3D geometry-based automatic landmark localization in presence of facial occlusions

3D Geometry-based Automatic Landmark Localization in Presence of Facial Occlusions

Enrico Vezzetti, Federica Marcolin, Stefano Tornincasa, Luca Ulrich, and Nicole Dagnes
Department of Management and Production Engineering
Politecnico di Torino, corso Duca degli Abruzzi 24, 10129 Torino, Italy

Abstract

This study proposes a novel automatic method for facial landmark localization relying on geometrical properties of 3D facial surface working both on complete faces displaying different emotions and in presence of occlusions. In particular, 12 descriptors coming from Differential Geometry including the coefficients of the fundamental forms, Gaussian, mean, principal curvatures, shape index and curvedness are extracted as facial features and their local geometric properties are exploited to localize 13 soft-tissue landmarks from eye and nose areas. The method is deterministic and is backboned by a thresholding technique designed by studying the behaviour of each geometrical descriptor in correspondence to the locus of each landmark. Occlusions are managed by a detection algorithm based on geometrical properties which allows to proceed with the landmark localization avoiding the covered areas. Experimentations were carried out on 3132 faces of the Bosphorus database and of a 230-sized internal database, including expressive and occluded ones (mouth, eye, and eyeglasses occlusions), obtaining 4.75 mm mean localization error.

Keywords

3D Face; Face Analysis; Landmark Localization; Differential Geometry; Feature Extraction.

1 Introduction

Automatic landmarking is a key process for different Face Analysis applications such as Face Identification, Face Verification, and Face Expression Recognition. Locating landmarks means building an alternative compact representation of the face, which could be easily portable and replicable. A set of landmark is a facial map which, together with geometrical and texture features, contribute to determine typical facial traits, so that the purpose of subject or emotion recognition could be achieved.

Landmarks were defined by Farkas [1] [2] and validated by other Anthropometry handbooks such as Swennen at al.'s [3]. Dense facial landmark distributions allow high-level facial description and repeatability, and are core for biometric applications, as the face model is an alternative to iris and fingerprint. Despite some eminent bidimensional solutions [4], in terms of data, 3D and 2D+3D (or even 2.5D [5]) are recently more addressed than 2D alone, as the third dimension improves accuracy [6] and allows to overcome problems such as pose, different lightning conditions [7], make-up, and camouflages. The application of geometry to RGB channels of facial 2D images has also been recently investigated, thus producing an hybrid 2D+3D approach [8].

Various research groups recently worked on automatically localizing facial 3D landmarks by exploiting geometrical properties of the facial surface, in particular adopting mean, Gaussian, principal curvatures, and Koenderink and vanDoorn's shape index and curvedness [9] as features.

Bagchi et al. developed a system based on **mean** (H) and **Gaussian** (K) **curvatures** to extract eye and nose points from faces in different poses relying on convexity and concavity information. A so-called HK classification of the points is obtained. Experimentations, carried out on the FRAV3D database, gave acceptable results, as eye corners were correctly detected in 72.2% cases and nose tip in 98.8% cases [10]. Similarly, **mean** and **Gaussian curvature** information were exploited by Boukamcha et al. for extracting nose tip, nose corners, nose base, and eye corners on faces with different expressions on both FRGC 1.0 and FRGC 2.0 databases. After a facial segmentation phase, the sign of H and K curvatures is studied for each region and landmark points are localized accordingly. The method worked properly on 99% of the tested facial images [11]. De Giorgis et al. adopted the **Gaussian curvature** for facial characterization to detect 13 landmarks, which "occur at important maxima and minima" of this geometrical descriptor. Experiments were run on the Face Warehouse database with 111 neutral expressions frontal-view faces. Nose points achieved 90% accuracy within 11 mm; the same accuracy was obtained with 3 mm limit for eye points and with 8 mm for mouth ones [12]. Similarly, Li et al. exploited **Gaussian curvature** point-by-point behaviour to localize nose tip. The method was not training-based nor model-based and its testing on the GavabDB database achieved a 100% localization accuracy in 10 mm precision [13].

Canavan and Yin proposed a method for landmark detection and tracking relying on the **shape index** in the scale proposed by Dorai and Jain [14]. In particular, they introduced the so-called Shape Index based Statistical Shape Model (SI-SSM) which made use of the global shape of the facial data as well as local patches. The method models both the global shape of landmarks and the local curvatures from patches around the landmark points. Mean squared errors were 3.2 one-point spacings (≈1.6 mm) on the BU-4DFE database and 2.9 spacings (≈1-45 mm) on the BP4D-Spontaneous database (with posed and spontaneous expressions, respectively) [15]. The same method was also tested on BU-3DFE, FRGC2.0, and Eurecom Kinect Face databases [16]. Li et al. developed a new landmark localization method called incremental Parallel Cascade of Linear Regression (iPar–CLR). Even if, in this case, the landmarking does not specifically rely on geometrical properties, the **shape index** was used to study the local facial 3D shape around each extracted landmark with the final aim of facial expression recognition [17]. Instead, the **shape index** was adopted as descriptor for developing a 3D landmark localization methodology robust to expression variations by Perakis et al. [18]. The candidate landmarks are identified and labeled by matching them with a Facial Landmark Model (FLM) of facial anatomical landmarks. Experimentations were carried out on the FRGCv2 and UND databases showing a mean error under 6.3 mm, with standard deviation under 2.6 mm on all tested faces. In particular, the mean error was under 10 mm in 90.4% cases and the facial side was correctly estimated on over 98.9% faces. Then, they improved their methodology by proposing a feature fusion technique for automatic landmarking (based on the shape index). In particular, "the proposed framework maps each feature into a similarity score and combines the individual similarity scores into a resultant score, used to select the optimal solution for a queried landmark" [19]. The method was tested on the same datasets and achieved a landmark localization error within a range of 3.5–5.5 mm. Lei et al. extracted 16 ear fiducial points by exploiting the descriptiveness of **curvedness** on a novel 3D ear tree-structured graph (ETG) model. On the UND F, UND45LR, UND60LR, and UND J2 datasets, the mean localization errors of the triangular fossa and incisure intertragica were smaller than 4 mm, thus obtaining a 100% detection accuracy [20].

Creusot et al. presented an automatic landmarking machine learning framework for 3D objects, applied in this study on faces. A 'pool' of 8 local shape descriptors, including **principal**, **mean**, **Gaussian curvatures**, the **shape index** and **curvedness**, is calculated and relative scores evaluated in order to compute landmark score functions; only the most successful among them will be selected and identified as the final landmark model. Saliency maps supported the localization of landmarks. The algorithm was tested on BFM and FRGC databases and obtained landmark locations were compared to those manually-annotated [21]. A similar method adopting the same geometrical descriptors was tested on FRGC v2 and Bosphorus 3D face datasets to localize 14 landmarks [22]. Gilani et al. used **Gaussian**, **mean**, and **principal curvatures**, and **curvedness** to extract morphologically significant facial landmarks on 6507 faces from FRGCv2 and BU3DFE databases. In particular, they combined level set curve evolution with geometric speed functions in order to automatically extract effective seed points for dense correspondence. Then, a deformable model is built and fitted to unseen faces to transfer

correspondences and, subsequently, landmarks. The proposed algorithm localized the 85 points available for the BU3DFE dataset on the whole database within an error of 5.85 ± 4.26 mm [23]. **Mean** and **Gaussian curvatures**, **shape index**, and **curvedness** were used for geometrically describing nose area for automatically detecting nose tip. A 99.9% nose tip true acceptance rate with 6.71% false acceptance rate was achieved on the FRGC v2.0 dataset when mean curvature and shape index along with curvedness were used as the input to the SVM classifier [5].

The method proposed in this study is a deterministic technique for automatically extracting 13 soft-tissue eye and nose landmarks from standard, expressive, and occluded 3D faces. Local properties of the 12 geometrical descriptors are taken into consideration to design a thresholding methodology which, for each landmark, identifies the region of interest, narrows it by applying different geometrical conditions, and finally extracts the landmark. The method is robust to facial expressions and occlusions involving mouth and eyes, including glasses. When the algorithm processes an occluded face, only the landmarks of the non-occluded parts are taken into consideration.

The innovativeness of this method with respect to our previous landmarking ones [24] [25] [26] relies on some newly-introduced elements and advances: i) robustness to different facial expressions; ii) robustness to occlusions; iii) use of the whole Bosphorus database as testing dataset; iv) overall revision and improvement of the method with tangible enhanced results.

2 Method

Figure 1 shows the global process of landmark localization, which extracts the landmarks one-by-one. Single landmark localization procedures are in blue blocks, while correctness checks (of the previously localized landmarks) are red-coloured. The facial occlusion algorithm is green, as it is a separate function which allows to proceed on the localization of the points avoiding the covered areas. Figure 2 explains the meaning of the landmarks presented in this chart.

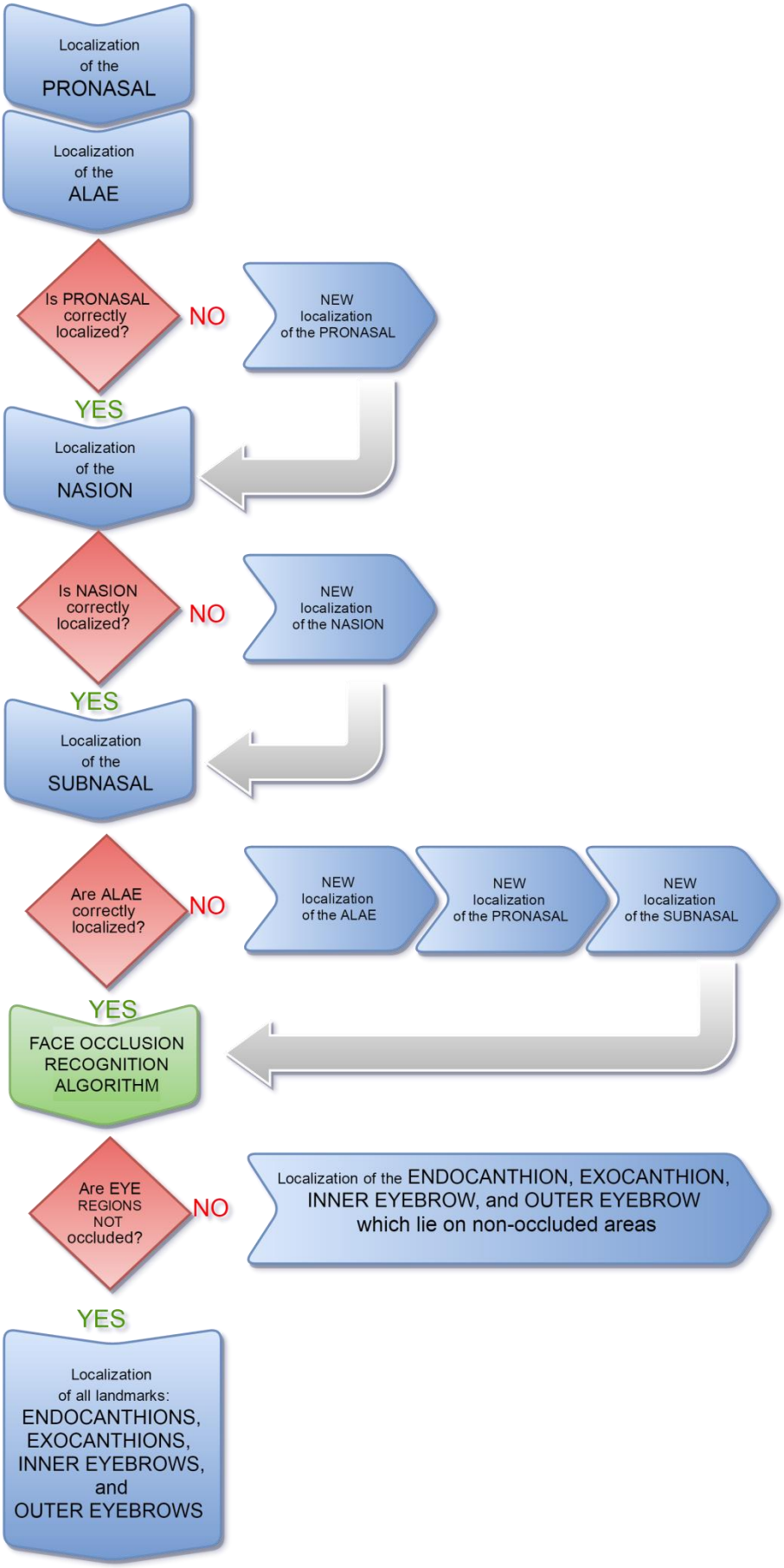


Figure 1. Diagram showing every step of the global landmark extraction algorithm.

Concerning pronasale, nasion, and alae localizations, a double check is made to verify whether these points have been correctly localized, otherwise a new localization process starts with a new region of research, while geometrical conditions remain unchanged. The reason why this double check is made is that, when cases with occlusions happen, occlusions may vary some conditions. In the case of the pronasale, this usually happens when the occluded area is close to the nose. Thus, the algorithm checks whether the pronasale vertical coordinate lies within left and right alae; otherwise the region of research is changed and the localization process restarts. Similar double checks are made for the nasion and the alae.

