y_k, z_k]^T,$$

where (x_i, y_i, z_i) represents each landmark. Aligned shape vectors form a distribution in the $k * d$ dimensional shape space, where k is the number of landmarks and d the dimension of each landmark. If landmark points do not represent a certain class of shapes, then they will be totally uncorrelated (i.e., purely random). On the other hand, if landmark points present a certain class of shapes, then they will be correlated with some degrees.

The idea is to estimate a vector of parameters that describes shape deformations (Cootes *et al.*, 1995; Cootes *et al.*, 2001; Cootes *et al.*, 2005; Stegman *et al.*, 2002) in a space where landmarks coordinates are totally uncorrelated from each other: this will be exploited by applying PCA. The correlated space is the space where landmarks coordinates are correlated from each other, while the uncorrelated space will be the space where landmarks coordinates are totally uncorrelated. Furthermore, we shall indicate with s the generic example shape in the correlated space, while r will be the generic example shape in the uncorrelated space. For each space, we may determine the covariance matrix of N example shapes according to:

$$C_x = \frac{1}{N-1}\sum_{i=1}^{N}(s_i - x_m)(s_i - x_m)^T, \quad C_y = \frac{1}{N-1}\sum_{i=1}^{N}(r_i - y_m)(r_i - y_m)^T,$$

where C_x is the covariance matrix in the correlated space, C_y is the covariance matrix in the uncorrelated space (C_y will be a diagonal matrix since landmarks coordinate are totally uncorrelated), s_i is the i^{th} among the example shapes in the correlated space, r_i is the i^{th} among the example shapes in the uncorrelated space, x_m is the mean shape in the correlated space, and y_m is the mean shape in the uncorrelated space. Between the two covariance matrices exists the following relationship holds:

$$C_x \cdot A = A \cdot C_y.$$

Therefore, C_y can be computed with the following formula:

$$C_y = A^T \cdot C_x \cdot A.$$

The resulting transform is known as the **Karhunen-Loéve Transform**, and achieves our original goal of creating mutually uncorrelated features. In fact, A is a matrix that contains (in columns) the $n = k * d$ eigenvectors of C_x; therefore, projecting aligned original example shapes to the eigenspace, we uncorrelate them as:

$$r = A^T \cdot (s - x_m).$$

To back-project uncorrelated shape vectors onto the correlated space, we can use:

$$s = x_m + A \cdot r.$$

Now, we want to consider a subspace of the uncorrelated space spanned by the most significant eigenvectors of C_x (principal component), neglecting the less significant ones, which are used to represent noise (Cootes *et al.*, 2001; Theodoris *et al.*, 2006). The idea is to consider the p eigenvectors associated to the p largest eigenvalues, so the mean square error between s and its approximation s' is minimized. Firstly, the eigenvalues are ordered in descending order; secondly, the eigenvalues are summed between them until the following relationship is verified:

$$\sum_{i=1}^{p} \lambda_i = f * \lambda_T,$$

where p is the number of considered eigenvalues, λ_i is the i^{th} eigenvalue, λ_T is the sum of all eigenvalues, and factor f is the percentage of total variance incorporated into statistical model. In this work f is equal to 0.98.

Let's define the matrix ϕ as containing (in columns) the p considered eigenvectors, approximating any example shape s using:

$$s' \approx x_m + \phi \cdot b,$$

where b is a p-dimensional vector given by:

$$b = \phi^T \cdot (s - x_m).$$

The vector b is the projection of s onto the subspace spanned by the p most significant eigenvectors of the eigenspace (principal components) and represents the variations in the subspace of the uncorrelated space between the shape s and the mean shape.

So that the shapes generated to be eligible, it is necessary to limit the variation range of parameters; it could be useful to estimate from the landmark set the probability density of b, so that the algorithm will be able to establish if the calculated parameters belong to this distribution. We may set that the parameters are eligible if $p(b) \geq p_t$, where p_t is a threshold we established. A general idea consists in considering all parameters statistically independent of each other and to suppose that they have a **Gaussian probability distribution**. Remembering that each parameter b_i has a variance equal to λ_i, the probability density of the vector b is:

$$p(b) = \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi\lambda_i}} e^{-\frac{b_i^2}{2\lambda_i}}.$$

Under this assumption, a limitation able to ensure that the shape generated do not differ too much from those in the landmark set is given by:

$$-n\sqrt{\lambda_i} \leq b_i \leq n\sqrt{\lambda_i},$$

where n is equal to 2 or 3. As known, a Gaussian probability distribution has the confidence interval between $\mu - 3\sigma$ and $\mu + 3\sigma$ (where μ is the mean of the distribution and σ is the standard deviation), which comprises about 99% of the population extracted from itself, so $n = 3$ seems to be a good choice since it allows to consider almost all of the landmark set.

In conclusion, the eligibility of parameters is verified by the condition:

$$-3\sqrt{\lambda_i} \leq b_i \leq 3\sqrt{\lambda_i}.$$

Each parameter which does not satisfy the previous condition is truncated to the nearest allowable value. In particular:

1. if $b_i > 3\sqrt{\lambda_i}$ then the value $3\sqrt{\lambda_i}$ will be assigned to b_i;
2. if $b_i < -3\sqrt{\lambda_i}$ then the value $-3\sqrt{\lambda_i}$ will be assigned to b_i.

In brief, the approach is as follows:

1. determine the mean shape x_m;
2. determine the covariance matrix C_x of the shape vectors s_i;
3. compute the eigenvectors 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 tangent space of the mean shape.

2.2 **Estimating face orientation**

Face orientation is estimated through the landmarks *pronasal*, *subnasal*, and *endocanthions*. Once these landmarks are detected, face could be rotated to the standard pose. Since we do not know face orientation, in order to localize these landmarks, we can use only the descriptors which do not depend to the reference system used, since they rely on intrinsic proprieties of surfaces. These descriptors are the Principal Curvatures (k_1 and k_2), the Gaussian and Mean Curvature (K and H), the Tangent Map (T), and Shape and Curvedness Indexes. In particular, with these conditions on the geometrical descriptors, in this phase we extract the

candidate points and the initial point, i.e. the point at which the subsequent statistical model starts its computation. Finally, the statistical model, explained in 2.1, is computed on the four landmarks (*pronasal*, *subnasal*, and two *endocanthions*) and iteratively fitted on candidate points, in order to have a precise extraction of them.

Firstly, the algorithm computes a threshold used to cluster the *endocanthion* candidate points. The scope of the threshold is that the close candidate points are related to the same landmark, therefore they will not be considered as different landmarks at the same time. Given its scope, the threshold depends on face size and is computed with the following formula:

$$threshold = \frac{\sqrt{(\max \mathrm{Z} - \min \mathrm{Z})^2 + (\max \mathrm{Y} - \min \mathrm{Y})^2 + (\max \mathrm{X} - \min \mathrm{X})^2}}{20},$$

where $\max \mathrm{Z}$, $\min \mathrm{Z}$, $\max \mathrm{Y}$, $\min \mathrm{Y}$, $\max \mathrm{X}$, and $\min \mathrm{X}$ are the maximums and minimums of *x*-, *y*-, and *z*-coordinates, respectively. In other words, we calculate the diagonal of the face bounding box, i.e. the imaginary box in which the face is inscribed (shown in Figure 16), and we divide it by twenty.

Secondly, the algorithm normalizes the descriptors in the range $[0,1]$, so we can express the conditions on the descriptors with the form of a probability.

Thirdly, the sets of the candidate points for each landmark is computed. The sets are found filtering out the mesh points out of the through some conditions; in particular:

1. the *pronasal* candidate points satisfy these conditions:
    a. very low values of the Shape Index ($S \leq 0.1$);
    b. high values of the Principal Curvature ($k_1 \geq 0.65$ and $k_2 \geq 0.7$);
2. the *subnasal* candidate points satisfy these conditions:
    a. very high values of the Tangent Map ($T > 0.8$);
    b. high values of the Shape Index ($S \geq 0.4$);
3. the *endocanthion* candidate points satisfy these conditions:
    a. very high values of the Shape Index ($S > 0.9$);
    b. high values of the Curvedness Index normalized in the range $[0,1]$ in the points which satisfy the previous condition ($C_{EN} > 0.5$).

In Figure 5 are shown the candidate points for each landmark. We can note that there are many false candidates, therefore the initial choice of the points is very important since the

statistical model is fitted in a neighbourhood of the starting points; in fact, if the initial points are not close to the real landmarks and the points detected by the algorithm are completely wrong.

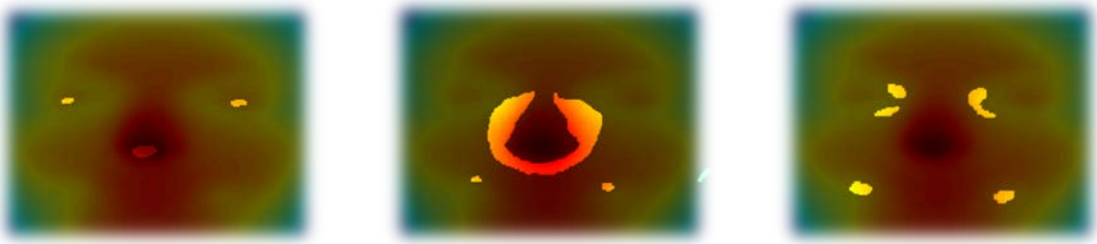


Figure 5. In the brightest regions there are the *pronasal* candidate points **(left)**, the *subnasal* candidate points **(centre)**, the *endocanthions* candidate points **(right)**.

Then, the initial points are chosen among candidate points. In particular:

a. the initial *pronasal* is the point which maximizes the Principal Curvature k_2;
b. the initial *subnasal* is the point closest to the initial *pronasal*;
c. the initial *endocanthions* are found this way:
    1. find the first initial *endocanthion* maximizing the Shape Index;
    2. the second initial *endocanthion* is searched maximizing the Shape Index between the candidate points having a distance from the first initial *endocanthion* greater than threshold and less than three times the threshold (threshold is the value initially computed by the algorithm);
    3. if the second initial *endocanthion* is not found, then the algorithm deletes from the set of *endocanthion* candidate points the first initial *endocanthion* and the ones which have a distance from it less than the threshold; subsequently, the algorithm returns to the step 1. The idea is that the distance between the two *endocanthions* belongs to the range $[threshold; 3 * threshold]$, therefore the first initial *endocanthion* and its neighbours are deleted because they are certainly some false candidate points.

