

Dispersion assessment in the location of facial landmarks on photographs

B. R. Campomanes-Álvarez · O. Ibáñez · F. Navarro ·
I. Alemán · O. Córdón · S. Damas

Received: 19 December 2013 / Accepted: 16 April 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract The morphological assessment of facial features using photographs has played an important role in forensic anthropology. The analysis of anthropometric landmarks for determining facial dimensions and angles has been considered in diverse forensic areas. Hence, the quantification of the error associated to the location of facial landmarks seems to be necessary when photographs become a key element of the forensic procedure. In this work, we statistically evaluate the inter- and intra-observer dispersions related to the facial landmark identification on photographs. In the inter-observer experiment, a set of 18 facial landmarks was provided to 39 operators. They were requested to mark only those that they could precisely place on 10 photographs with different poses (frontal, oblique, and lateral views). The frequency of landmark location was studied together with their dispersion. Regarding the intra-observer evaluation, three participants identified 13 facial points on five photographs classified in the frontal and oblique views. Each landmark location was repeated five times at intervals of at least 24 h. The frequency

results reveal that glabella, nasion, subnasale, labiale superius, and pogonion obtained the highest location frequency in the three image categories. On the contrary, the lowest rate corresponds to labiale inferius and menton. Meanwhile, zygion, gonion, and gnathion were significantly more difficult to locate than other facial landmarks. They produced a significant effect on the dispersion depending on the pose of the image where they were placed, regardless of the type of observer that positioned them. In particular, zygion and gonion presented a statistically greater variation in the three image poses, while the location of gnathion is less precise in oblique view photographs. Hence, our findings suggest that the latter landmarks tend to be highly variable when determining their exact position.

Keywords Forensic anthropology · Facial morphology · Facial landmarks · Landmark location · Observer dispersion

Electronic supplementary material The online version of this article (doi:10.1007/s00414-014-1002-4) contains supplementary material, which is available to authorized users.

B. R. Campomanes-Álvarez (✉) · O. Ibáñez · O. Córdón ·
S. Damas
European Centre for Soft Computing, 33600 Mieres, Asturias, Spain
e-mail: rosario.campomanes@softcomputing.es

F. Navarro · I. Alemán
Physical Anthropology Laboratory, University of Granada,
18012 Granada, Spain

O. Córdón
Department of Computer Science and Artificial Intelligence,
University of Granada, 18014 Granada, Spain

O. Córdón
Research Center on Information and Communication Technologies
(CITIC-UGR), University of Granada, 18014 Granada, Spain

Introduction

The study of facial morphology from photographs has produced a great interest in forensic anthropology over the years [5]. The analysis of the anthropometric landmarks to define dimensions, proportions, and facial characteristics from photographs has been considered in diverse forensic areas such as identification of living individuals [19], age estimation [10], or craniofacial identification [1, 23].

Regarding the identification of living individuals, face recognition is an important task that forensic experts perform during their investigation when there is a video or image available from a crime scene. Facial anthropometric features, calculated from facial landmarks, are used to verify the resemblance between two individuals by comparing photographs and living persons [11].

The facial geometry characterized by landmarks extracted from 2D images can also be a strong indicator of age progression. Based on this premise, several approaches have been proposed for age estimation and face verification across aging [3, 24, 25].

Within craniofacial identification, craniofacial superimposition (CFS) [26] involves the superimposition of a skull (or a skull model) with a number of *ante mortem* images of an individual and the analysis of their morphological correspondence. Many approaches use the craniometric and cephalometric landmark locations to guide the skull-face overlay process and/or to assess the skull-face relationship once they are superimposed [6].

The analysis of facial morphology using photographs show measurement errors due to subjective analysis, magnification and projection errors, or differences in head orientation [8]. Thus, landmarks present a challenge in terms of assessing observer-induced measurement variations.