When the localization of pronasale, nasion, alae, and subnasale are complete, the algorithm automatically checks whether occlusions are present. The occlusion detection methodology relies on geometrical properties via thresholding technique. If occlusions are present, the algorithm deletes the landmarks extracted on the occluded parts, so that only the landmarks present on the uncovered parts of the face are taken into consideration. Result analysis is undertaken only on the landmarks selected in this phase.

2.1 Geometrical descriptors

Differential Geometry is a branch of Geometry which allows differentiability/derivability of surfaces, so that point-by-point derivatives, and, subsequently, curvatures and other descriptors, are computable. A detailed previous study of our research group [27] provides definitions and in-depth explanations of the geometrical descriptors adopted in this study. They are briefly listed here below together with the formulas implemented in the method here proposed.

The three coefficients of the first fundamental form E, F, G and the three coefficients of the second fundamental form e, f, g are:

$$E = 1 + h_x^2, \tag{1.1}$$
$$F = h_x h_y, \tag{1.2}$$
$$G = 1 + h_y^2, \tag{1.3}$$
$$e = \frac{h_{xx}}{\sqrt{1+h_x^2+h_y^2}}, \tag{1.4}$$
$$f = \frac{h_{xy}}{\sqrt{1+h_x^2+h_y^2}}, \tag{1.5}$$
$$g = \frac{h_{yy}}{\sqrt{1+h_x^2+h_y^2}}, \tag{1.6}$$

where h_x,h_y,h_{xx},h_{yy},h_{xy} are the first and second (and mixed) derivatives of the surface h with respect of x and y. h stands as the facial surface in these formulations; nonetheless, in the algorithm, h is given by a matrix in which every point describes the facial depth map. In other words, h represents the z-coordinates.

Gaussian and mean curvatures are given by the formulas:

$$K = \frac{h_{xx}h_{yy}-h_{xy}^2}{\left(1+h_x^2+h_y^2\right)^2}, \tag{1.7}$$
$$H = \frac{\left(1+h_x^2\right)h_{yy}-2h_x h_y h_{xy}+\left(1+h_y^2\right)h_{xx}}{\left(1+h_x^2+h_y^2\right)^{3/2}}. \tag{1.8}$$

Principal curvatures rely on Gaussian and mean curvatures (and viceversa):

$$k_1 = H + \sqrt{H^2 - K}, \quad (1.9)$$
$$k_2 = H - \sqrt{H^2 - K}. \quad (1.10)$$

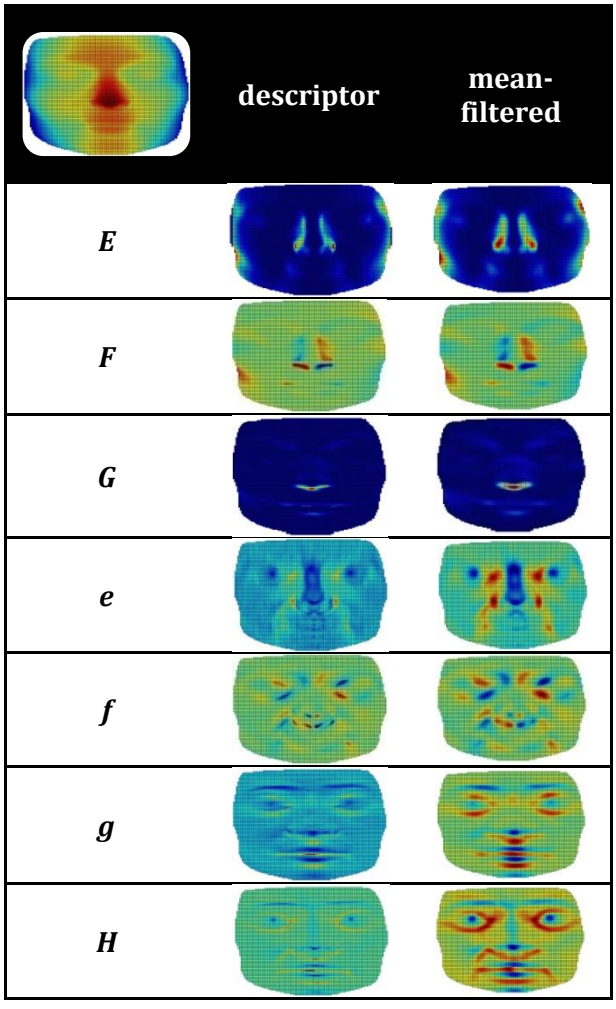
Koenderink and van Doorn's shape index and curvedness are defined in their original formulation [9]:

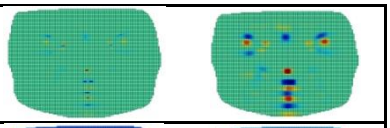
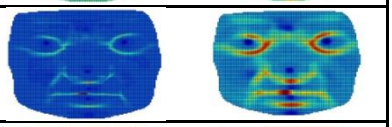
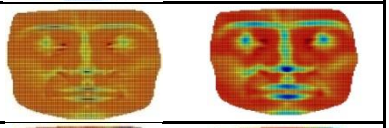
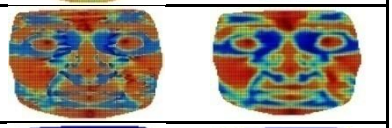
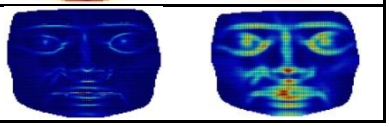
$$S = -\frac{2}{\pi} arctan \frac{k_1 + k_2}{k_1 - k_2}, \quad S \in [-1,1], \quad k_1 \geq k_2, \quad (1.11)$$
$$C = \sqrt{\frac{k_1^2 + k_2^2}{2}}. \quad (1.12)$$

These descriptors, whose facial point-by-point maps are shown in the second column of Table 1, have proven to be effective in describing facial shape at different ages and in extracting landmarks [24] [25] [26]. Diagnosis [28] and recognition [29] were also successfully addressed.

For manageability reasons [30], a mean filter has been applied to some geometrical descriptors. Means are calculated in squared neighbourhoods of side 4 around each point of the facial depth maps. The twelve geometrical descriptors and their respective mean-filtered versions are shown mapped on a face on Table 1. The images in this table regard only one person (female, aged 25, serious pose), whose depth map was obtained via Minolta Vivid 910 laser scanner.

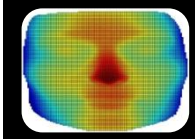
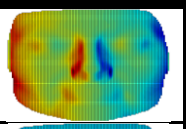
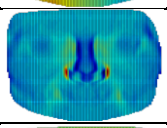
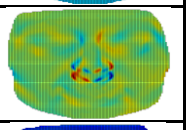
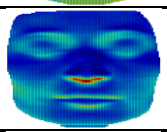
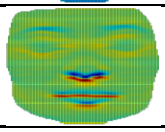
Table 1. Point-by-point maps on a face of descriptors. The second columns shows the twelve geometrical descriptors; the third column shows the mean-filtered versions.



K	
k_1	
k_2	
S	
C	

Together with the above mentioned geometrical descriptors, the derivatives have also been taken into consideration to describe the face point-by-point. Their mappings on the same face shown in Table 1 are reported in Table 2.

Table 2. Derivatives of the facial surface shown on a face map. They are computed point-by-point via Matlab® function "gradient".

	description	map
D_x	first derivative with respect of *x*	
D_{xx}	second derivative with respect of *x*	
D_{xy}	mixed derivative	
D_y	first derivative with respect of *y*	
D_{yy}	second derivative with respect of *y*	

2.2 Landmark localization

The set of 13 landmarks localized by this method is shown in Figure 2.

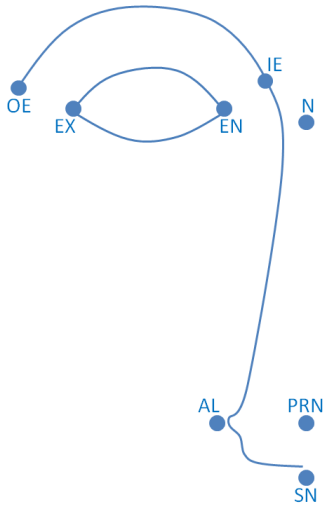


Figure 2. Landmarks used in this study: OE-outer eyebrow, IE-inner eyebrow, EX-exocanthion, EN-endocanthion, N-nasion, AL-alare, PRN-pronasale, SN-subnasale. Except for the eyebrow points, which are not considered real soft-tissue landmarks, their morphometric definitions are provided by Swennen et al. [3].

The method firstly defines an area of research (also called area/region of interest) where to start the localization of the point in exam. Regions of interest are rectangular and dimensioned according to the 3D model dimension, so that region and face size era proportional. These areas are narrowed thanks to the application of different geometrical conditions via thresholding technique on numerical values of descriptors. The thresholds have been experimentally set using a training facial dataset of 40 serious faces + 50 expressive/occluded faces of the Bosphorus dataset. The numerical values of the thresholds have been chosen with an automatic function which checks the numerical value of each geometrical descriptor on the locus of the ground truth landmark. Figures 3 and 4 show some generated graphs of descriptors *f* and *H*, respectively, which supports the choice of the thresholds for the extraction of the ***nasion***. Other similar representations have been obtained for the other descriptors and landmarks.

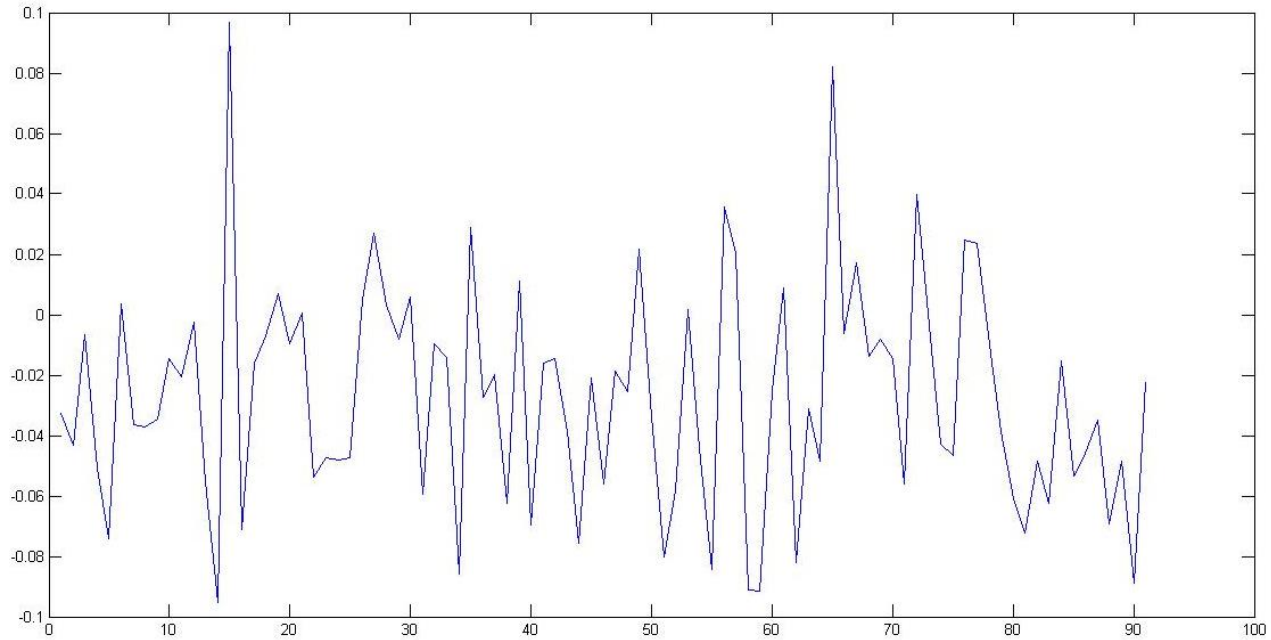


Figure 3. Numerical point values (*y*-axis) of descriptor *f* mapped on each of the 90 faces (*x*-axis) of the training dataset on the locus of the ground truth *nasion* landmark point. The behaviour of this graph supports the setting of the thresholding of descriptor *f* for the localization of this landmark, and in particular to the formulation of condition 3: "the second coefficient of the second fundamental form *f* is nearly equal to zero, i.e. f ∈ (-0.1; 0.1)."