Once the initial points are extracted, it is necessary to define the left and the right *endocanthion* between the two extracted. This operation is performed through Analytic Geometry rules in the space. In particular, we consider two oriented straight lines (vectors):

1. the first vector starts from initial *pronasal* and ends on the first initial *endocanthion*;
2. the second vector starts from initial *pronasal* and ends on the second initial *endocanthion*.

As shown in Figure 6, the vector cross products between the first and the second vector produces two possible vectors, which lie on the same straight line but have opposite directions.

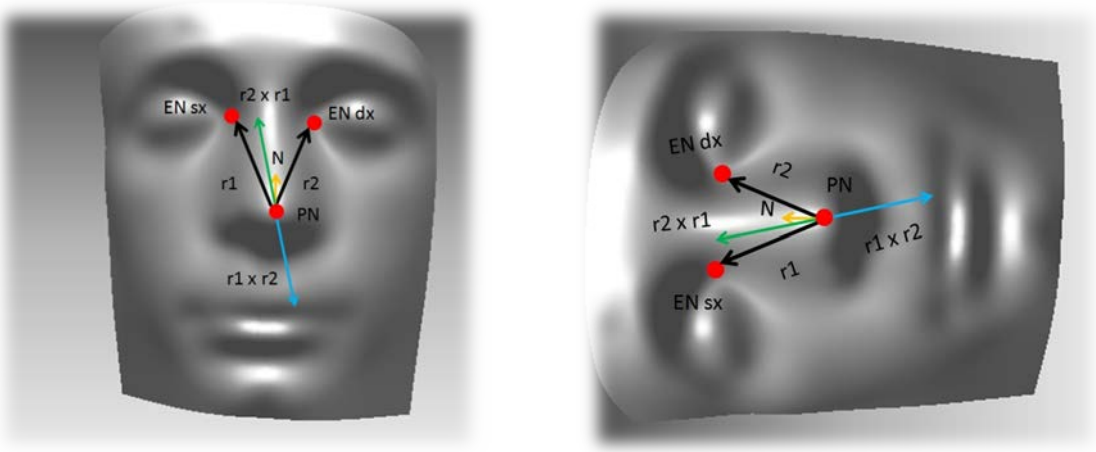


Figure 6. (left) the black lines are the vectors which link each *endocanthion* with the *pronasal*; in yellow the surface normal to the *pronasal* point, while the green and blue vectors are the result of the vector cross products between the black vectors; naming $\boldsymbol{r_1}$the vector which links the *pronasal* with the left *endocanthion* and $\boldsymbol{r_2}$ the vector which links the *pronasal* with the right *endocanthion,* the blue vector is the result of the vector cross product between $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$, while the green vector is the result of the vector cross product between $\boldsymbol{r_2}$ and $\boldsymbol{r_1}$. The green and yellow vectors have nearly the same direction, while blue and yellow vectors have nearly opposite directions; **(right)** the same situation happens on the rotated face, therefore this method is invariant with respect to face orientation.

The general idea is to perform the vector cross product between the first and the second vector and to compute the vector inner product between the surface normal at the *pronasal* and the result of the previous operation. There are two cases:

1. if the vector inner product is negative, then the first initial *endocanthion* will be the left one;
2. if the vector inner product is positive, then the first initial *endocanthion* will be the right one.

Now it is possible to perform the fitting procedure, i.e. applying the statistical model introduced in section 2.1. The procedure consists in a serial steps repeated until it converges. The convergence is reached when the landmarks detected at the i^{th}-iteration are very close to the landmarks detected at the $(i^{th} - 1)$-iteration. Considering that the landmark shape x_l is composed by the initial point extracted previously, the following steps are:

1. assign the mean shape of the statistical model to the reference shape x_r;
2. align the landmark shape x_l to the reference shape x_r with the Procrustes function, to obtain the linear transformation T;

3. compute the vector b using the formula: $b = \phi^T \cdot (s - x_m)$;
4. verify if the vector b is eligible: if the condition is not verified, then make the vector b eligible, how said previously;
5. compute the new reference shape x_r through the formula: $x_r = x_r + \Phi \cdot b$;
6. transform the reference shape x_r using the linear transformation T^{-1}, to obtain the new landmark shape x_{l1};
7. for each landmark in the shape x_{l1} search in its candidate points the closest point, which will be the new landmark in the shape x_{l1};
8. if the landmarks in the shape x_{l1} are very close to the landmarks in the shape x_l then exit; else assign the shape x_{l1} to the landmark shape x_l and return to the step 2.

Some results of this step are shown in Figure 7.

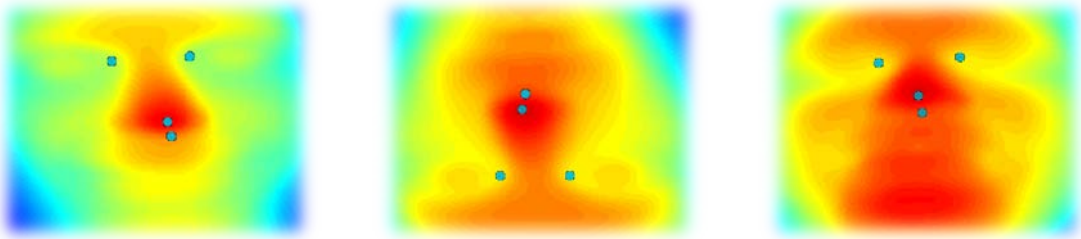


Figure 7. Some results of the fitting phase.

Once the fitting procedure is done, the *pronasal*, the *subnasal*, and the two *endocanthions* are detected. At this point, the algorithm rotates the face in a specific pose through Analytic Geometry. In this work we use a reference system like the one shown in the Figure 8.

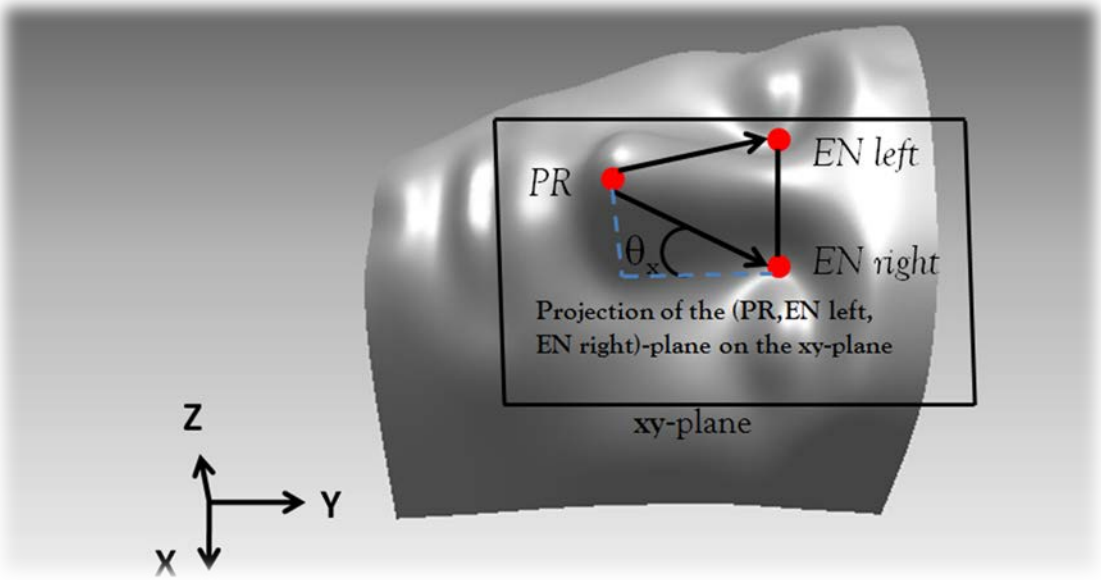


Figure 8. The new reference system used. The plane defined by the *pronasal* and the two *endocanthions* is rotated around $\boldsymbol{x}$-axis of $\boldsymbol{\theta_x}$; the $\boldsymbol{y}$-axis is a vector parallel to the vector which starts from *pronasal* and finishes to the midpoint between the two *endocanthions*; the $\boldsymbol{x}$-axis is a vector whose direction is toward the right side of the face; the $\boldsymbol{z}$-axis is a vector which completes the frame; the axes origin coincides with the mesh centroid.

In order to find the transformation matrix which rotates the face into standard pose, we use Analytic Geometry. Firstly, we have to define a reference value to θ_x that could be:

$\theta_x = -0{,}8573578984258491\, rad.$

Subsequently, we compute these steps:

1. translate the mesh so that its centroid coincides with the origin of axes;
2. compute the unit normal of the plane defined by the *pronasal* and the two *endocanthions* ($n = \frac{r_2 \times r_1}{|r_2 \times r_1|}$);
3. compute the unit vector (m) which starts from the *pronasal* and finishes to the midpoint of the two *endocanthions*;
4. compute a third vector (l) that is perpendicular to the vectors n and m using the formula: $l = m \times n$;
5. define the rotation matrix in this way:

$$R = [\, l \mid m \mid n \,]^{-1};$$

6. the rotation matrix R rotates the face in a new reference system where the plane defined by the *pronasal* and the two *endocanthions* is parallel with the xy-plane; therefore, it is necessary to rotate the face around x-axis of θ_x;

7. the final rotation matrix is defined in this way:

$$R_f = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix} * R.$$

So, face is rotated to the standard pose and we can proceed with the detection of the remaining landmarks.

2.3 Detecting all landmarks

The detection of the other landmarks is performed through the use of a statistical model computed on all landmarks and, since the face is in a standard pose, with the aid of the other geometrical descriptors, such as the derivatives and E, F, G, e, f, g. In this phase the Tangent Map, the Shape Index, and the Principal Curvature k_2, which are used to choose the candidate points, have been normalized in the range $[0,1]$.