Meanwhile, it has long been recognized that not all landmarks are equally identifiable. Bookstein distinguishes three types of landmarks in [2]. Type 1 includes landmarks at which three different tissues meet, e.g., nasion, bregma, and asterion. Type 2 defines points of maximum curvature or other local morphogenetic processes, usually with a biomechanical implication like a muscle attachment site, such as ectocanthion and prosthion. Finally, type 3 refers to extremal landmarks, which belong to a curve or surface. Some type 3 landmarks are gnathion, gonion, and glabella. Therefore, there is a good reason to suspect that the identification precision differs among landmarks [4].

The correct position of facial landmarks is fundamental for any further analysis of faces [5]. Hence, the study and quantification of the accuracy in locating those landmarks is necessary when either photographs become the only resort or they play an important role for the forensic assessment.

Many studies have analyzed the precision and repeatability of anthropometric landmarks directly acquired from crania or extracted from surface imaging methods as conventional cephalometric views, computed tomography, cone-beam computed tomographic volumes [17, 21, 22]. However, the quantification of the error caused by identifying facial landmarks on photographs has only just been recently studied by Cummaudo et al. in [5]. In that work, they evaluated the inter- and intra-observer landmark dispersions on two photographs of the same person in frontal and lateral views and on eight photographs of different sex and age subjects. In the inter-observer experiment, 24 operators located 18 facial landmarks on the frontal view photographs and 11 points on the lateral ones. For the intra-observer analysis, three operators repeated the latter procedure on the

same photographs 20 times at intervals of 24 h. Their results showed that gonion, zygion, and frontotemporale were placed with the highest dispersion in frontal view images, while gnathion, pogonion, and tracion carried the largest dispersion in lateral view photographs. On the other hand, the least dispersion corresponded to pupil, cheilion, endocanthion, and stomion in frontal view photographs and selion, pronasale, and subnasale in lateral ones. The authors confirmed that few anatomical points can be defined with the highest accuracy and showed the importance of the preliminary investigation of reliability in positioning facial landmarks. However, the latter study did not analyze the landmark identification accuracy in terms of statistical significance.

The aim of our work is to extend the latter study by evaluating the dispersion related to the location of facial landmarks on 2D images including a statistical test, as well as analyzing the frequency of marked landmarks depending on the image type. We have performed statistical inter- and intra-observer analyses to assess the dispersion related to the landmark identification on photographs taking into account the specific located landmark type, the orientation of the face in the photograph, and the observer experience. The inter-observer dispersion has been calculated for 18 landmarks, three image poses (frontal, oblique, and lateral views), and two types of observers (expert and student). In the case of the intra-observer experiment, we have analyzed the dispersion for 13 landmarks in five photographs classified in the frontal and oblique views. Three observers have located the facial points five times at intervals of at least 24 h.

Materials and methods

In order to mimic the conditions of a real forensic scenario, the photographs considered in the experiments corresponded to CFS cases addressed by the staff of the Physical Anthropology Laboratory at the University of Granada, Spain, in collaboration with the Spanish Scientific Police. The resolution of the 2D images was at least 640×480 pixels as suggested in [5, 18]. A set of facial landmarks commonly employed in CFS was provided to the observers, and they were requested to mark only those that they could precisely locate.

The 2D coordinates of the facial landmarks were placed on the photographs using the Landmarker™ software. Landmarker™ is an open-license application developed by our team to allow users to handle and incorporate photographs of forensic cases to the database for a subsequent analysis. Facial landmarks can be located in either a precise (using a point) or an imprecise (using an ellipse) way. In the current study, we only consider precise landmark location. Once a landmark is marked, the user has to select its name from a predefined list. At any time, the user can save the work done,

obtain a file with the landmark position in coordinates, and/or save the image with the landmarks. For each photograph and user, all landmark positions are stored in a remote database that will be used for the whole study analysis. Figure 1 shows a snapshot of the landmark location function in Landmarker™.