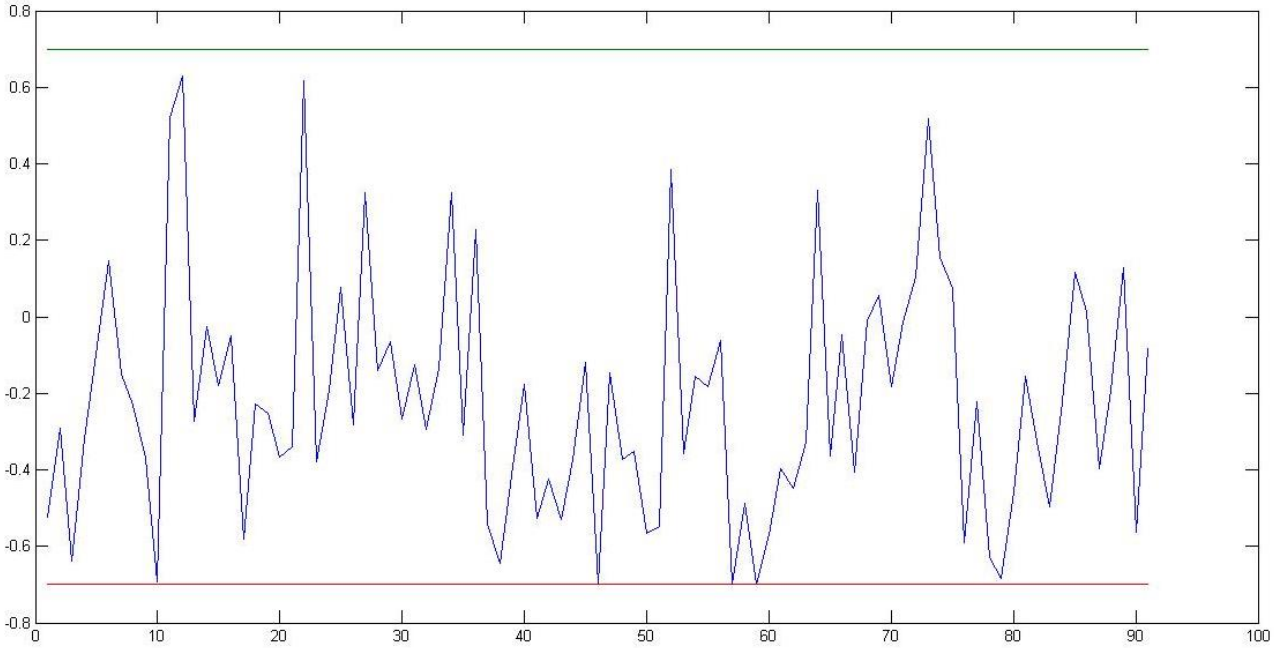


Figure 4. Numerical point values (*y*-axis) of descriptor *H* mapped on each of the 90 faces (*x*-axis) of the training dataset on the locus of the ground truth *nasion* landmark point. The behaviour of this graph supports the setting of the thresholding of descriptor *H* for the localization of this landmark, and in particular to the formulation of condition 2: "the mean curvature is approximately equal to zero, namely H ∈ (-0.7; 0.7)." In this representation, the minimum of the descriptor (-0.7) is given by a red line; the maximum (+0.7) by a green line.

The local behaviour of each descriptor on the locus of a landmark is reported in Table 3, which was also obtained thanks to previous studies of geometrical descriptors on faces [24]. The table reports updated values and behaviours with respect to that in [24] and holds for all faces belonging to the training dataset.

Table 3. Behaviour of each geometrical descriptor in correspondence to each landmark. Acronyms on the second row: c is cup, ru is rut, sru is saddle rut, s is saddle, sri is saddle ridge, r is ridge, d is dome. If there are two notes in the same box, the first one is referred to the landmark on the left side of the face, the other to the right, if we consider the point of view of someone who is looking to the face, in front of it. The x-axis is vertical, y is horizontal and z enters the sheet.

	OE	IE	EX	EN	AL	PRN	N	SN
Class	d	d	ri	c/ru	ri	d	sru/s/sri/ri	ru/sru/s/sri
E	local max	≈0	local max	≈0	local max	≈0	<3	≈0
F	>0, <0	>0, <0	≈0	≈0	>0, <0	≈0	≈0	Є (-2;2)
G	>3	>10	≈0	≈0	≈0	≈0	local max	≈0
e	≈0	local min	local max	local max	<0.4	local min	local min	local max
f	local min local max	local min local max	≈0	Є (-0.4;0.4)	≈0	≈0	Є (-0.1;0.1)	≈0
g	local min	>-3	local max	local max	local max	<-2.9	local max	local max
H	local min	local min	≈0	>-0.6	≈0	<2	Є (-0.7;0.7)	<1.1
K	≈0	local max	local max	local max	local max	≈0	≈0	≈0
k_1	≈0	local max	≈0	local max	local max	local min	local max	local min
k_2	local min	local min	≈0	local max	local min	≈0	≈0	local min

S	Є (0.625;1)	Є (0.625;1)	Є (-0.625;-0.125)	Є (-1;-0.375)	Є (-0.625;-0.125)	Є (0.56;1)	Є (-0.375;0.625)	Є (-0.47;0.47)
C	≈0	local max	local max	local max	local max	local min	≈0	local min
$\mathbf{D_x}$	>0, <0	>0, <0	≈0	Є (-3;3)	>0, <0	≈0	≈0	Є (-1.5;1.5)
$\mathbf{D_{xx}}$	local min	≈0	>0	>0	local max	≈0	local min	<0
$\mathbf{D_{xy}}$	local min	local min	Є (-3;3)	local max	≈0	local max	>1	≈0
$\mathbf{D_y}$	>2	>2	>0	Є (-3.5;3.5)	<0	local max	<0	local max
$\mathbf{D_{yy}}$	local min	local min	Є (-3;3)	local max	≈0	local max	>1	≈0

The algorithm starts with the localization of the ***pronasale***, whose key geometrical features are summed up here:

1. the mean-filtered shape index S lies in the range (0.56; 1), i.e. the point belongs to the point whose geometrical shape is "cap";
2. the mean curvature H is lower than a threshold value, i.e. $H < 2$, which was experimentally set;
3. the coefficient g is lower than a threshold value, i.e. $g < -2.9$;
4. the coefficient e has a local minimum in this area;
5. k_1 has a local minimum in correspondence to the point.

A primary area of research, called Pron1, is identified by focusing on a central area of research, in which the point with maximum Z value is identified and called P_{max}. A neighbourhood of this point is created, which becomes the new area of research called Pron2. Conditions 1 to 4 were used to gradually narrow the area of interest, while the final condition carries out point extraction. Also, when the minimum of k_1 is obtained, the algorithm searches for another minimum, lower than a threshold value ($k_1 < 0.9$) in the neighbourhood of the one already found. If a point satisfying this condition is obtained, the final pronasale is located in the middle of the two points, otherwise the pronasale is the first minimum is kept. Figure 5 shows the diagram of the process.

The images showing the narrowing of the area of interest are generated on the whole faces instead of on the initial region of interest. This allows an improved understanding on how the descriptor condition behaves on the whole facial shape. In the actual algorithm, conditions are applied to the region of interest only. This holds for all other landmarks and their related images.

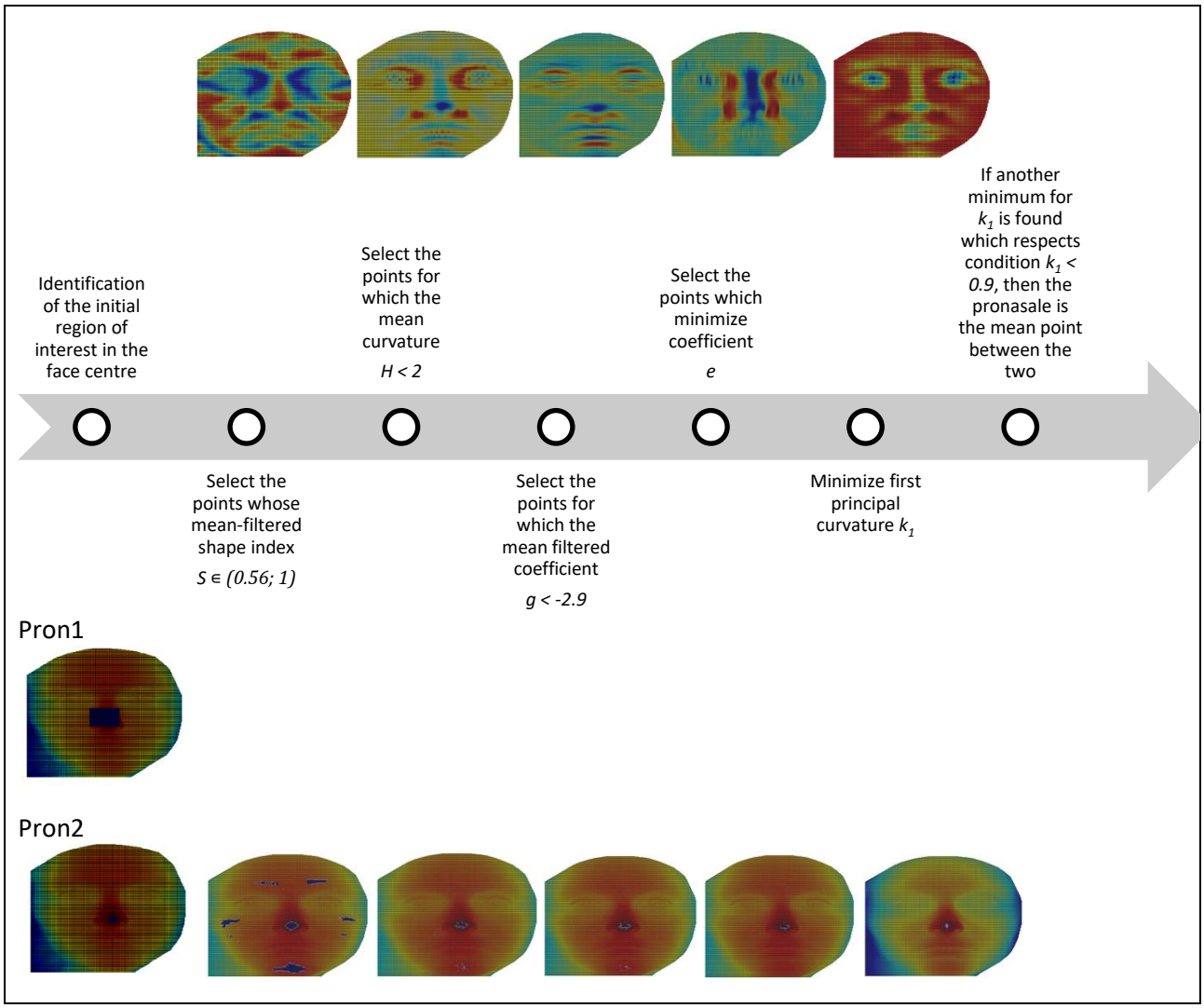


Figure 5. Process for localizing the pronasale. The first row shows the descriptors (mapped on a face of the Bosphorus database) involved in the instruction below: (from left to right) shape index, mean curvature, coefficient *g*, coefficient *e*, first principal curvature. The final row shows the gradual narrowing of the area of interest when the steps above are gradually applied. The landmark is extracted at the final step.

The branch of the global algorithm which verifies the correctness of the PRN localization re-defines the initial area of research of the point, in case of anomalies. This re-definition of the area of research is set via alae positions, by creating a neighbourhood of the middle point between the AL points.

Then, the ***alae*** are extracted. The right AL is detected first, by identifying an area of interest at approximately a similar *y* value of PRN. The area of research of the left AL is defined symmetrically. The PRN position is also used to define left and right facial sides. Alae's salient geometrical features are:

1. the coefficient *e* is lower than a threshold value, i.e. $e < 0.4$;
2. the coefficient *E* has a local maximum in correspondence to the alae points.

Condition 1 refines the area of interest and condition 2 localizes the points. The steps are shown in Figure 6.