Firstly the algorithm finds the parametric coordinates (u and v) of the landmarks previously detected and computes the sets of the candidate points for each landmark. In particular:

1. the *pronasal* candidate points are the points which have parametric coordinates in the ranges:

   $$u \in [u_{PN} - 3, u_{PN} + 3],\ v \in [u_{PN} - 3, v_{PN} + 3],$$

   where u_{PN} and v_{PN} are the *pronasal* parametric coordinates;
2. the *subnasal* candidate points are the points which have parametric coordinates in the ranges:

   $$u \in [u_{SN} - 3, u_{SN} + 3],\ v \in [u_{SN} - 3, v_{SN} + 3],$$

   where u_{SN} and v_{SN} are the *subnasal* parametric coordinates;
3. the left *endocanthion* candidate points are the points which have parametric coordinates in the ranges:

   $$u \in [u_{ENsx} - 6, u_{ENsx} + 6],\ v \in [u_{ENsx} - 6, v_{ENsx} + 6],$$

   where u_{ENsx} and v_{ENsx} are the left *endocanthion* parametric coordinates;
4. the right *endocanthion* candidate points are the points which have parametric coordinates in the ranges:

   $$u \in [u_{ENdx} - 6, u_{ENdx} + 6],\ v \in [u_{ENdx} - 6, v_{ENdx} + 6],$$

   where u_{ENsx} and v_{ENdx} are the right *endocanthion* parametric coordinates;

5. the *nasion* candidate points are the points which have parametric coordinates in the ranges:

$$u \in [u_{ENsx} + 6, u_{ENdx} - 6],$$
$$v \in [\min(v_{ENsx}, v_{ENdx}) - 15, \max(v_{ENsx}, v_{ENdx}) + 15];$$

furthermore, these points are filtered with the following conditions:

   a. low values of the Shape Index ($S < 0.5$);
   b. the Mean Curvature satisfies the condition: $H \in (-0.5, +0.5)$;

6. the left *alae* candidate points are the points which have parametric coordinates in the ranges:

$$u \in [u_{PN} - 40, u_{PN} - 10],\ v \in [u_{PN} - 8, v_{PN} + 8];$$

7. the right *alae* candidate points are the points which have parametric coordinates in the ranges:

$$u \in [u_{PN} + 10, u_{PN} + 40],\ v \in [u_{PN} - 8, v_{PN} + 8];$$

8. the left *exocanthion* candidate points are the points which have parametric coordinates in the ranges:

$$u \in [10, u_{ENsx} - 25], \quad v \in [v_{ENsx} - 8, v_{ENsx} + 6];$$

furthermore, these points are filtered with the following conditions:

   a. the Tangent Map T must be in the range between 0.6 and 0.9 ($T \in [0.6,0.9]$);
   b. the Shape Index must be equal or less than 0.75 ($S \leq 0.75$);
   c. the Coefficient e must be in the range between -0.3 and 0.3 ($-0.3 \leq e \leq 0.3$);
   d. the Coefficient E must be in the range between 1.5 and 3.5 ($1.5 \leq E \leq 1.5$);

9. the right *exocanthion* candidate points are the points which have parametric coordinates in the ranges:

$$u \in [u_{ENdx} + 25,140], \quad v \in [v_{ENdx} - 8, v_{ENdx} + 6];$$

furthermore, these points are filtered with the following conditions:

   a. the Tangent Map T must be in the range between 0.6 and 0.9 ($T \in [0.6,0.9]$);
   b. the Shape Index must be equal or less than 0.75 ($S \leq 0.75$);
   c. the Coefficient e must be in the range between -0.3 and 0.3 ($-0.3 \leq e \leq 0.3$);
   d. the Coefficient E must be in the range between 1.5 and 3.5 ($1.5 \leq E \leq 1.5$);

Subsequently, the initial points must be extracted. Initial points for *pronasal*, *subnasal*, and *endocanthions* are the same detected in section 2.2, while the others are extracted this way:

1. the initial *alae* (one for side) are the two points which maximize the Coefficient e between their candidate points;
2. the initial *endocanthions* (one for side) are the two centroids of their candidate points;
3. the initial *nasion* is the centroid of its candidate points.

Figures 9, 10, 11, 12 show the regions-of-interest of each landmark.

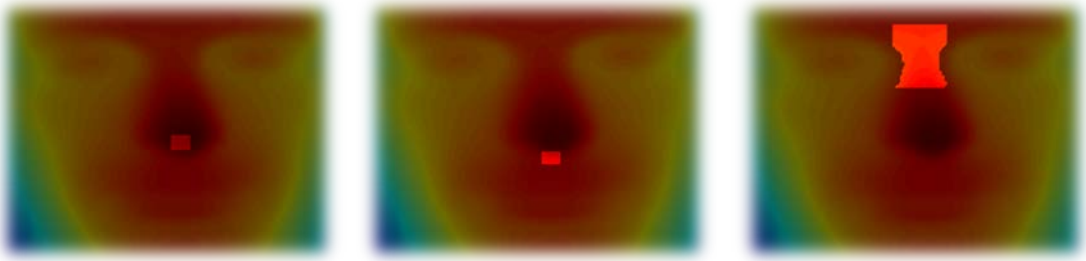


Figure 9. In the brightest regions there are the *pronasal* candidate points **(left)**, the *subnasal* candidate points **(center)**, the *nasion* candidate points **(right)**.

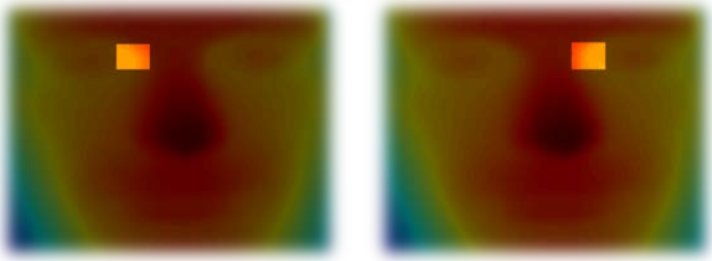


Figure 10. In the brightest regions there are the left *endocanthion* candidate points **(left)** and the right *endocanthion* candidate points **(right)**.

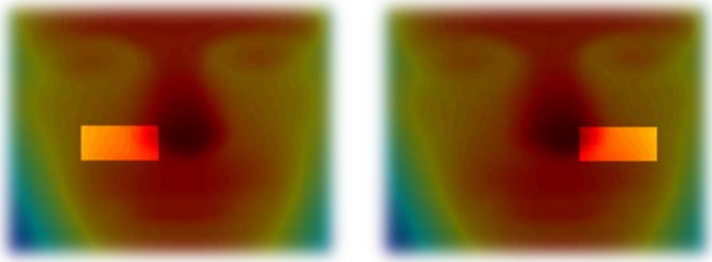


Figure 11. In the brightest regions there are the left *alae* candidate points **(left)** and the right *alae* candidate points **(right)**.

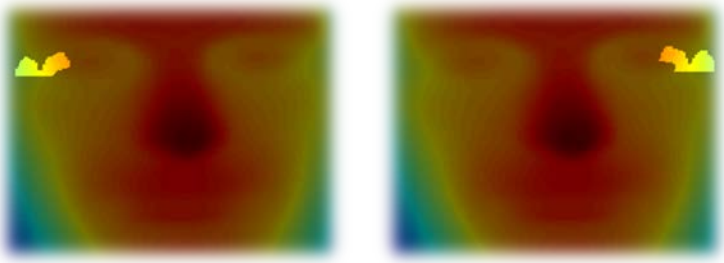


Figure 12. In the brightest regions there are the left *exocanthion* candidate points **(left)** and the right *exocanthion* candidate points **(right)**.

Once the initial points are extracted, the statistical model intervenes and the fitting phase begins. It is similar to the fitting procedure previously described, but it differs from the previous one in one point: when the statistical model is fitted to the mesh, the algorithm now searches among the candidate points not the closest points to the initial ones, but the points which satisfy some geometrical conditions. These conditions are:

1. the *pronasal* is the point which maximizes Z in the range:
$$u \in [u_{PN} - 2, u_{PN} + 2],\ v \in [u_{PN} - 2, v_{PN} + 2], with\ (u, v) \in CP(PN),$$
where u_{PN} and v_{PN} are the parametric coordinates of the *pronasal* extracted and $CP(PN)$ is the set of the its candidate points;
2. the *subnasal* is the point which maximizes the Coefficient g between the five points (one for each v-parameter) which maximize Z in the range:
$$u \in [u_{SN} - 2, u_{SN} + 2],\ v \in [u_{SN} - 2, v_{SN} + 2], with\ (u, v) \in CP(SN),$$
where u_{SN} and v_{SN} are the parametric coordinates of the *subnasal* extracted and $CP(SN)$ is the set of the its candidate points;
3. the two *alae* are the two points which maximize the Coefficient e between the five points (one for each u-parameter) which maximize Z in the range:
$$u \in [u_{Alae} - 2, u_{Alae} + 2], v \in [u_{Alae} - 2, v_{Alae} + 2], with\ (u, v) \in CP(Alae),$$
where u_{Alae} and v_{Alae} are the parametric coordinates of the left or right *alae* extracted and $CP(Alae)$ is the sets (one for side) of the their candidate points;
4. the two *endocanthions* are the two points which maximize the Coefficient g between the three points (one for each u-parameter) which minimize Z in the range:
$$u \in [u_{EN} - 1, u_{EN} + 1],\ v \in [u_{EN} - 1, v_{EN} + 1], with\ (u, v) \in CP(EN),$$
where u_{EN} and v_{EN} are the parametric coordinates of the left or right *endocanthion* extracted and $CP(EN)$ is the sets (one for side) of the their candidate points;

5. the two *exocanthions* are the two midpoints between two points which maximize the Coefficient g and minimize the Shape Index in the range:

$$u \in [u_{EX} - 2, u_{EX} + 2],\ v \in [u_{EX} - 2, v_{EX} + 2], with\ (u, v) \in CP(EX)$$

where u_{EX} and v_{EX} are the parametric coordinates of the left or right *exocanthion* extracted and $CP(EX)$ is the sets (one for side) of the their candidate points;

6. the *nasion* is the point which minimizes the Coefficient g between the seven points (one for each v-parameter) which maximize Z in the range:

$$u \in [u_N - 3, u_N + 3],\ v \in [u_N - 3, v_N + 3], with\ (u, v) \in CP(N),$$

where u_N and v_N are the parametric coordinates of the *nasion* extracted and $CP(N)$ is the set of the its candidate points.