Once all the coordinates were recorded, they were exported to a Microsoft Excel® spreadsheet. Inter- and intra-observer analyses were performed to assess the error related to the location of facial landmarks. These statistical tests were conducted using the XLStats® Excel® add-in [12].

Inter-observer study design

Thirty-nine observers participated in this study. Thirty-one were forensic experts and eight master students. Ten facial photographs classified by face orientation (frontal=3, right oblique=3, left oblique=3, left lateral=1) were considered for the experiment. The images correspond to Mediterranean subjects of different sex, aged between 15 to 85 years old, without facial disorders. Eighteen landmarks (Table 1 and Fig. 2) were included in the analysis.

Intra-observer study design

Three observers (two forensic experts and one master student) identified 13 facial landmarks (Table 2) on five photographs of different subjects. The images were classified by face orientation (frontal=2, right oblique=1, left oblique=2). Each landmark location was repeated by the observers five times at intervals of at least 24 h. As in the other study, the photographs corresponded to Mediterranean subjects of different sex, aged between 15 to 85 years old, without facial disorders.

Statistical analysis

In both the inter- and intra-observer analyses, the dispersion has been calculated as follows.

Each expert places a landmark in a certain coordinate x and y , being the reference coordinates ($x=0$, $y=0$) on the top-left corner of the photograph. The mean coordinates in the X - and in the Y -axes were determined for every landmark. The mean in the X - and in the Y -axes refers to the mean of the x and y coordinates respectively, located by all the experts. The absolute value of the difference between each observed coordinate respect to the mean (DRM) was computed in the two axes [17]. Once all this information was gathered, we calculated a