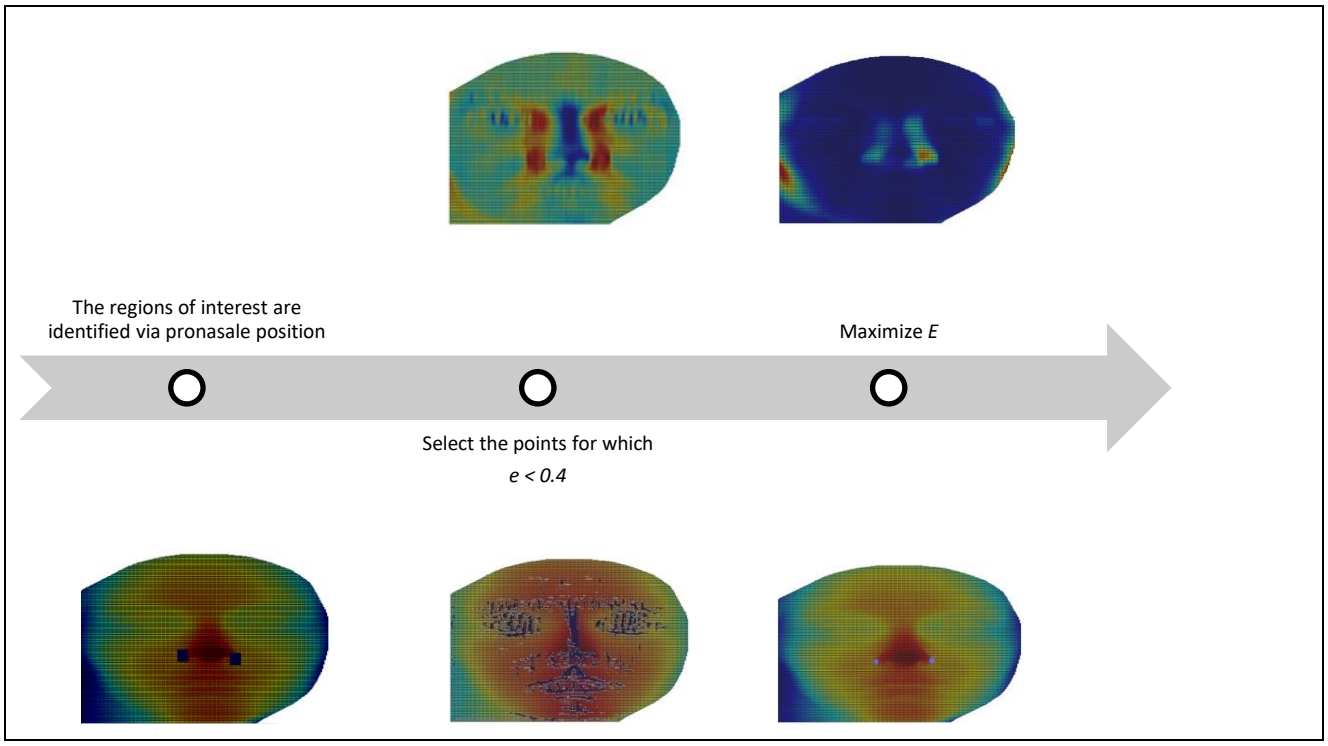


Figure 6. Process for localizing the alae. The first row shows the descriptors (mapped on a face of the Bosphorus database) involved in the instruction below: (from left to right) derivative with respect of *x*, coefficient *e*, coefficient *E*. The final row shows the gradual narrowing of the area of interest when the steps above are gradually applied. The landmarks are extracted at the final step.

The pronasale coordinates are also used to identify the area of research of the ***nasion***. The nasion is searched in the same vertical direction of the pronasale and distant from it approximately one third the length of the face. Its geometrical features are:

1. the point belongs to the points whose shape index value lies in the range corresponding to saddle rut, saddle, saddle ridge, and ridge, i.e. $S \in (-0.375; 0.625)$;
2. the mean curvature is approximately equal to zero, namely $H \in (-0.7; 0.7)$, where 0.7 is a threshold experimentally set;
3. the second coefficient of the second fundamental form *f* is nearly equal to zero, i.e. $f \in (-0.1; 0.1)$, where 0.1 is a threshold experimentally set;
4. the first coefficient of the first fundamental form *E* is lower than a threshold value, i.e. $E < 3$;
5. the coefficient g has a maximum in correspondence to this point.

Conditions 1 to 4 support the area of interest localization, while the final condition extracts the point. Figure 7 exposes the process.

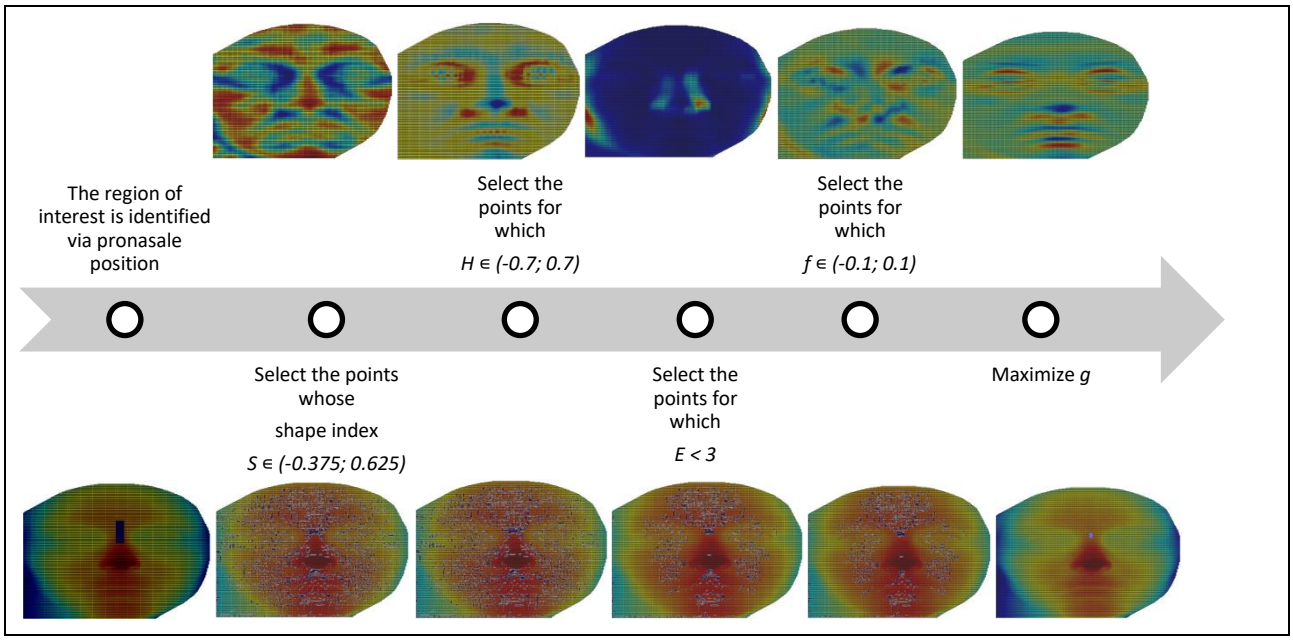


Figure 7. Process for localizing the nasion. The first row shows the descriptors (mapped on a face of the Bosphorus database) involved in the instruction below: (from left to right) shape index, mean curvature, coefficient *E*, coefficient *f*, coefficient *g*. The final row shows the gradual narrowing of the area of interest when the steps above are gradually applied. The landmark is extracted at the final step.

The ***subnasale*** position is defined horizontally between the alae and in the vertical direction of the pronasale. Its geometrical features are the following.

1. The shape index lies in the range (-0.47; 0.47), i.e. this point could be associated to a saddle shape;
2. the first derivative with respect of x is approximately equal to zero, i.e. $D_x \in (-1.5; 1.5)$;
3. the second derivative with respect of x is negative;
4. the mean curvature is lower than an experimentally set threshold, namely $H < 1.1$;
5. the first fundamental form coefficient F lies in the range (-2; 2);
6. the first derivative with respect of y has a local maximum in correspondence to the subnasale point.

Conditions 1-5 are used to refine the area of interest and condition 6 is adopted to extract the point. The steps are shown in Figure 8.

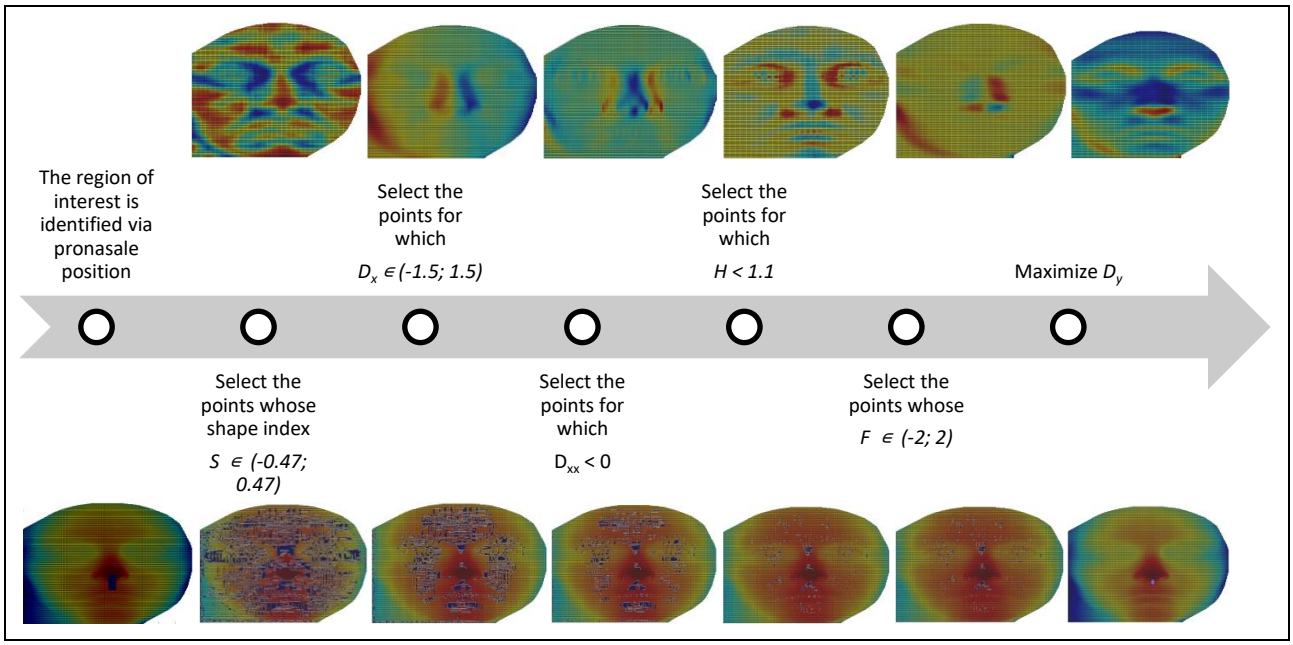


Figure 8. Process for localizing the subnasale. The first row shows the descriptors (mapped on a face of the Bosphorus database) involved in the instruction below: (from left to right) shape index, first derivative with respect of *x*, second derivative with respect of *x*, mean curvature, coefficient *F*, first derivative with respect of *y*. The final row shows the gradual narrowing of the area of interest when the steps above are gradually applied. The landmark is extracted at the final step.

Endocanthions' areas of interests are localized relying on the positions of alae and nasion previously localized. These conditions are used to refine the regions of interest and then to extract the points:

1. endocanthions belong to the set of points whose geometrical shape is cup or rut, i.e. $S \in (-1; -0.375)$;
2. the mixed derivative is lower than a threshold value, $D_{xy} < 0.6$;
3. the first derivative with respect of *x* ranges from -3 to 3, i.e. $D_x \in (-3; 3)$;
4. the first derivative with respect of *y* ranges from -3.5 to 3.5, i.e. $D_y \in (-3.5; 3.5)$;
5. the coefficient f is nearly equal to zero, namely $f \in (-0.4; 0.4)$, and is positive on the right endocanthion and negative on the left;
6. the mean curvature H is higher than threshold value -0.6;
7. the mean-filtered coefficient e has a maximum in correspondence to these points.

Conditions 1 to 6 support the refinement of the areas of research, while the final condition detects the two points. Figure 9 shows steps and images.

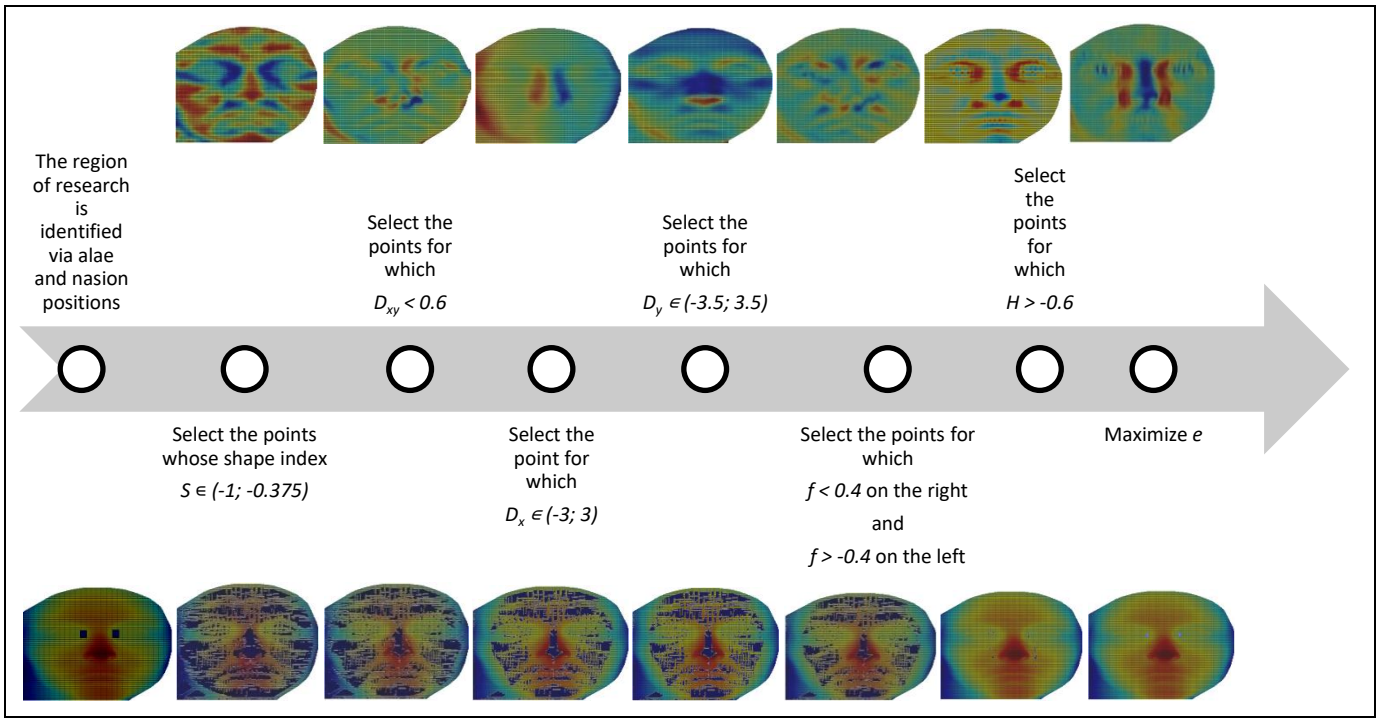


Figure 9. Process for localizing the endocanthions. The first row shows the descriptors (mapped on a face of the Bosphorus database) involved in the instruction below: (from left to right) shape index, mixed derivative, first derivative with respect of *x*, first derivative with respect of *y*, coefficient *f*, mean curvature, coefficient *e*. The final row shows the gradual narrowing of the area of interest when the steps above are gradually applied. The landmarks are extracted at the final step.