In summary, the fitting procedure performs the following steps:

1. the landmark shape x_l is composed by the initial point extracted previously;
2. assign the mean shape of the statistical model to the reference shape x_r;
3. searches in small neighborhoods of the points in the landmark shape x_l the points (one for landmark) which have some typical features of a landmark;
4. align the landmark shape x_l to the reference shape x_r with the Procrustes Function, to obtain the linear transformation T;
5. compute the vector b using the formula: $b = \phi^T \cdot (s - x_m)$;
6. verify if the vector b is eligible; if the conditions is not verified, then make the vector b eligible, how said previously;
7. compute the new reference shape x_r through the formula: $x_r = x_r + \Phi \cdot b$;
8. transform the reference shape x_r using the linear transformation T^{-1}, to obtain the new landmark shape x_{l1};
9. for each landmark in the shape x_{l1} search in its candidate points the closest point, which will be the new landmark in the shape x_{l1};
10. if the landmarks in the shape x_{l1} are very close to the landmarks in the shape x_l then exit; else assign the shape x_{l1} to the landmark shape x_l and return to the step 3.

In Figure 13 we can see the intermediate steps of the fitting procedure. In general, after six iterations, the landmarks extracted are very close to the real landmarks; in fact, the landmarks shown in the centre figures and in the right figures are nearly in the same position.

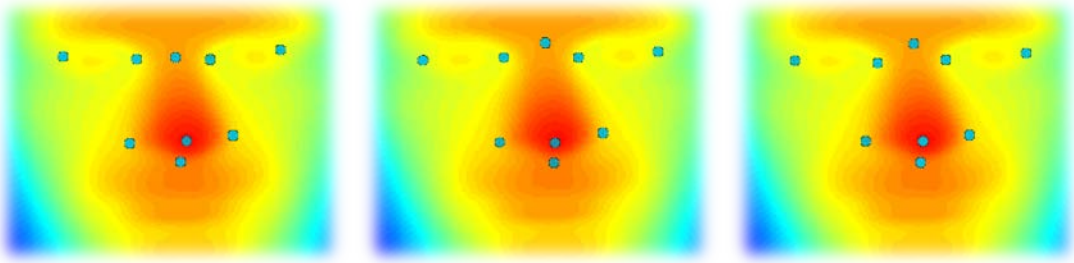


Figure 13. The intermediate steps of the fitting procedure. In particular: **(left)** the initial state, where the *nasion* extracted is wrong; **(center)** the intermediate state, where the landmarks are very close to the real landmarks; **(right)** the landmarks extracted.

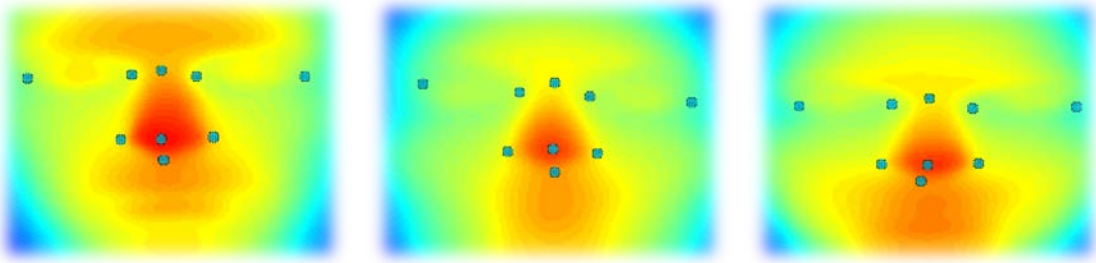
3. Results

The method was elaborated and implemented in Matlab®. Thirty-three faces of nine people with different facial expressions were scanned through a Minolta Vivid 910 and used for the experimentation. For each person, 7 facial expressions were taken, namely a straight expression and the 6 main emotional expressions, meaning anger, disgust, enjoyment, fear, sadness, and surprise, according to the theory of "basic emotions" of Ekman (Ekman, 1970; Ekman and Keltner, 1997). The scanned people were all Caucasian, male and female, from 20 to 40 years old.

After the scanning, the facial shells have been triangulated with a square mesh. The method was directly run on matrices collecting three-dimensional coordinates of the triangulated facial point clouds. The computing times could be divided in two types:

1. the computing time of the algorithm, which is about 15 seconds;
2. the processing time of the parametric surface fitting on the point cloud, which is about 10 seconds.

The results of the algorithm on nine faces belonging to different people are shown in Figure 14, while in Figure 15 there are the results of the algorithm on three faces of the same person performing three different facial expressions.



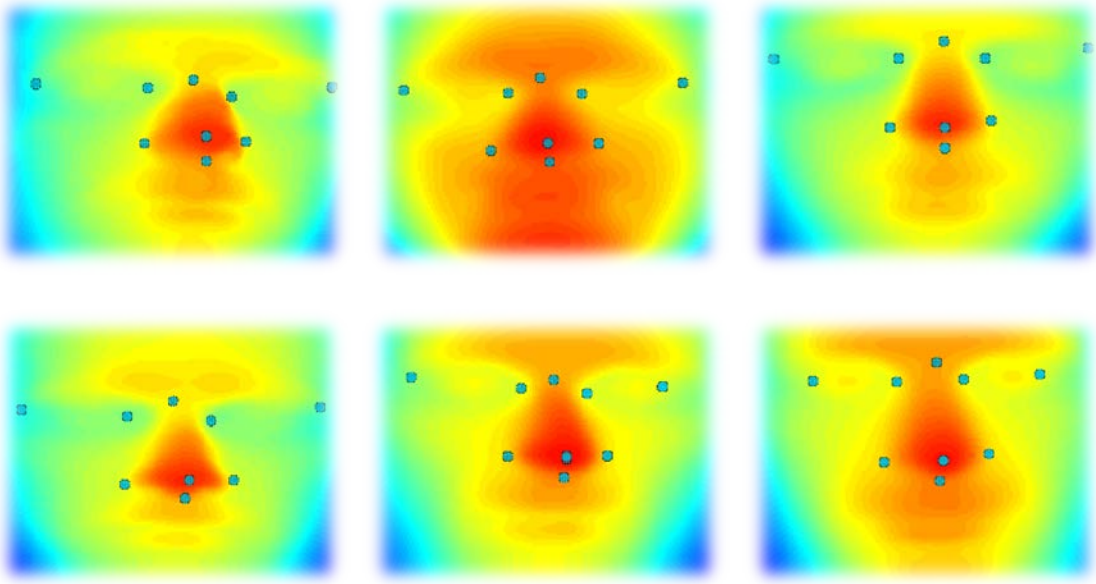


Figure 14. The extracted landmarks for nine of the thirty-three faces; the faces belong to different people.

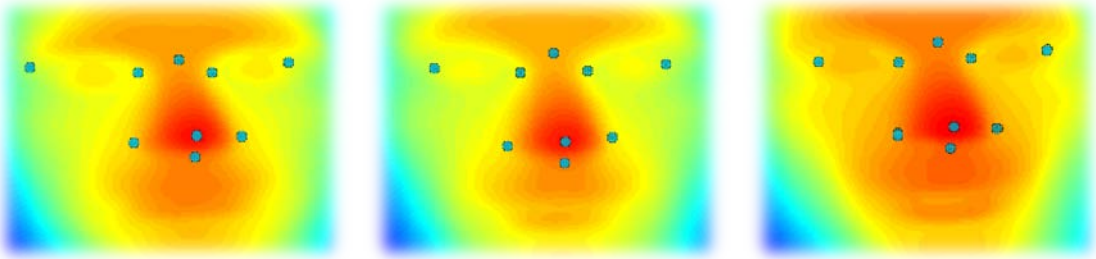


Figure 15. The extracted landmarks from faces belonging to the same person but with different expressions. In particular, the face expressions are sadness **(left)**, fear **(center)**, and surprise **(right)**.

To verify the goodness of the extracted landmarks, a brief statistical study was performed. Firstly, the landmarks of thirty-three faces were hand-detected from a plastic surgeon, so that we could compare them with the extracted ones. Subsequently, Euclidean distances between the correct landmarks and the respective points given by the new algorithm and the previous one are computed. However, in order to compare them, a normalizing operation is necessary; the idea was to normalize the distances by dividing them by the diagonal of the face bounding box. A bounding box is an invisible rectangular 3D box in which the face is somehow inscribed, as Figure 16 shows.

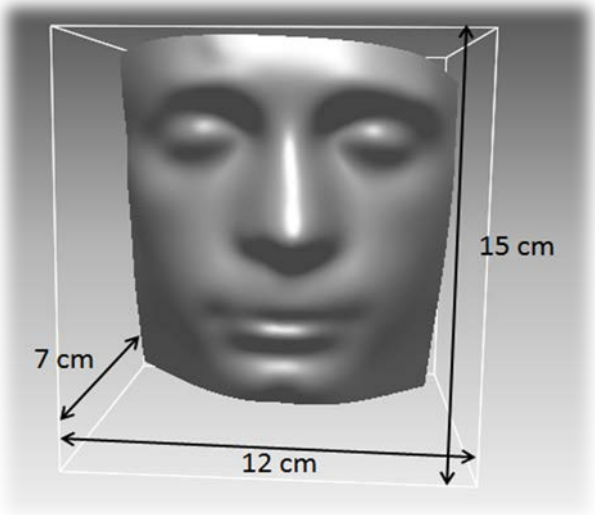


Figure 16. The measures of the bounding box of a face; face width is about 12 cm, height is about 15 cm and depth is about 7 cm. Applying the theorem of Pythagoras we can compute the diagonal of the bounding box ($\boldsymbol{D \approx 20}$ cm).

As shown in Figure 16, the sides of this box have a standard length, therefore, through a simple proportion, a normalization could be performed. The proportion is the following:

$$e : d_l = D : d_f,$$

where d_l is the distance between the correct landmark and the detected landmark, e is the normalized distance which must be computed (we can call it error), D is the diagonal of the bounding box of the face standard and d_f is the diagonal of the bounding box of the face where the landmarks are detected.