Exocanthions' areas of research are identified based on endocanthions' positions. These are the conditions adopted to extract their locations:

1. the first derivative with respect of *y* is positive;
2. the second derivative with respect of *x* is positive;
3. the second derivative with respect of *y* lies in the range (-3; 3);
4. the first coefficient of the first fundamental form E has two maximums in correspondence to the two landmarks.

Conditions 4 is adopted the extract the landmarks in the regions of interest identified at steps 1 to 3. Figure 10 shows the process.

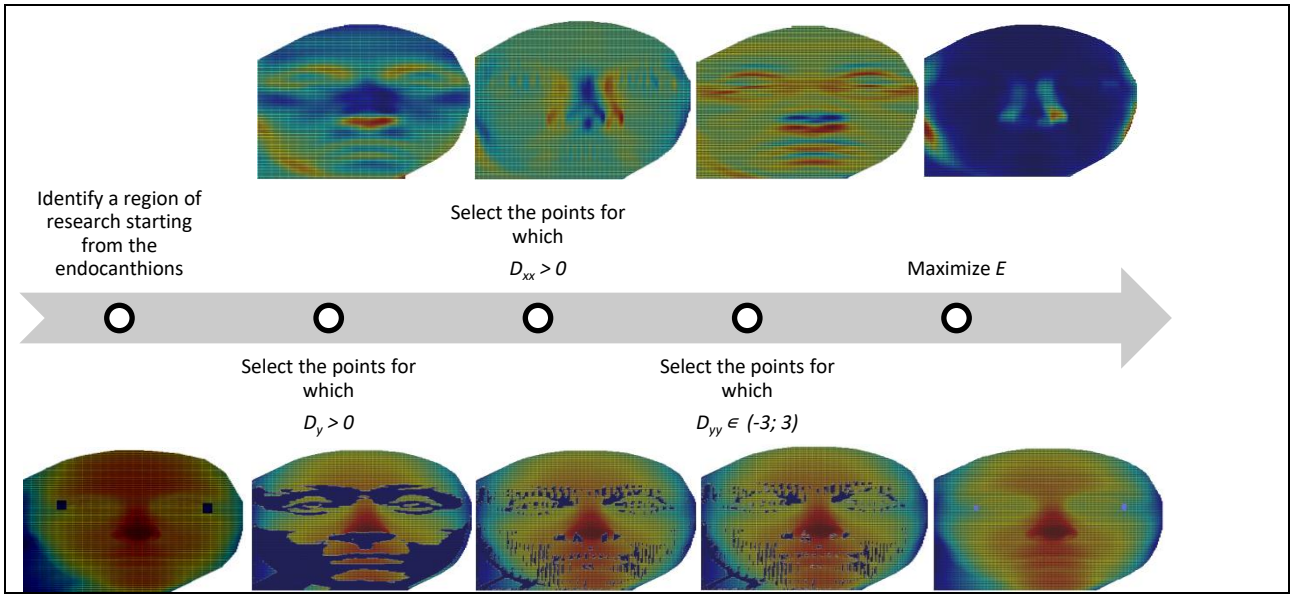


Figure 10. Process for localizing the exocanthions. The first row shows the descriptors (mapped on a face of the Bosphorus database) involved in the instruction below: (from left to right) first derivative with respect of *y*, second derivative with respect of *x*, second derivative with respect of *y*, coefficient *E*. The final row shows the gradual narrowing of the area of interest when the steps above are gradually applied. The landmarks are extracted at the final step.

The ***inner eyebrows***, whose areas of research are set via endocanthions' positions, have these geometrical conditions:

1. the coefficient *g* is greater than the threshold value -3;
2. the first derivative with respect of *y* is greater than 2;
3. coefficient *G* is greater than a threshold value 10;
4. coefficient *f* has a maximum in correspondence to the right inner eyebrow and a minimum to the left.

Left and right sides of the face are identified via PRN coordinates. Conditions 1-3 are used to narrow the areas of interest; condition 4 extracts the points. Figure 11 shows the process step by step.

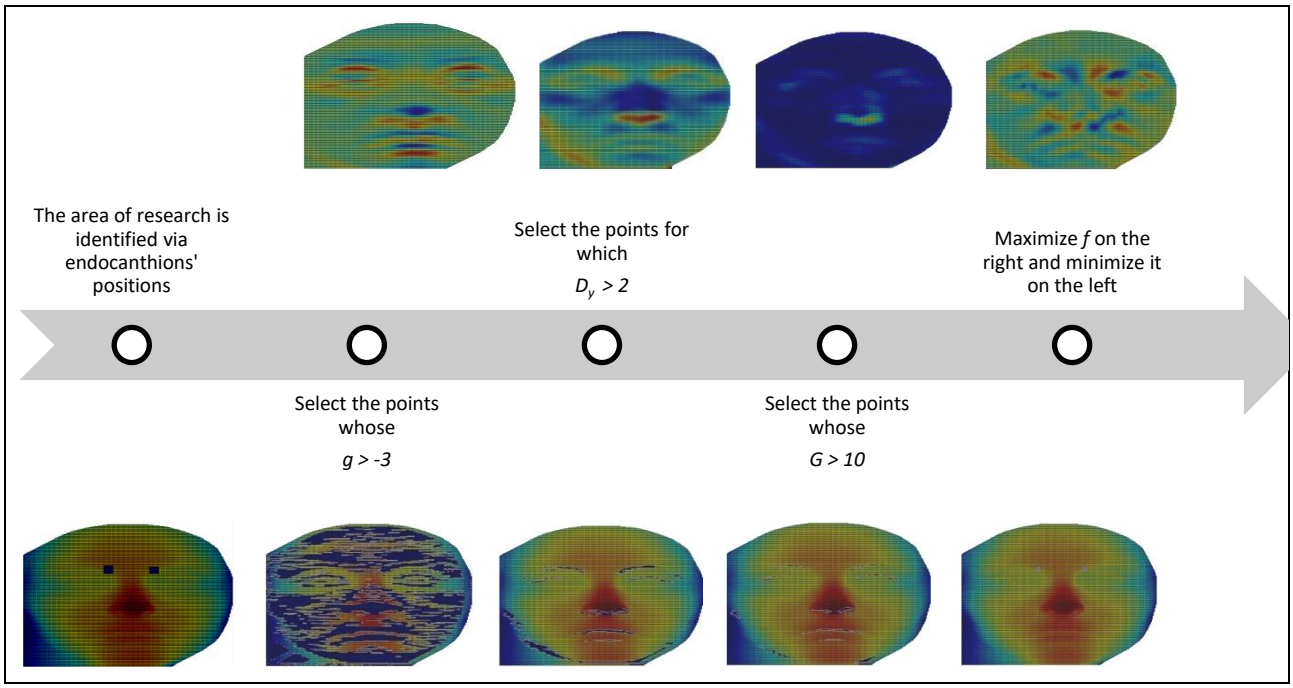


Figure 11. Process for localizing the inner eyebrows. The first row shows the descriptors (mapped on a face of the Bosphorus database) involved in the instruction below: (from left to right) coefficient *g*, first derivative with respect of *y*, first derivative with respect of *x*, coefficient *G*, coefficient *f*. The final row shows the gradual narrowing of the area of interest when the steps above are gradually applied. The landmarks are extracted at the final step.

The ***outer eyebrows***, whose areas of research are set via exocanthions' and inner eyebrows' positions, have these geometrical conditions:

1. the coefficient G is greater than a threshold value ($G > 3$);
2. coefficient *f* has a maximum in correspondence to the right outer eyebrow and a minimum to the left.

Left and right sides of the face are identified via PRN coordinates. Conditions 1 is used to narrow the region of interest; condition 2 extracts the point. The steps are shown in Figure 12.

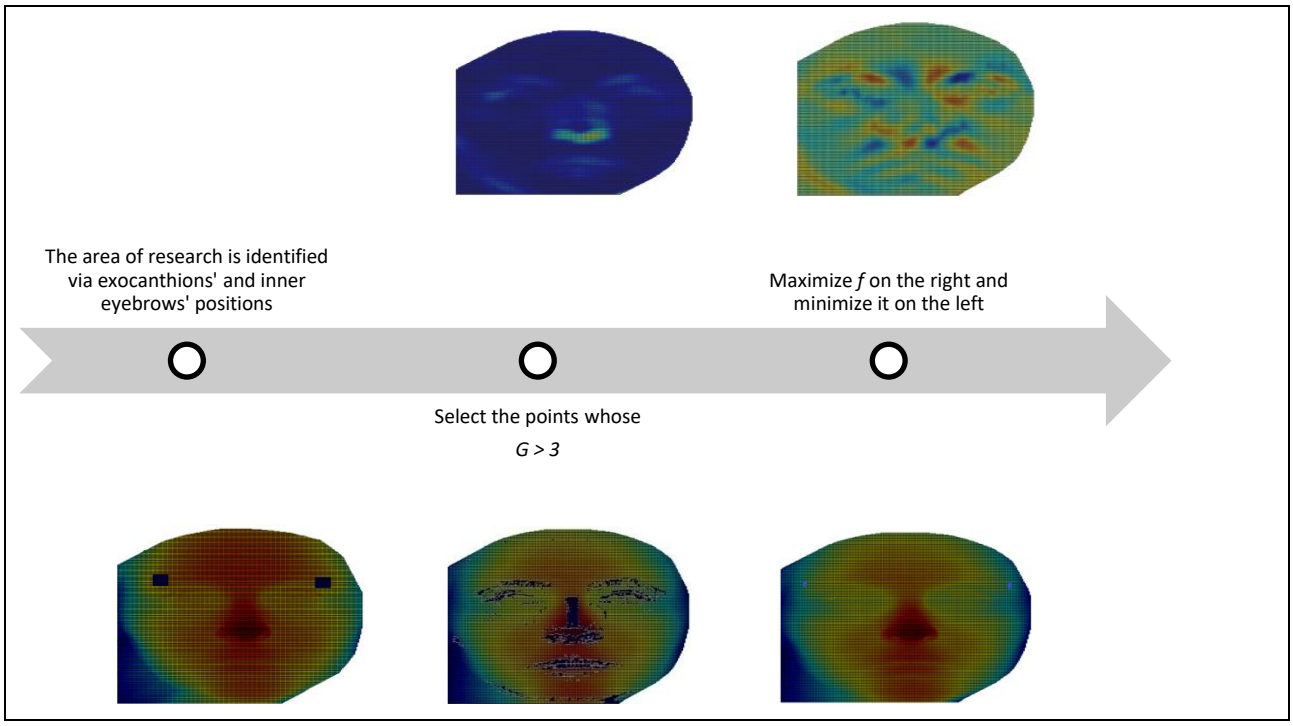


Figure 12. Process for localizing the outer eyebrows. The first row shows the descriptors (mapped on a face of the Bosphorus database) involved in the instruction below: (from left to right) first derivative with respect of *x*, coefficient *G*, coefficient *f*. The final row shows the gradual narrowing of the area of interest when the steps above are gradually applied. The landmarks are extracted at the final step.

2.3 Occlusion detection

The method is designed to detect two types of occlusions: eye- and eyeglasses-based. Eye occlusions are provoked by hands covering these facial parts. These kinds of occlusions are chosen relying on the available occluded faces of the Bosphorus database.