Once the normalized distances were computed, the sample mean μ and sample variance σ of these errors e_i were calculated:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} e_i, \qquad \sigma = \frac{1}{N-1}\sum_{i=1}^{N}(e_i - \mu)^2.$$

Since the diagonal of the bounding box of the standard face is given in centimetres, the errors and the mean will be in centimetres, while the variance will be in square centimetres.

Table 1 shows the errors on the detected landmarks with our previous geometric method, while Table 2 shows the errors on the detected landmarks with the statistical method.

Face	PN	SN	Left *alae*	Right *alae*	N	Left EN	Left EX	Right EN	Right EX
A disgust	0,000000	0,462384	1,377968	2,136248	4,781825	0,822180	2,965512	0,000000	3,198966
A	0,092863	0,273609	0,800119	1,107198	4,995738	0,189145	2,773832	0,387509	3,526048

enjoyment									
A fear	5,302633	6,037101	1,618077	2,046464	2,769072	nd	1,174162	2,095775	0,879138
A anger	0,000000	0,614204	0,831798	1,579049	5,792273	0,198926	3,301386	0,394830	3,251960
A straight	5,578411	6,032036	1,501161	1,220283	3,734194	nd	0,723541	nd	1,478767
A surprise	0,096215	0,690938	1,215083	1,814219	1,148803	0,000000	1,938825	0,812242	1,345096
A sadness	0,091621	0,660839	0,639964	1,342685	5,193155	0,854284	3,191624	0,000000	3,183644
B disgust	0,000000	0,095451	1,222273	1,429403	5,487973	0,552146	0,000000	0,415520	3,545822
B enjoyment	0,000000	0,000000	1,562388	1,411442	1,870658	0,406317	0,000000	0,407503	0,000000
B fear	0,000000	0,093815	1,385636	1,694579	1,083590	0,966525	0,160395	0,898588	0,699424
B straight	0,000000	0,159025	1,513443	1,529998	1,320302	0,565691	3,827539	0,627688	3,646372
B surprise	0,097797	0,000000	1,509481	1,420211	2,265855	0,837827	0,310812	0,655393	0,598336
C straight	3,404893	2,625729	1,084339	1,497613	1,197322	0,000000	0,836250	0,000000	0,523755
D disgust	0,000000	0,000000	1,247381	1,552350	0,267925	0,000000	0,778690	0,000000	0,554704
D enjoyment	0,000000	0,000000	1,046686	1,115114	0,682151	0,135822	0,949677	0,000000	1,516129
D fear	0,000000	0,094580	1,325889	1,374958	0,678932	0,000000	0,894567	0,000000	1,296277
D anger	0,136921	0,094441	1,400497	1,567730	0,703805	0,688054	0,897806	0,000000	0,950379
D straight	0,000000	0,000000	1,116347	1,298311	1,250866	0,419763	1,201163	0,138033	1,021320
D surprise	0,000000	0,000000	1,144163	1,439502	0,819671	0,000000	0,967131	0,000000	0,818975
E straight	0,192742	0,283780	1,142788	1,406316	1,128696	0,261565	1,390415	0,326844	0,763752
G straight	1,992700	1,088195	nd	nd	5,804329	0,000000	3,331233	2,548691	3,158839
H straight	1,553049	0,573259	nd	nd	5,060891	0,948022	4,051764	0,788119	2,902754
I disgust	0,000000	0,980362	2,150249	0,594056	nd	0,464803	0,827523	0,630289	3,688349
I enjoyment	5,644865	4,960352	2,043826	1,871001	2,727443	nd	1,717792	nd	1,196462
I fear	0,157299	0,572124	1,408674	2,035180	5,903441	0,764994	3,350541	0,264345	3,417757
I anger	0,000000	0,478551	2,066248	2,241016	nd	0,359521	3,467147	0,258844	3,494618
I straight	0,000000	0,429635	1,781803	2,035776	5,781285	0,615706	2,989192	0,498771	2,695455
I surprise	5,777090	6,386505	1,828581	1,971882	3,518407	nd	nd	nd	nd
I sadness	0,000000	1,035113	2,079713	2,111924	nd	0,376088	3,420194	0,484326	2,318734
L disgust	0,000000	0,000000	3,268479	1,530134	2,397277	0,648415	2,235085	0,263401	2,435169
L enjoyment	0,102192	0,000000	0,628435	1,014100	1,336682	0,481138	1,094927	0,847771	0,801169
L fear	0,000000	0,493908	0,789340	1,640910	2,978528	0,254219	2,673390	0,463005	2,472777
L anger	0,000000	0,109172	0,977452	1,861438	3,010985	2,500448	2,857788	0,576667	2,422708
L straight	5,298176	5,910212	1,922673	3,373614	3,174130	1,431659	1,294068	nd	3,776546
L surprise	0,000000	0,857835	0,914187	1,821694	nd	0,240388	3,943875	0,612449	3,007881
L sadness	0,210671	0,326624	0,963366	1,458054	2,512631	0,603883	2,793119	0,255596	2,287509
Mean	0,992504	1,178327	1,397309	1,633660	2,855589	0,518360	1,952313	0,489131	2,082160
Variance	3,863761	3,919188	0,286411	0,224900	3,474020	0,252777	1,596996	0,313147	1,410543

Table 1: the errors on the detected landmarks with the geometric method, their mean and variance. Expression "nd" means "non-detected".

Face	PN	SN	Left *alae*	Right *alae*	N	Left EN	Left EX	Right EN	Right EX
A disgust	0.119912	0.069305	0.272232	0.145103	0.562062	0.198433	0.638396	0.221982	1.368039
A enjoyment	0.115224	0.000000	0.330383	0.128060	0.691479	0.280069	0.609055	0.239248	0.937778

A fear	0.000000	0.261754	0.285558	0.167483	0.283187	0.118895	0.814031	0.223069	0.875557
A anger	0.068019	0.067840	0.317425	0.127756	0.373065	0.121045	1.048911	0.198529	0.354305
A straight	0.070978	0.142837	0.249122	0.092028	0.350696	0.295314	0.786398	0.072137	0.373377
A surprise	0.069216	0.178347	0.225010	0.000000	0.293966	0.069466	0.830696	0.166457	0.660707
A sadness	0.189888	0.316459	0.170815	0.276928	0.745121	0.225297	1.764005	0.120606	0.115695
B disgust	0.074126	0.127731	0.303298	0.185031	0.214433	0.451646	0.403834	0.465039	0.471244
B enjoyment	0.101334	0.146723	0.309229	0.000000	0.722083	0.371147	0.175087	0.478155	0.280046
B fear	0.000000	0.121855	0.104388	0.105569	0.630785	0.375556	0.169223	0.356599	0.289186
B straight	0.096023	0.078173	0.287198	0.126602	0.497605	0.772505	1.452381	0.679804	0.970747
B surprise	0.104497	0.074759	0.231582	0.000000	0.311847	0.607073	0.730034	0.454929	0.810364
C straight	nd	nd	nd	nd	nd	nd	nd	nd	nd
D disgust	0.113053	0.206538	0.206485	0.000000	0.857784	0.356454	0.828676	0.453860	0.972222
D enjoyment	0.000000	0.070892	0.170826	0.172387	0.436631	0.113006	0.739948	0.070688	1.348366
D fear	0.071371	0.000000	0.305307	0.122839	0.442739	0.241222	1.285168	0.237685	1.255865
D anger	0.111832	0.136028	0.301369	0.000000	1.158026	0.514454	1.398116	0.651338	1.536640
D straight	0.143417	0.140926	0.104247	0.000000	0.728403	0.227235	0.690899	0.432799	1.096160
D surprise	0.073178	0.131339	0.224683	0.134657	0.358332	0.185365	0.605188	0.105499	0.680972
E straight	0.073169	0.244127	0.196156	0.099137	1.148158	0.402856	0.967758	0.256571	1.497062
G straight	0.068192	0.597205	0.341056	0.000000	0.509181	0.000000	0.226649	0.304350	1.012444
H straight	0.085970	0.707965	0.288289	0.086060	0.869624	0.070986	0.377199	0.292100	0.676612
I disgust	0.000000	0.150083	0.185107	0.154308	0.368813	0.096839	0.342839	0.234596	0.480783
I enjoyment	0.223603	0.000000	0.277093	0.143357	0.339257	0.151070	0.516459	0.084974	1.132872
I fear	0.092795	0.142259	0.180820	0.000000	0.170914	0.197647	0.753700	0.183487	1.348624
I anger	0.000000	0.159891	0.176740	0.000000	0.849104	0.403738	0.154000	0.420008	0.183185
I straight	0.088433	0.156265	0.374561	0.159684	0.437838	0.217423	0.357316	0.244613	1.247267
I surprise	0.070314	0.237623	0.365620	0.000000	0.730362	0.313089	0.304874	0.293685	0.487215
I sadness	0.089571	0.264788	0.310307	0.000000	0.558094	0.115472	0.460715	0.226056	1.031738
L disgust	0.000000	0.076241	0.180025	0.162794	0.389798	0.656511	1.587377	0.297684	0.790354
L enjoyment	0.000000	0.078773	0.284515	0.000000	0.576306	0.313976	1.014045	0.245200	0.872978
L fear	0.000000	0.301959	0.250947	0.294727	0.401598	0.470922	0.988432	0.457919	0.184601
L anger	0.116365	0.119067	0.416612	0.197801	0.330016	0.384196	1.179574	0.600284	1.432810
L straight	0.000000	0.403928	0.183385	0.209943	0.728083	0.849994	1.359453	0.797295	0.884264
L surprise	0.067935	0.533888	0.324656	0.117573	0.903614	0.683912	1.361990	0.532576	1.290946
L sadness	0.000000	0.120888	0.433828	0.148644	0.167776	0.404452	1.049219	0.267900	1.249883
Mean	0.071383	0.187613	0.261968	0.101671	0.546765	0.321636	0.799190	0.324792	0.862883
Variance	0.003146	0.025970	0.006400	0.007327	0.066118	0.043483	0.188688	0.032514	0.175168

Table 2. Errors on the detected landmarks with the statistical method, their mean and variance. Expression "nd" means "non-detected".

In Figures 17, 18, and 19 are shown the graphs of sample mean and standard deviation ($\pm\sqrt{\sigma}$) for both the methods.

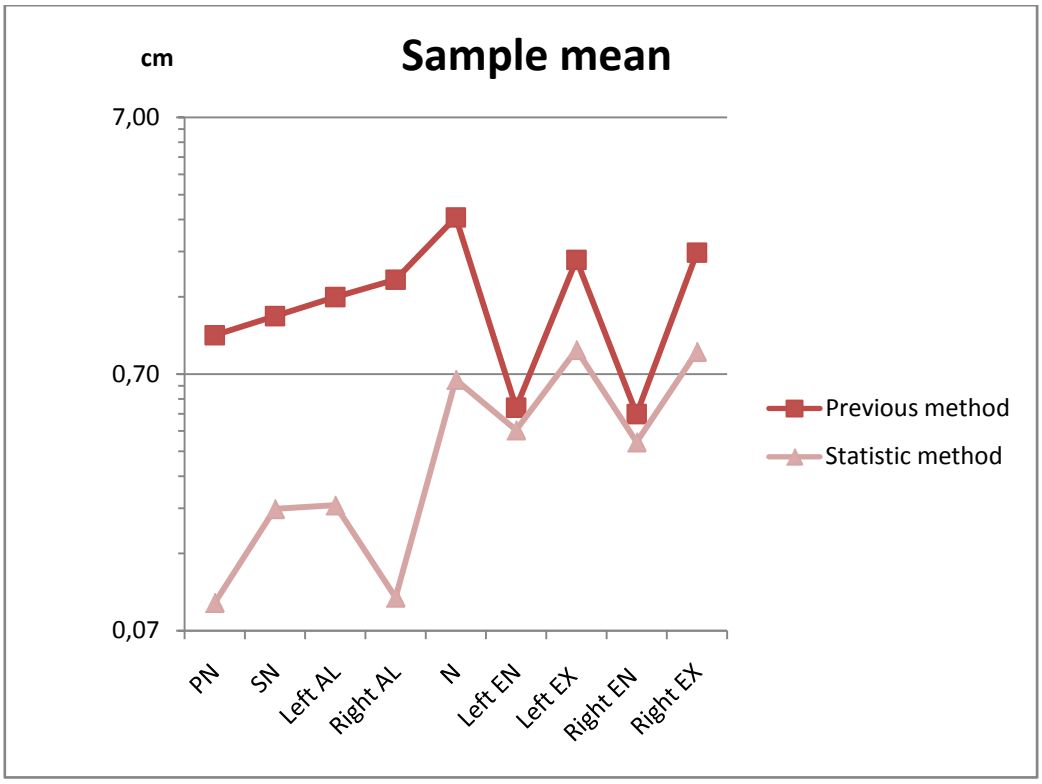


Figure 17. Graph of sample mean $\boldsymbol{\mu}$ for the previous geometric method and the new statistical method here proposed.

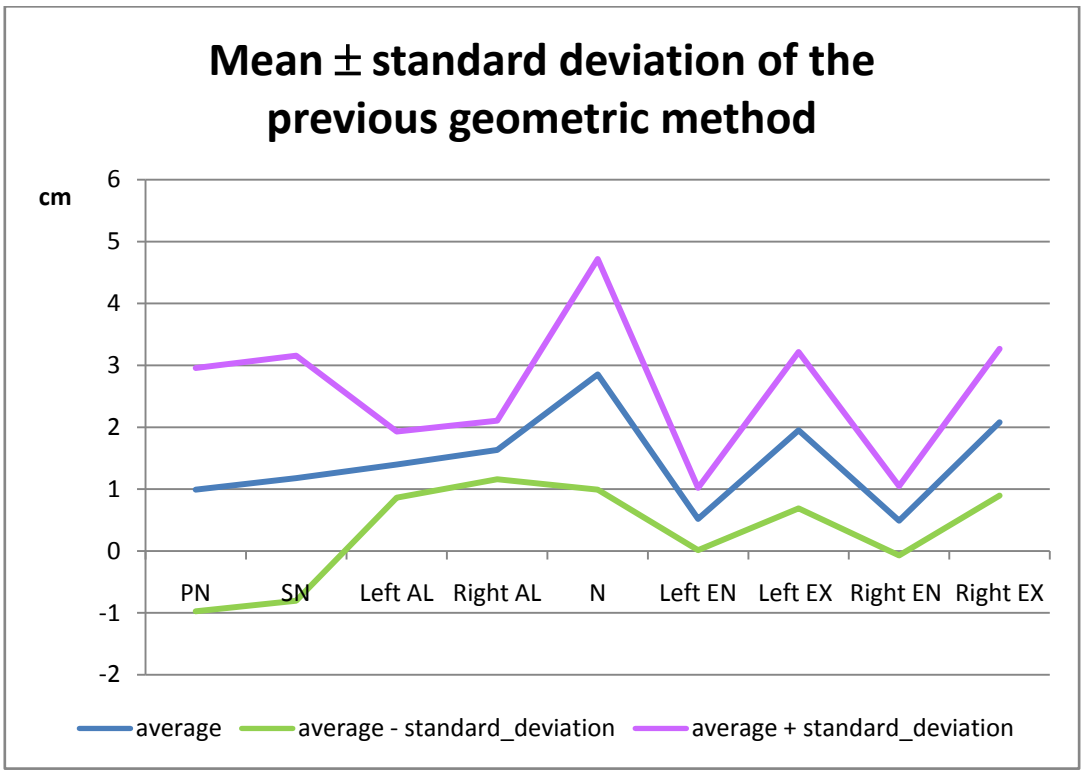


Figure 18. Graph of sample mean (blue line in the middle) and standard deviation (lilac and green lines above and below), equal to $\pm\sqrt{\boldsymbol{\sigma}}$ for our previous geometrically-based method.

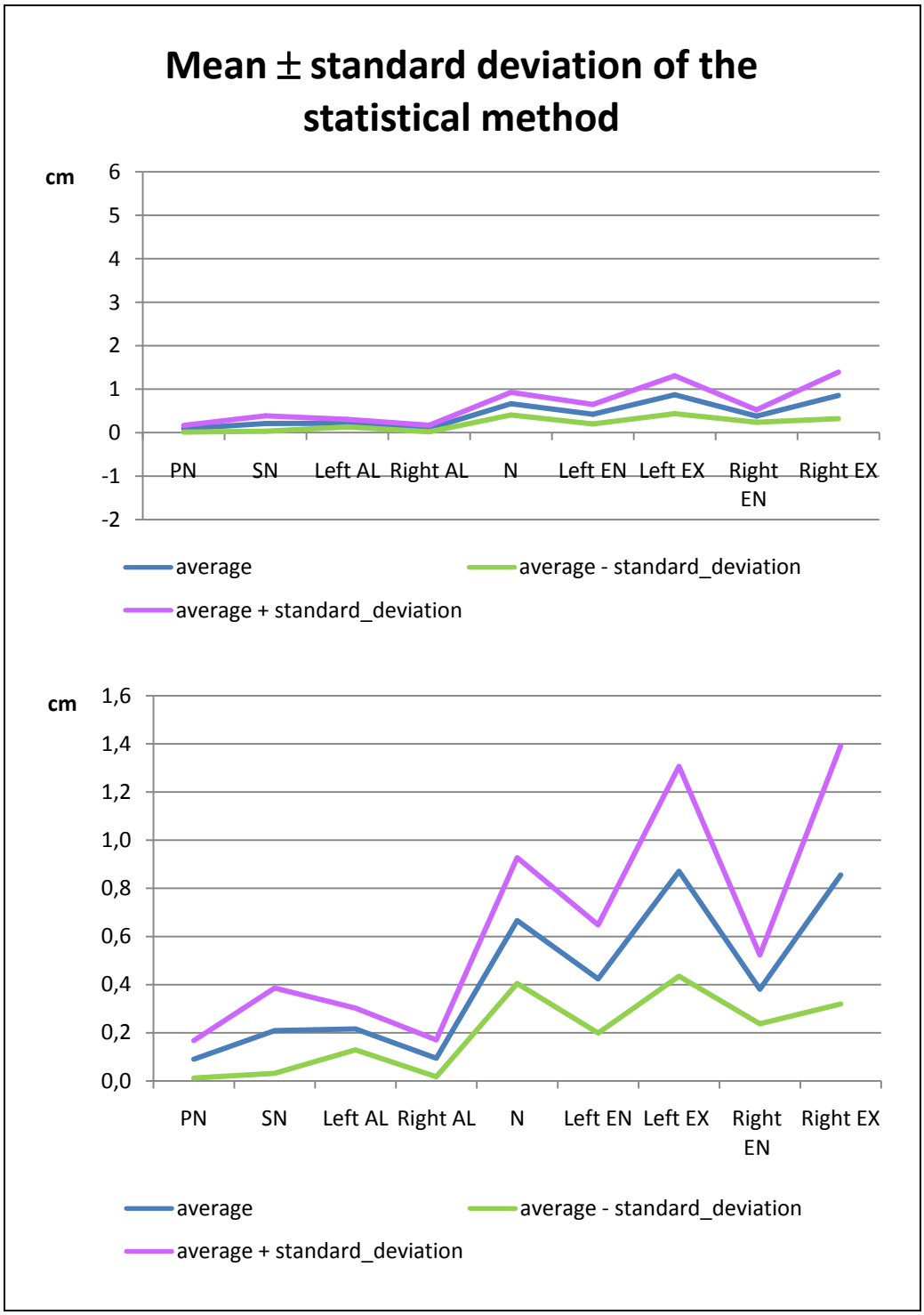


Figure 19. Graphs of sample mean (blue line in the middle) and standard deviation (lilac and green lines above and below), equal to $\pm\sqrt{\sigma}$, for the new statistical method. In the figure above the axis unit is kept equal to the Figure 16, in order to have a visual comparison between the two methods, while the figure below shows the same behavior with an axis unit more suitable for the numerical values of this method.

The graphical representations show that in the new statistical method the overall sample mean is always lower than the mean of the previous method. The lowest sample mean values concern the right *endocanthion* (0.489131 cm) for the previous method and the *pronasal*

(0.090042 cm) for the statistical one; the highest values are reached for the extraction of the *nasion* (2.855589 cm) in the geometrical method and for the left *exocanthion* (0.870817 cm) in the new one. This global behavior of the sample mean shows that the new geometrically- and statistically-based method is an improvement of the previous method, which only relies on Differential Geometry.