The **eye-based occlusion** detection process is described in Figures 13 and 14. The algorithm:

1. divides the face into left and right parts;
2. computes five features for each facial part (left and right):
    a. the 7-bins histogram of the point-by-point facial map of the shape index;
    b. the number of points which respectively satisfy these conditions:
        i. $D_x > 15$, $D_y > 10$; $D_{xy} > 2$; $D_x < -15$; $D_y < -10$, and $D_{xy} < -2$;
        ii. $D_{yy} > 15$ and $D_{yy} < 15$;
        iii. $H > 0.05$ and $G > 3$;
        iv. $g < 0$ and $D_y > 2$;
3. evaluates the difference between left and right features. The obtained features are: *ShapeIndex*, $Derivative_1$, $Derivative_2$, $Coefficient_1$, and $Coefficient_2$;
4. compares these features with thresholds:
    a. *ShapeIndex* > 1000 and *ShapeIndex* < 500;
    b. $Derivative_1 \equiv 0$;
    c. $Derivative_2 > 600$ and $Derivative_2 \leq 0$;
    d. $Coefficient_1 > 1000$, $Coefficient_1 > 500$, and $Coefficient_1 \leq 0$;
    e. $Coefficient_2 > 1000$, $Coefficient_2 > 500$, $Coefficient_2 \leq 500$, $Coefficient_2 \leq 0$;
    f. $Derivative_1 > 20$ and $Derivative_2 < 200$;
    g. $Derivative_1 < 20$ and $Derivative_2 < 300$;
    h. $Derivative_1 < 50$ and $Derivative_2 < 350$;
    i. $Tot = ShapeIndex + Derivative_1 + Derivative_2 + Coefficient_1 + Coefficient_2 \geq 15000$.

5. after this first part, divides the face into three areas, shown in Figure 15: left eye area, right eye area, and mouth area;
6. evaluates, for each area, the number of points with the z-coordinate greater than the pronasale one. These features are called, respectively, $Area_1$, $Area_2$, and $Area_3$, depending on the name of the area which they are computed on;
7. on the whole facial area, evaluates the number of points with the z-coordinate greater than the pronasale one. This feature is called *Peack*;
8. on the whole facial area, evaluates the number of points with the z-coordinate largely lower than their neighbouring points. This feature is called *Ditch*;
9. calculates $\%Occluded = \frac{(Area_1+Area_2+Area_3)\cdot 100}{N \cdot M}$ where N and M are dimensions of the facial depth map;
10. compares these features with the following thresholds:
    a. $\%Occluded > 2.4$;
    b. $Area_1 > Area_2$ and $Area_1 > Area_3$;
    c. $Area_2 > Area_1$ and $Area_2 > Area_3$;
    d. $Peack > 100$;
    e. $Ditch > 100$.

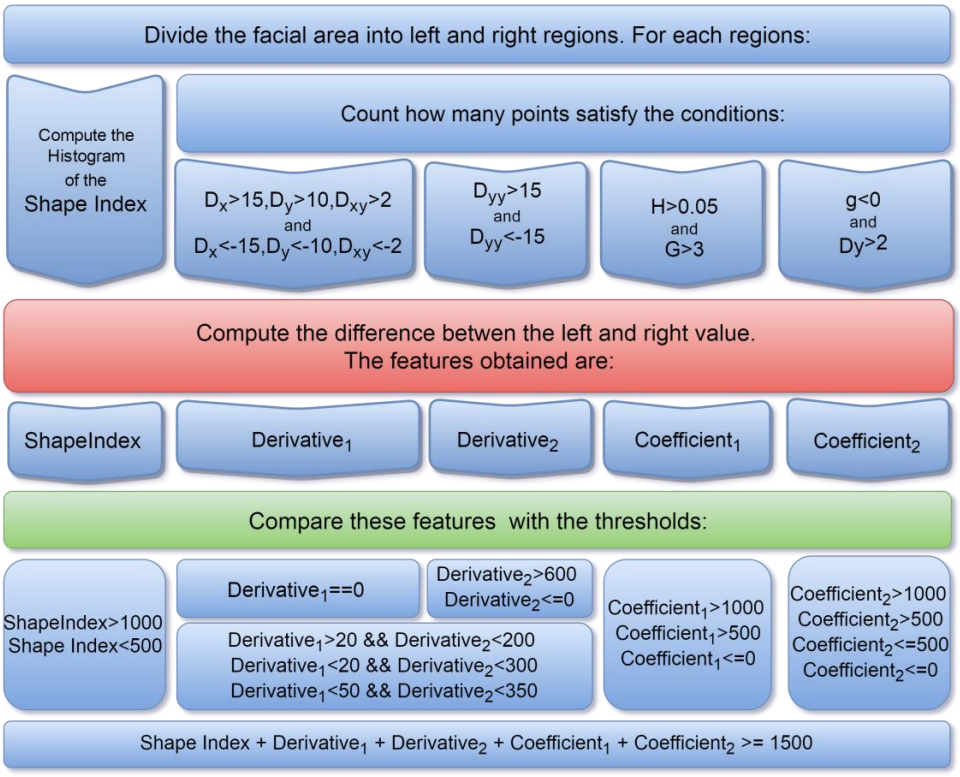


Figure 13. First part of the eye-based occlusion detection process.

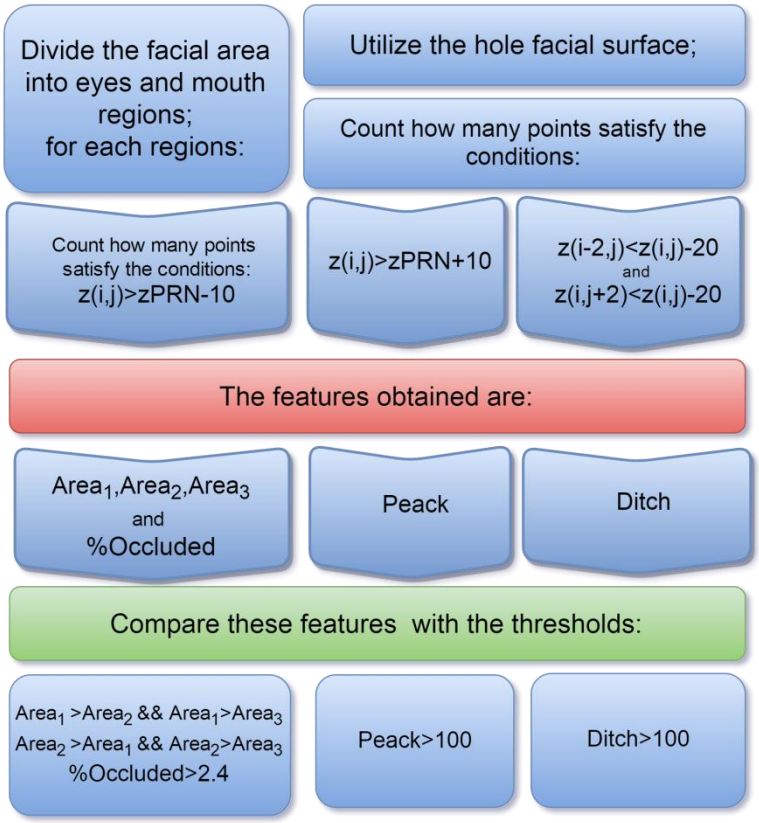


Figure 14. Second part of the eye-based occlusion detection process.

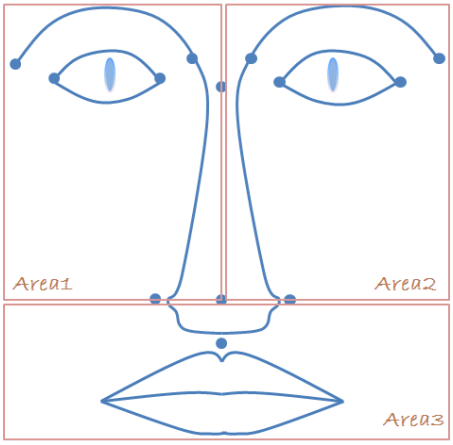


Figure 15. Areas defined at the first step of second part of overall occlusion detection process.

These features present specific behaviour in presence of eye occlusions provoked by a hand, as shown in Figure 16.

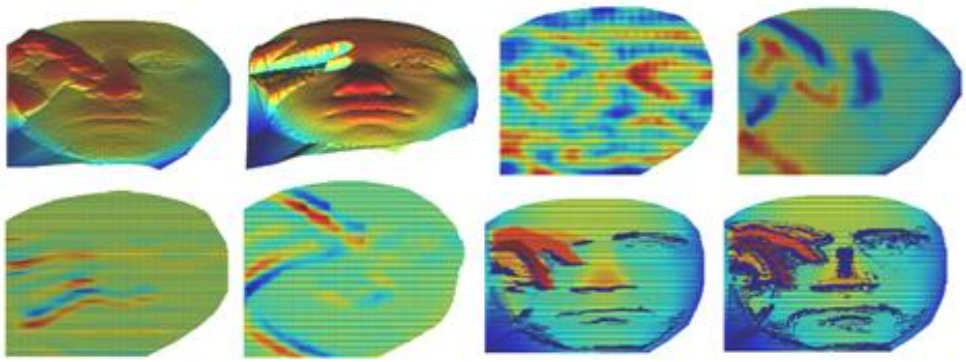


Figure 16. Facial maps of the Bosphorus database highlighting specific behaviours in the case of eye occlusion: depth maps showing respectively Peack zone and Ditch zone, shape index, first derivative with respect of x, second derivative with respect of y, mixed derivative, and two depth maps showing the applied geometrical conditions ($H > 0\text{-}05$ and $g > 3$ on the penultimate face; $g < 0$ and $D_y > 2$ on the last face).

The presence of **eyeglasses** is detected with an algorithm schematized in Figure 17. The algorithm:

1. taking into consideration only the eyes areas shown in Figure 18a, evaluates the number of points satisfying all conditions: $D_{xx} > 2$, $D_{xx} < -2$, $D_{xy} > 0.5$, $D_{xy} < -0.5$, $D_{yy} > 1$, $D_{yy} < -1$, $G > 20$. This feature is called *Coefficient*;
2. considering eyeglasses area shown in Figure 18b, computes:
    a. the number of points with the z-coordinate greater than the pronasale one. This feature is called *Count*;
    b. on the whole facial area, evaluates the number of points with the z-coordinate largely lower than their neighbouring points. This feature is called *Ditch*;
3. compares these features with thresholds:
    a. $Coefficient > 400$:
    b. $Count \in (0; 300)$ and $Ditch > 0$.

If these conditions are satisfied, eyeglasses are present. Figure 19 shows the behaviour of the involved descriptors when eyeglasses are present.

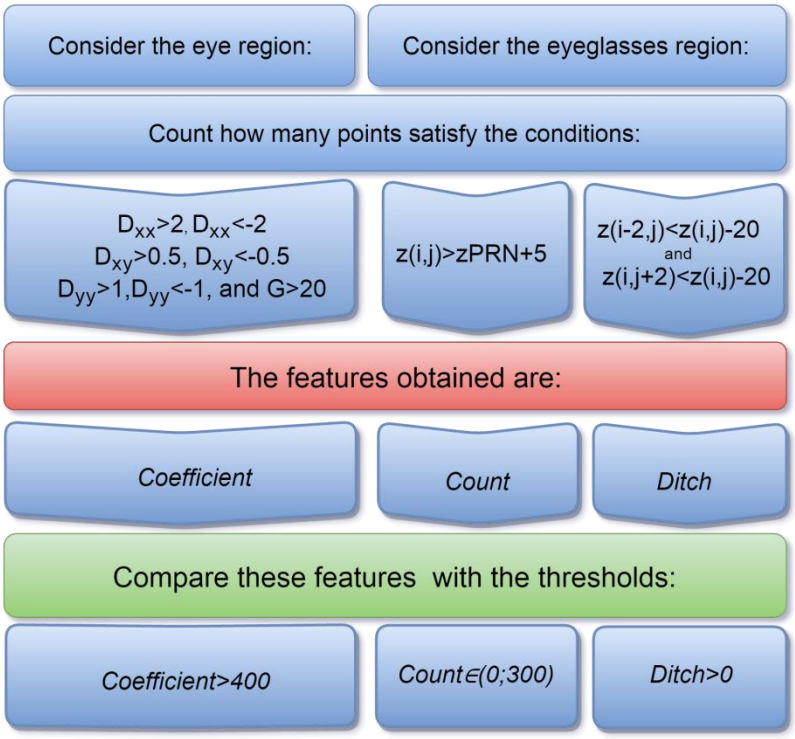


Figure 17. Eyeglasses-based occlusion detection process.

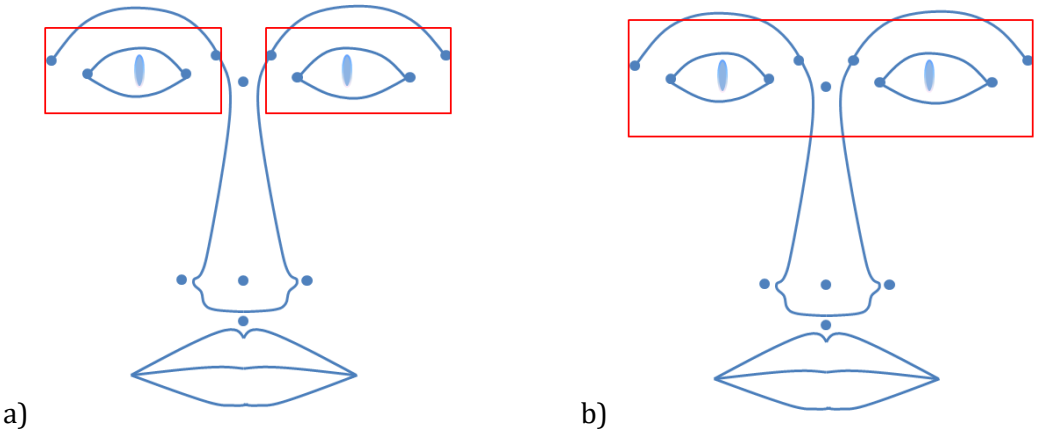


Figure 17. a) Eyes areas used to calculate feature *Coefficient*. b) Eyeglasses area used to calculate features *Count* and *Ditch*.