Another result feature to point out is that, while in the previous method the double and symmetrical landmarks, i.e. *alae*, *endocanthions*, and *exocanthions*, gained similar results, meaning the right and left *alae*, right and left *endocanthions*, and right and left *exocanthions* reached comparable sample mean numerical values, the same cannot be said for the statistically-based method, in which results of the double landmarks are not so comparable.

Finally, it can be seen from both Figures 18 and 19 that for *alae* and *endocanthions* the numerical value of the standard deviation is kept low. This is a behaviour which concerns both the methods, although in the statistical method the numerical values are lower. Since it was a good result for the previous method, it is important that this trend is maintained also in the new one.

The quality of results could also be recorded with other graphical representations of the results, i.e. scatter plots and distribution functions for both the methods, shown in Figures 20, 21, 22, and 23. The scatter plot is a by-points representation on the Cartesian plane of the positions of the obtained landmarks. It is likely to put the found points in the same reference system, where the origin stands for the correct landmark position. The scattered points on the plane pretend to show the position of the points obtained by our algorithms, in particular the direction and the distance from the correct landmark.

The distribution function shows how many landmarks (n, on the ordinate axis) takes a particular distance value ($|e|$, on the abscissa axis). For simplicity sake, the distribution is not continuous, but discrete. To obtain it, we discretized the set of error distances splitting them up into 50 short ranges for both the old and the new method results. A minimum and a maximum distance has been chosen: the minimum is equal to 0 and concerns the landmarks obtained by the algorithms which are in the same position of the correct landmarks; the maximum is equal to 5 cm and concerns the points whose distances from the correct landmark is equal or greater than 5 cm. Then, we subdivided the range between 0 and 5 in short intervals all equal to 0.1 cm, thus obtaining a discretization of the range of distances in 50 short ranges. Then, a

graphical representation was done with the split-up distance range from 0 to 5 on the abscissa axis and the number of occurrences of the distance ranges on the ordinate.

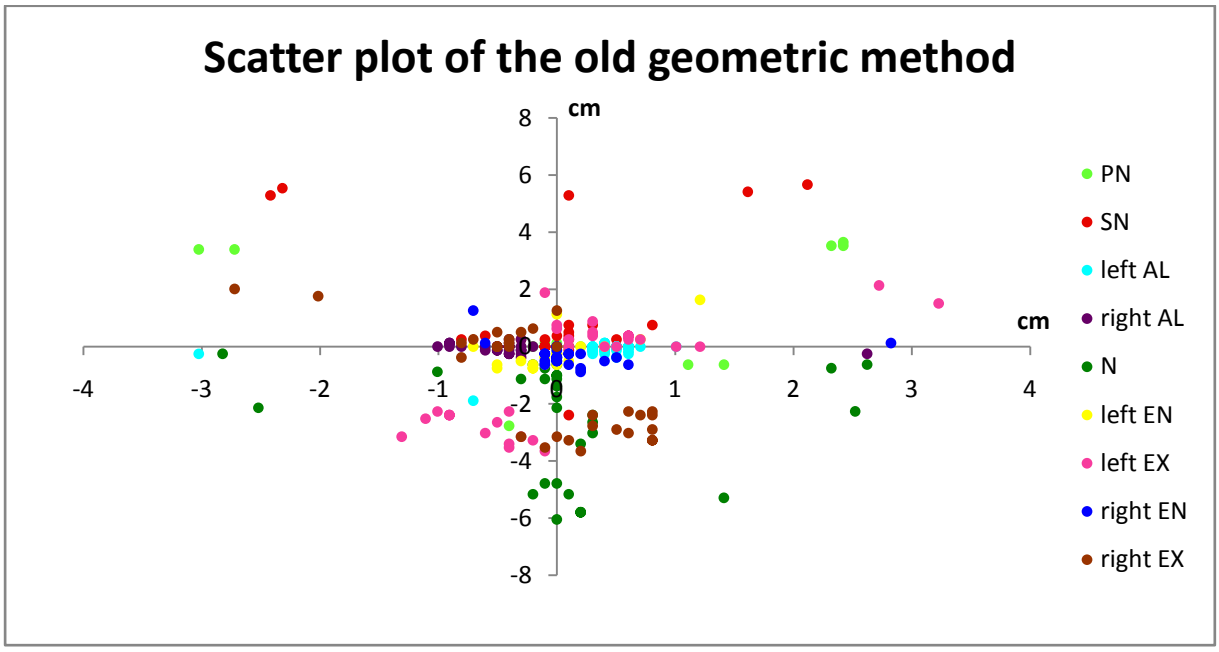


Figure 20. Scatter plot of the previous method results. The origin of the axis represents the correct location of every landmark. The scattered points are the positions of the landmarks obtained with the previous geometrically-based algorithm. In this representation, the direction of the positioning of the obtained landmark is kept equal to the real one, while the absolute value of the distance the bi-dimensional distance on x and y axis, namely an approximation of the 3D distance.

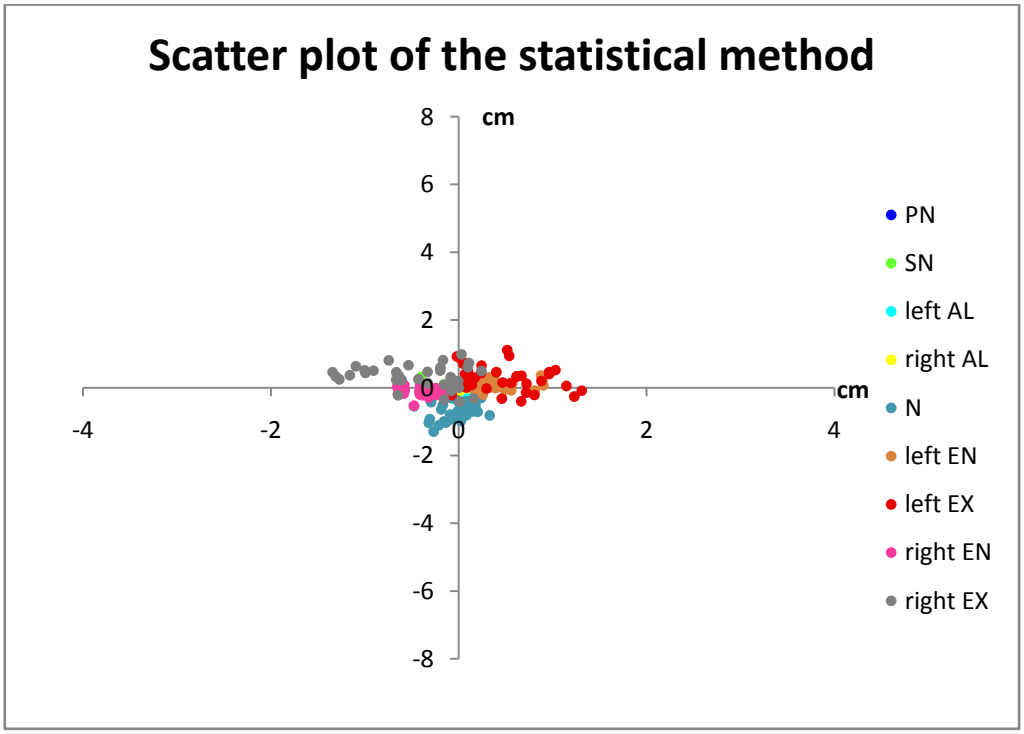


Figure 21. Scatter plot of the new statistical method results. The origin of the axis represents the correct location of every landmark. The scattered points are the positions of the landmarks obtained with our new algorithm.

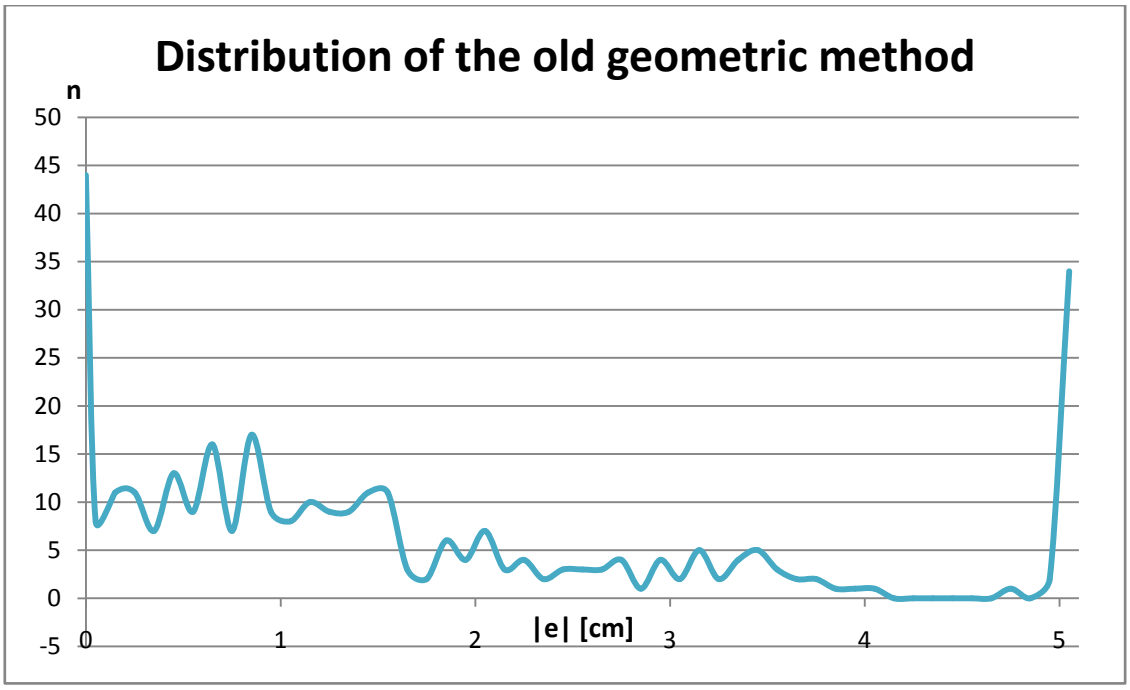


Figure 22. Distribution function for the results of the old method. The discrete distribution function shows how many landmarks ($\boldsymbol{n}$, on the ordinate axis) takes a particular distance value ($|\boldsymbol{e}|$, on the abscissa axis).

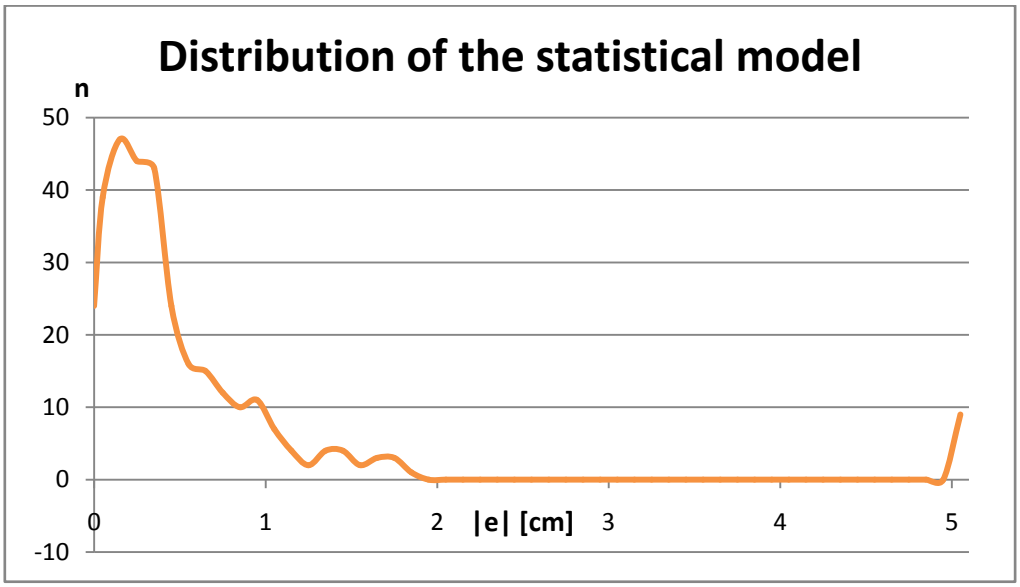


Figure 23. Distribution function for the results of the statistical method.

Scatter plots clearly show that the new method gained more precise landmark positioning, namely the obtained landmarks are contained in a neighbourhood of $2 \times 2\ cm^2$ of the real landmark, while in the old method the neighbourhood was wider. The distributions give clarification of the number of landmarks whose distance "between true and obtained" is short or large. In particular, while the landmark distances are well distributed in the range $0 \leq |e| \leq 5$ for the previous method, for the new method landmark distances are concentrated

in the range $0 \leq |e| \leq 2$, with a peak in correspondence of 0,1 cm, whose related number of landmark is 48. This means that most of the obtained landmarks are approximately very close to the real ones.

4. Conclusions

In this study we proposed a new statistical and geometrically-based landmarking method for 3D facial scans. In particular, we improved the geometrical conditions on descriptors taken from Differential Geometry that we used in our previous work for building up a new pose-independent algorithm for computing candidate points. Then, we fitted a statistical model based on Procrustes Analysis and PCA on the candidate points for a precise landmark extraction. Due to its structure, this method can be considered as both geometrical and statistical. The correctness of results was confirmed by a plastic surgeon and discussed through a brief statistical study.

5. References

ALKER, M., FRANTZ, S., ROHR, K. AND STIEHL, H. S. (2001) "Improving the Robustness in Extracting 3D Point Landmarks from 3D Medical Images Using Parametric Deformable Models", *Lecture Notes in Computer Science* 2208: 582-590.

CALIGNANO, F. (2009) "Morphometric methodologies for bio-engineering applications", PhD Degree Thesis, Politecnico di Torino, Department of Production Systems and Business Economics.

COOTES, T., TAYLOR, C., COOPER, D. AND GRAHAM, J. (1995) "Active shape models – their training and application", *Computer Vision and Image Understanding* 61(1): 38-59.

COOTES, T. AND TAYLOR, C. (2001) "Statistical Models of Appearance for Computer Vision", Technical Report, Wolfson Image Analysis Unit, Imaging Science and Biomedical Engineering, University of Manchester.

COOTES, T., TAYLOR, C., KANG, H. AND PETROVIC, V. (2005) "Modeling Facial Shape and Appearance", Book Chapter of LI, S.Z. AND JAIN, A.K., *Handbook of Face Recognition*, Springer, 2005: 39-63.

DALAL, A. B. AND PHADKE, S. R. (2007) "Morphometric analysis of face in dysmorphology", *Computer Methods and Programs in Biomedicine* 85: 165-172.

DELLA CROCE, U., CAPPOZZO, A., AND KERRIGAN, D. C. (1999) "Pelvis and lower limb anatomical landmark calibration precision and its propagation to bone geometry and joint angles", *Medical and Biological Engineering and Computing* 37(1): 155–61.

DRYDEN, I. AND MARDIA, K., *Statistical Shape Analysis*, Wiley, 1998.

EKMAN, P. (1970) "Universal Facial Expressions of Emotions", *Califormia Mental Health Research Digest* 8(4): 151-158.

EKMAN, P. AND KELTNER, D., *Facial Expressions of Emotions*, Lawrence Erlbaum Associates Publisher, Mahwah, New Jersey, 1997.

FRANTZ, S., ROHR, K. AND STIEHL, H. S. (1998) "Multi-Step Procedures for the Localization of 2D and 3D Point Landmarks and Automatic ROI Size Selection", *Lecture Notes in Computer Science* 1406: 687-703.

FRANTZ, S., ROHR, K. AND STIEHL, H. S. (1999) "Improving the Detection Performance in Semi-automatic Landmark Extraction", *Lecture Notes in Computer Science* 1679: 253-262.

FRANTZ, S., ROHR, K. AND STIEHL, H. S. (2000) "Localization of 3D Anatomical Point Landmarks in 3D Tomographic Images Using Deformable Models", *Lecture Notes in Computer Science* 1935: 492-501.

FRANTZ, S., ROHR, K. AND STIEHL, H. S. (2005) "Development and validation of a multi-step approach to improved detection of 3D point landmarks in tomographic images", *Image and Vision Computing* 23(11): 956-971.

GRIFFIN, F.M, MATH, K., SCUDERI, G.R., INSAL, J.N., AND POILVACHE, P.L. (2000) "Anatomy of the epicondyles of the distal femur MRI analysis of normal knees", *The Journal of Arthroplasty* 15(3): 354–359.

KOENDERINK, J. J. AND VAN DOORN, A. J. (1992) "Surface shape and curvature scales", *Image and Vision Computing* 10(8): 557-564.

LIU X., KIM, W., AND DRERUP, B. (2004) "Foot 3D characterization and localization of anatomical landmarks of the foot by FASTscan", *Real-Time Imaging* 10(4): 217–28.

MAHOOR, M. H. AND ABDEL-MOTTALEB, M. (2009) "Face recognition based on 3D ridge images obtained from range data", *Pattern Recognition* 42: 445-451.

MathWorks http://www.mathworks.it/help/toolbox/stats/procsrustes.html.

MAUDGIL, D.D., FREE, S.L., SISODIYA, S.M., LEMIEUX, L., WOERMANN, F.G., FISH, D.R., AND SHORVON S.D. (1998) "Identifying homologous anatomical landmarks on reconstructed magnetic resonance images of the human cerebral cortical surface", *Journal of Anatomy* 193(4): 559–71.

MENA-CHALCO, J. P., MACÊDO, I., VELHO, L., AND CESAR, R. M. JR. (2008) "PCA-based 3D Face Photography", *XXI Brazilian Symposium on Computer Graphics and Image Processing*: 313–320.

MUTSVANGWA, T. AND DOUGLAS, T. S. (2007) "Morphometric analysis of facial landmark data to characterize the facial phenotype associated with fetal alcohol syndrome", *Journal of Anatomy* 210: 209-220.

NAIR, P. AND CAVALLARO, A. (2009) "3-D Face Detection, Landmark Localization, and Registration Using a Point Distribution Model", *IEEE Transactions on Multimedia* 11(4): 611-623.

[a] PERAKIS. P., PASSALIS, G., THEOHARIS, T., TODERICI, G. AND KAKADIARIS, L. A. (2009) "Partial Matching of Interpose 3D Facial Data for Face Recognition", *IEEE 3rd International Conference on Theory, Applications, and Systems*: 1-8.

[b] PERAKIS. P., PASSALIS, G., THEOHARIS, T., AND KAKADIARIS, L. A. (2009) "Automatic 3D Facial Region Retrieval from Multi-pose Facial Datasets", *Eurographics Workshop on 3D Object Retrieval*: 37-44.

STEGMAN, M, AND GOMES, D. (2002) "A Brief Introduction to Statistical Shape Analysis", Technical Report, Informatics and Mathematical Modelling, Technical University of Denmark.

THEODORIS, S. AND KOUTROUMBAS, K., *Pattern Recognition – Third Edition*, Academic Press, Orlando, Florida, 2006.

VAN SINT JAN, S., *Color Atlas of Skeletal Landmark Definitions; Guidelines for Reproducible Manual and Virtual Palpations*, Churchill Livingstone Elsevier, Edinburgh, 2007.

VEZZETTI, E. AND MARCOLIN, F. "Three-dimensional Face Morphology Analysis: Soft-Tissue Landmark Formalization", under review.

WÖRZ, S. AND ROHR, K. (2005) "Localization of anatomical point landmarks in 3D medical images by fitting 3D parametric intensity models", *Medical Image Analysis* 10(1): 41-58.

YANG, J., LING, X., LU, Y., WEI, M., AND DING, G. (2001) "Cephalometric image analysis and measurement for orthognathic surgery", *Medical and Biological Engineering and Computing* 39(3): 279–84.

YAHARA, H., HIGUMA, N., FUKUI, Y., NISHIHARA, S., MOCHIMARU, M., KOUCHI, M. (2005) "Estimation of anatomical landmark positions from model of 3-dimensional foot by the FFD method", *Systems and Computers in Japan* 36(6): 26–38.