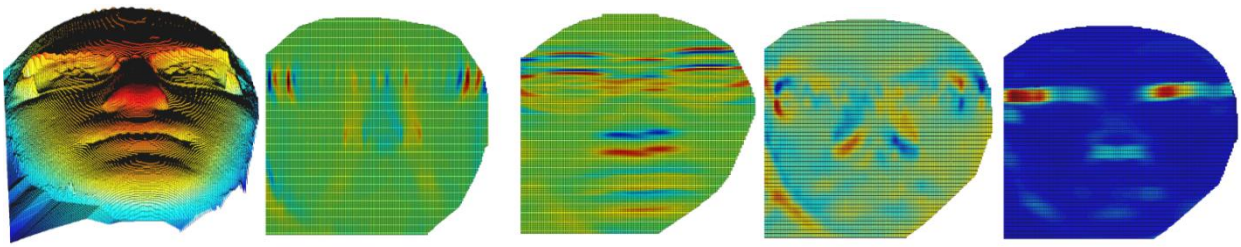


Figure 19. Facial maps of the Bosphorus database highlighting specific behaviours in the case of eyeglasses: facial depth map, second derivative with respect of *x*, second derivative with respect of *y*, mixed derivative, coefficient *G*.

When occlusions, if present, are detected and classified, the algorithm deletes the landmarks on the occluded parts, keeping only the ones on actual facial parts.

3 Results

The method was tested on 3362 faces overall. Among them, 3132 belong to the Bosphorus database, 230 belong to our private database. Even if we have developed a methodology for rotating the face into a standard pose [31], the faces tested in the present study are all front-view.

The faces belonging to the Bosphorus database include 286 serious, 68 angry, 63 disgusted, 64 fearful, 99 happy, 59 sad, 65 surprised, 92 eyes-occluded, 93 mouth-occluded, 94 glasses-occluded faces. The remaining ones describe Action Units positions. In particular, serious, expressive, occluded faces have been included in the testing dataset.

Our private database, acquired via Minolta Vivid 910 laser scanner, collects 32 subjects, each acquired with the serious pose and other expressions among the 6 basic emotions. Figure 20 shows a selection of the faces.

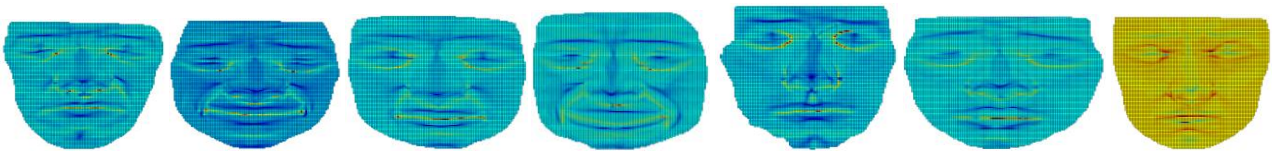


Figure 20. Faces of our database belonging to different subjects performing different expressions. From left to right: disgust, joy, fear, anger, serious, surprise, sadness.

Figures 21 and 22 show some landmark localization results on different faces of the Bosphorus and internal databases, respectively.

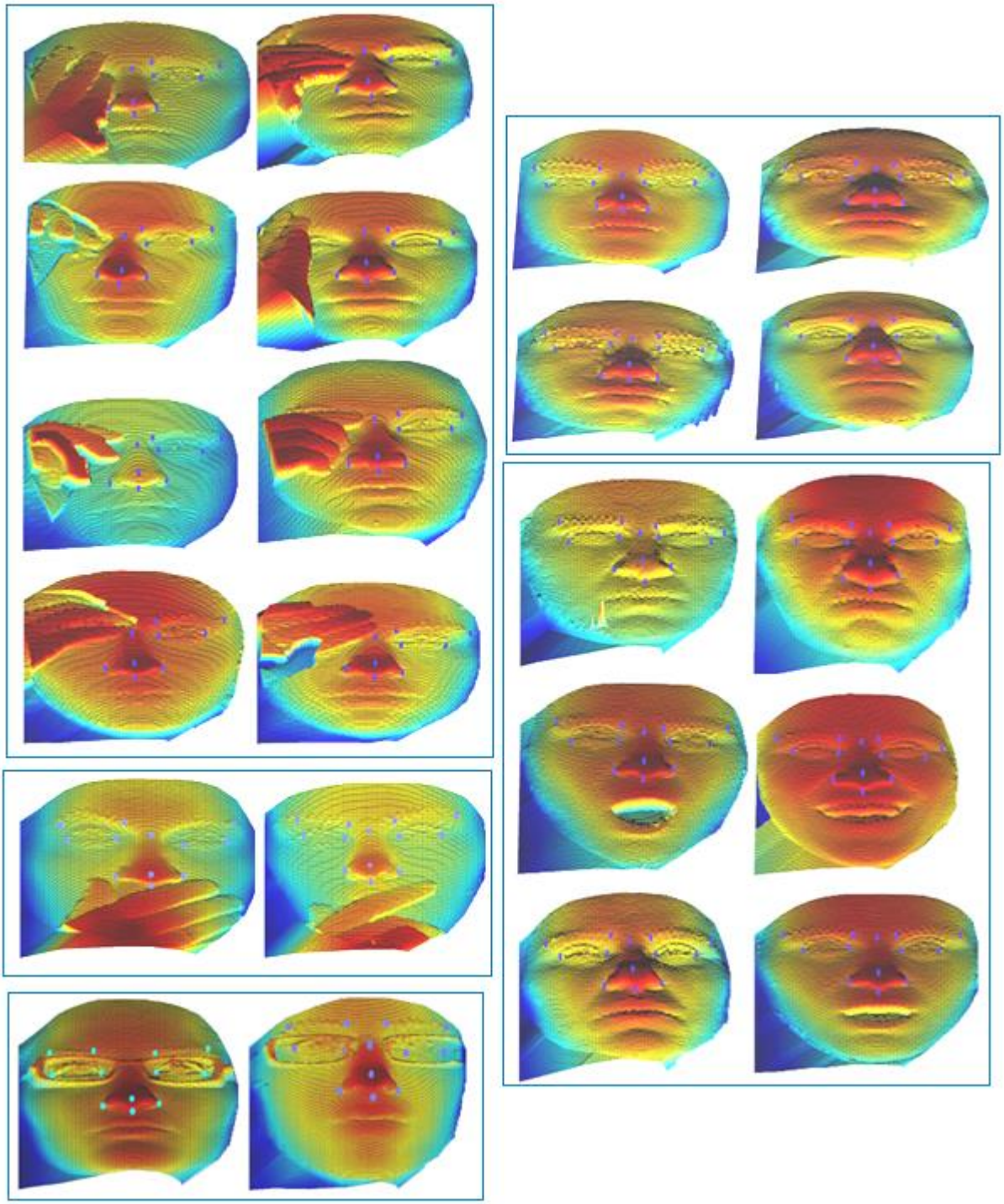


Figure 21. Landmarks localized by our algorithm on different Bosphorus faces. The first two columns refer to occluded faces (set on the top: eye-occluded faces; set in the centre: mouth-occluded faces; set on the bottom: eyeglasses-occluded faces). The third and fourth columns show four serious faces (set on the top) and six expressive faces, one per expression (set on the bottom).

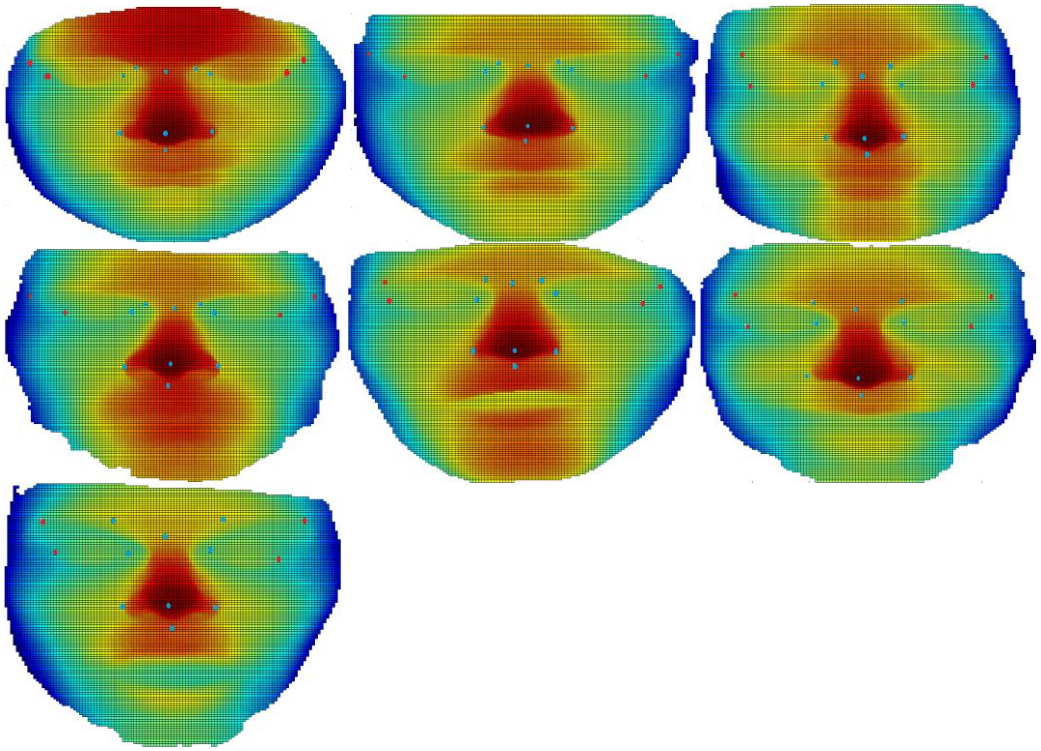


Figure 22. Landmarks localized by our algorithm on different subjects with various expressions of our internal database.

Table 4 shows the detailed landmark localization errors (in millimetres) computed as Euclidean distances between the ground truth landmarks, whose coordinates are reported in the Bosphorus database, and the obtained ones. Concerning our database, the ground truth landmarks have been manually allocated on the faces and compared to those obtained with the algorithm.

Table 4. Detailed landmark localization errors.

Database	face type	dataset size	OE	IE	EX	EN	AL	PRN	SN	N	mean	
Bosphorus database	Neutral	286	7,60	4,71	4,90	3,36	4,70	2,54	2,86	2,97	4,21	
	Expressive	418	8,02	5,55	5,48	5,56	4,64	2,69	3,75	4,04	4,97	4,93
	Action Units	2149	8,06	5,75	5,77	4,15	5,14	2,64	3,51	4,20	4,90	
	Eye occluded	92	7,20	5,49	4,64	3,73	5,24	2,45	2,84	3,74	4,42	4,68
	Mouth occluded	93	7,75	4,65	5,47	3,72	5,14	2,88	3,16	3,32	4,51	
	Glasses occluded	94	9,94	5,55	6,20	3,94	4,93	2,40	2,83	5,22	5,13	
	Mean *Bosphorus*	3132	8,10	5,46	5,56	3,91	5,05	2,61	3,31	4,07	4,76	
internal database	Mean *internal database* (neutral + expressive)	230	5,29	5,42	6,16	5,02	3,57	2,78	6,54	3,05	4,73	
Global testing database	Global mean	3362	7,16	5,45	5,76	4,28	4,55	2,67	4,39	3,73	4,75	

Analyzing the two databases separately, we can see that the error obtained on the Bosphorus (4.76 mm) is comparable to that of our internal database (4.73 mm). Occluded faces, belonging only to the Bosphorus database, show a 4.68 mm mean error, which is given by a mean between eye-occluded (4.42 mm), mouth occluded (4.51 mm), and glasses occluded (5.13 mm) faces. Expressive faces of the Bosphorus database, including those displaying the six basic emotions (4.97 mm) and those

representing the Action Units one by one (4.90 mm), gave 4.93 mm mean error. Overall, the errors are very similar among databases and face types, thus proving the robustness of the method under different conditions.

Figure 23 shows global landmark localization errors.

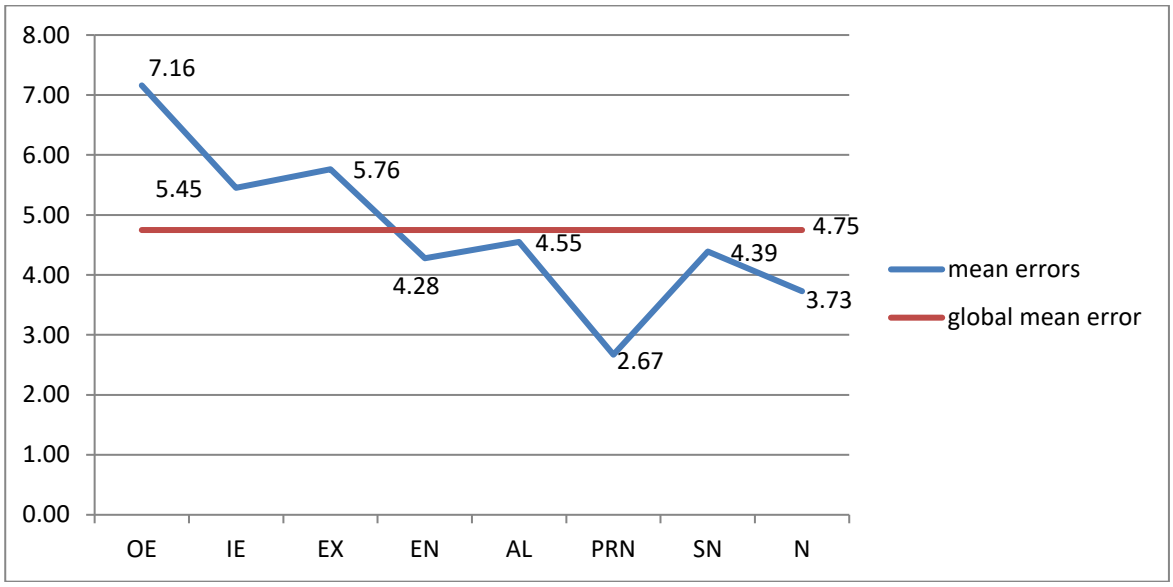


Figure 18. Mean localization errors [mm] for each landmark.

4 Discussion

The worst results were obtained for the outer eyebrows (7.16 mm), while the best result was gained for the pronasale (2.67 mm), followed by the nasion (3.73 mm), the endocanthion (4.28 mm), and the subnasale (4.39 mm). Overall, the mean global error for all landmark is 4.75 mm. Only outer eyebrows were out of the accuracy limit, while all other points ranged within 6 mm. Overall, the mean error is under 10 mm for 94.90 % of the tested facial scans; is under 6.3 mm for 82.52 %.

Despite the evident advances of the newly proposed method with respect to those of our group presented in [24] [25] [26], a comparison between obtained results is due. The study presented in [24] proposes a 9 landmark extraction methodology which was tested on our private database of 79 adult front-view faces. 95% of landmarks have been evaluated as correctly localized by a maxillofacial surgeon. The study was completed by that presented in [25], which introduced 8 new landmarks in the extraction methodology. 85% of the landmarks, evaluated on 58 adult faces of our private database, obtained mean error < 5 mm. A similar methodology was adopted in [26] to detect 13 landmarks on 9 foetuses' faces. Global mean error was 2 mm, which gives a 100% localization accuracy, considering foetuses' faces size (the foetuses were at 24th gestational week). The summed-up comparisons are reported in Table 5.

Table 5. Comparison between results of landmarking approaches presented in our previous studies.

Study	#landmarks	#tested faces	tested database	accuracy
[24]	9	79	private database	95% considered "well localized"
[25]	8	58+54	private database + Bosphorus (expressive)	85% landmarks with mean error < 5 mm
[26]	13	9	private database	mean error = 2 mm (foetuses's faces are half-sized than adults)
This study	13	3132 + 230	Bosphorus (expressive and occluded) + private database	mean error = 4.75 mm 94.90% faces with mean error < 10 mm 82.52% faces with mean error < di 6.3 mm

The results gained with the method presented in this study show enhanced accuracy (4.7 mm mean error) and soundness (3362 faces dataset). While previous contributions were preliminary and heuristic approaches to landmarking, this study was obtained via large experimentations and its results are effective.

Localization accuracies could not be directly compared to other current works of other authors due to the adoption of different testing databases, landmark sets, and localization error types. A few examples: a 11/3/8 mm threshold was used to determine the accuracy depending on facial area in [12]; a 10 mm threshold was used in [13]; reported success rates are obtained by accepting points within 10 per cent of inter-ocular distance to the ground truth in [32].

Considering the studies with comparable scenarios, we may mention: Perakis et al. [18], who obtained a mean error under 6.3 mm (the mean error was under 10 mm in 90.4% cases); Lei et al. [20], with mean localization errors smaller than 4 mm; Gilani et al. [23], with errors ranging from 4.26 to 5.85 mm. Sukno et al. [33] achieved an overall error of 4.81 and 4.25 mm for data with and without occlusions, respectively, on the Bosphorus database. Our mean localization error of 4.75 mm is consistent with these state-of-the-art accuracies and the percentage of faces under 10 mm localization error is 94.90%, thus outperforming the result obtained in [18].

5 Conclusions

An automatic 3D landmark localization method is presented in this study. The algorithm has been designed to be robust to occlusions and different facial expressions. The method relies on thresholds, set experimentally on a training set of 90 faces, applied to point-by-point facial maps of the geometrical descriptors and was tested on 3362 faces (expressive, neutral, eyes-occluded, mouth-occluded, and with glasses) of the 3D Bosphorus database and of an internal database including 32 subjects performing the 6 basic emotions. The final global results show a mean localization error of 4.75 mm, which lays within the state of the art accuracy.

The methodology potentials could be fostered with the embedding of bidimensional information such as RGB channels, whose feasibility has already been tested [8] and are currently under further investigation for enhancing accuracy, especially for Face Recognition purposes. New geometrical descriptors have recently been designed [30]. The purpose was to build descriptors based on primary ones which could be even more tailored for describing landmark local surface properties. Their applicability is now under testing. Both RGB channels and new geometrical descriptors information could merge into an hybrid 2D+3D feature-based approach. They could be adopted as features for a

convolutional neural network methodology both for automatic landmarking and recognition. Also, we planned to widen both training and testing datasets by embedding FRGC v2.0 and BU-3DFE databases.

The idea is to build a sound landmarking methodology which could be transversal through different facial conditions and data types and that could be adopted as first step of Face Verification/Identification and emotion recognition procedures. The final aim is to reach real-time applicability in different scenarios, including real-world conditions "in the wild".

Reference

[1] Leslie G. Farkas, *Anthropometry of the Head and Face in Medicine*. New York: Elsevier North Holland, Inc, 1981.

[2] Leslie G. Farkas, *Anthropometry of the head and face (ed 2)*. New York: Raven Press, 1994.

[3] G. R. Swennen, F. A. Schutyser, and J. E. Hausamen, *Three-dimensional cephalometry: a color atlas and manual.*: Springer Science & Business Media, 2005.

[4] S. Ren, X. Cao, Y. Wei, and J. Sun, "Face alignment at 3000 fps via regressing local binary features," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1685-1692, 2014.

[5] P. Szeptycki, M. Ardabilian, and L. Chen, "Nose tip localization on 2.5 D facial models using differential geometry based point signatures and SVM classifier," *BIOSIG-Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG)*, pp. 1-12, September 2012.

[6] I.A. Kakadiaris, "A third dimension in face recognition," *DOI: 10.1117/2.1201205.004214*, 2012.

[7] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: A survey," *Pattern Recognition Letters*, vol. 28, no. 14, pp. 1885-1906, 2007.

[8] E. Vezzetti, F. Marcolin, S. Tornincasa, and P. Maroso, "Application of geometry to RGB images for facial landmark localisation - a preliminary approach," *International Journal of Biometrics*, vol. 8, no. 3/4, pp. 216-236, 2016.

[9] J. J. Koenderink and A. J. van Doorn, "Surface shape and curvature scales," *Image and vision computing*, vol. 10, no. 8, pp. 557-564, 1992.

[10] P. Bagchi, D. Bhattacharjee, M. Nasipuri, and D. K. Basu, "A novel approach to nose-tip and eye corners detection using HK curvature analysis in case of 3D images," *Third International Conference on Emerging Applications of Information Technology (EAIT)*, pp. 311-315, November 2012.

[11] H. Boukamcha, M. Elhallek, M. Atri, and F. Smach, "3D face landmark auto detection," *World Symposium on Computer Networks and Information Security (WSCNIS)*, pp. 1-6, September 2015.

[12] N. De Giorgis, L. Rocca, and E. Puppo, "Scale-Space Techniques for Fiducial Points Extraction from 3D Faces," *Image Analysis and Processing—ICIAP*, pp. 421-431, 2015.

[13] Y. Li, Y. Wang, B. Wang, and L. Sui, "Nose tip detection on three-dimensional faces using pose-invariant differential surface features," *Computer Vision IET*, vol. 9, no. 1, pp. 75-84, 2014.

[14] C. Dorai and A. K. Jain, "COSMOS-A representation scheme for 3D free-form objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 10, pp. 1115-1130, 1997.

[15] S. Canavan and L. Yin, "Feature detection and tracking on geometric mesh data using a combined global and local shape model for face analysis," *IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)* , pp. 1-6, September 2015.

[16] S. Canavan, P. Liu, X. Zhang, and L. Yin, "Landmark localization on 3D/4D range data using a shape index-based statistical shape model with global and local constraints," *Computer Vision and Image Understanding*, vol. 139, pp. 136-148, 2015.

[17] H., Ding, H., Huang, D., Wang, Y., Zhao, X. Li, J. M. Morvan, and L. Chen, "An efficient multimodal 2D+ 3D feature-based approach to automatic facial expression recognition," *Computer Vision and Image Understanding*, vol. 140, pp. 83-92, 2015.

[18] P. Perakis, G. Passalis, T. Theoharis, and I. A. Kakadiaris, "3D facial landmark detection under large yaw and expression variations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1552-1564, 2013.

[19] P. Perakis, T. Theoharis, and I. A. Kakadiaris, "Feature fusion for facial landmark detection," *Pattern Recognition*, vol. 47, no. 9, pp. 2783-2793, 2014.

[20] J. Lei, X. You, and M. Abdel-Mottaleb, "Automatic Ear Landmark Localization, Segmentation, and Pose Classification in Range Images," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no. 2, pp. 165 - 176, 2016.

[21] C. Creusot, N. Pears, and J. Austin, "IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)," *3D landmark model discovery from a registered set of organic shapes*, pp. 57-64, June 2012.

[22] C. Creusot, N. Pears, and J. Austin, "A machine-learning approach to keypoint detection and landmarking on 3D meshes," *International Journal of Computer Vision*, vol. 102, no. 1-3, pp. 146-179, 2013.

[23] S. Zulqarnain Gilani, F. Shafait, and A. Mian, "Shape-Based Automatic Detection of a Large Number of 3D Facial Landmarks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4639-4648, 2015.

[24] E. Vezzetti and Marcolin F., "Geometry-based 3D face morphology analysis: soft-tissue landmark formalization," *Multimedia Tools and Applications*, pp. 1-35, 2012.

[25] E. Vezzetti and F. Marcolin, "3D Landmarking in Multiexpression Face Analysis: A Preliminary Study on Eyebrows and Mouth," *Aesthetic Plastic Surgery*, vol. 38, pp. 796–811, 2014.

[26] E. Vezzetti, D. Speranza, F. Marcolin, and G. Fracastoro, "Exploiting 3D Ultrasound for Fetal Diagnosis Purpose through Facial Landmarking," *Image Analysis & Stereology*, vol. 33, no. 3, pp. 167-188, 2014.

[27] E. Vezzetti and Marcolin F., "Geometrical descriptors for human face morphological analysis and recognition," *Robotics and Autonomous Systems*, vol. 60, no. 6, pp. 928-939, 2012.

[28] E. Vezzetti, D. Speranza, F. Marcolin, and G. Fracastoro, "Diagnosing cleft and lip palate pathology in 3D ultrasound: a landmarking-based approach ," vol. under review.

[29] E. Vezzetti, F. Marcolin, and G. Fracastoro, "3D face recognition: An automatic strategy based on geometrical descriptors and landmarks," *Robotics and Autonomous Systems*, vol. 62, no. 12, pp. 1768-1776, 2014.

[30] E. Vezzetti and F. Marcolin, "Novel descriptors for geometrical 3D face analysis," *Multimedia Tools and Applications*, vol. X, pp. 1-30, 2016.

[31] E. Vezzetti, S. Moos, F. Marcolin, and V. Stola, "A pose-independent method for 3D face landmark formalization," *Computer Methods and Programs in Biomedicine*, vol. 198, no. 3, pp. 1078-1096, 2012.

[32] H. Dibeklioglu, A. A. Salah, and T. Gevers, "Robust Landmark Localization and Tracking for Improved Facial Expression Analysis".

[33] F. M. Sukno, J. L. Waddington, and P. F. Whelan, "3-d facial landmark localization with asymmetry patterns and shape regression from incomplete local features," *IEEE transactions on cybernetics*, vol. 45, no. 9, pp. 1717-1730, 2015.