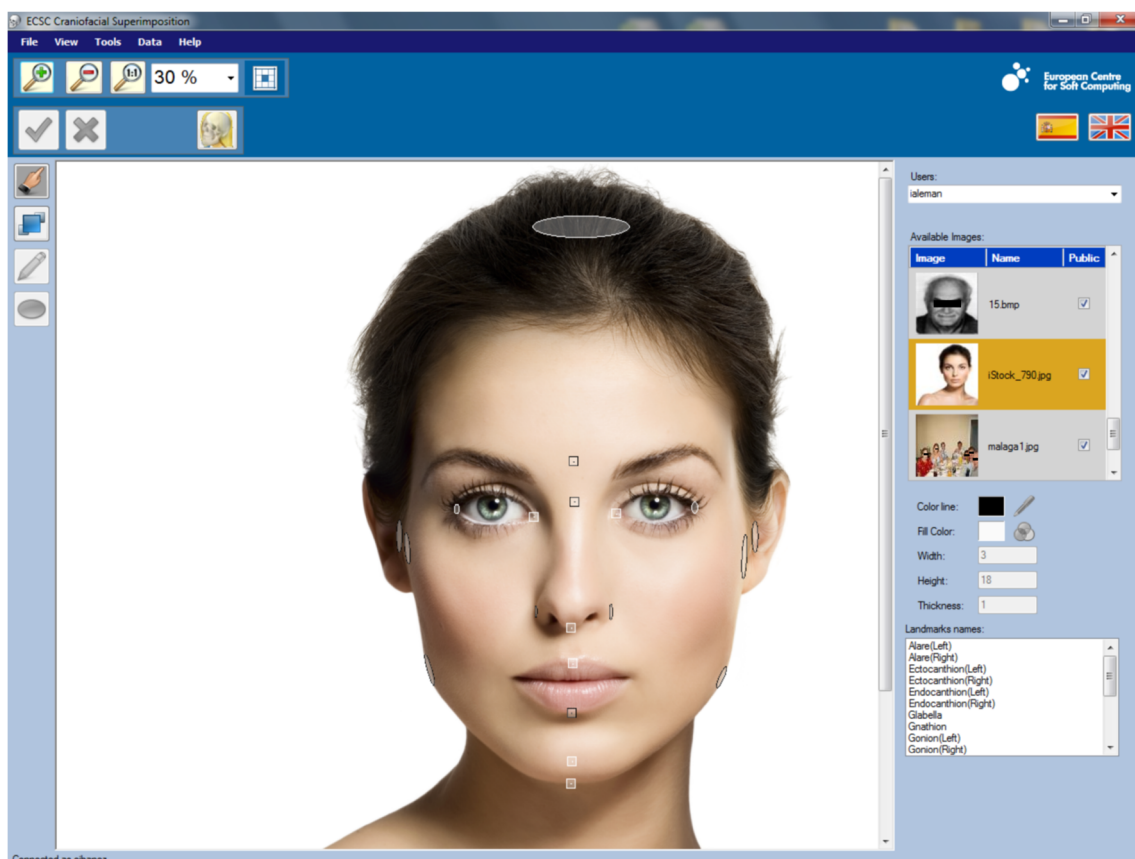


Fig. 1 Facial landmarks marked on a photograph using the Landmarker software

Table 1 The 18 landmarks used in the inter-observer study, together with their abbreviations, the type as defined in [2], and the anatomical description taken from [8, 14]

Landmarks	Type	Anatomical description
Glabella (g)	3	In the midline, the most prominent point between the eyebrows
Nasion (n)	1	The midpoint on the soft tissue contour of the base of the nasal root at the level of the frontonasal suture
Endocanthion left, right (enl, enr)	1	The point at the inner commissure (medial canthus) of each palpebral fissure
Exocanthion left, right (exl, exr)	2	The soft tissue point located at the outer commissure of each eye fissure
Alare left, right (alal, alar)	3	The most lateral point on each alar contour where nose meets the skin of the philtrum and cheek
Subnasale (sn)	1	The midpoint on the nasolabial soft tissue contour between the columella crest and the upper lip
Zygion left, right (zyl, zyr)	3	The most lateral point on the soft tissue contour of each zygomatic arch
Labiale superius (ls)	2	The midpoint on the vermilion line of the upper lip
Labiale inferius (li)	2	The midpoint of the vermilion line of the lower lip
Gonion left, right (gol, gor)	3	The most lateral point of the jaw line at the mandibular angle
Pogonion (pg)	3	The most anterior midpoint on the soft tissue chin
Gnathion (gn)	3	The most inferior point on the soft tissue contour of the chin
Menton (me)	3	The most inferior point on the soft tissue contour of the chin ^a

^a Due to the fact that some experts have distinguished between gnathion and menton, we have considered gnathion as the midpoint point on the soft tissue chin between pogonion and menton for those cases

single spread parameter that corresponds to the total difference mean (TDM) in the X and Y directions. As explained in [5], the final dispersion values were determined by normalizing the TDMs. In our case, the highest dispersion was equivalent to 1.00.

Table 3 summarizes the operations to obtain the final dispersion: observation refers to a certain location of a landmark by an observer; n is the total number of observers that marked a particular landmark (defined in the second column of the table), x_i and y_i are the stored coordinates for each observation, \bar{x} and \bar{y} are the mean values of the x and y coordinates respectively, DRM_{x_i} and DRM_{y_i} correspond to the difference respect to the mean of x_i and y_i respectively, TDM_i is the total difference mean of the observation i , $Norm$ refers to the normalization (i.e., the scaling of the values

between zero and one), and $Dispersion_i$ is the dispersion for the observation i .

The inter-observer dispersion was calculated for 18 landmarks, three types of images (frontal, oblique, and lateral views), and two types of observers (expert and student).

Due to the inequality of number of observations per cell, an analysis of variance (ANOVA) for unbalanced data [9, 15, 16] was computed. In the inter-observer study, we modeled our experiment using three factors. The landmark was defined as a factor (also called main effect) whose levels are the different landmarks located by the experts; the type of image was a three-level factor: front view, lateral view, and oblique view, and the type of observer was a factor with two levels, expert and student. Hence, the qualitative independent variables

Fig. 2 Facial landmarks considered in the experiment and their location in frontal (a) and lateral view (b)

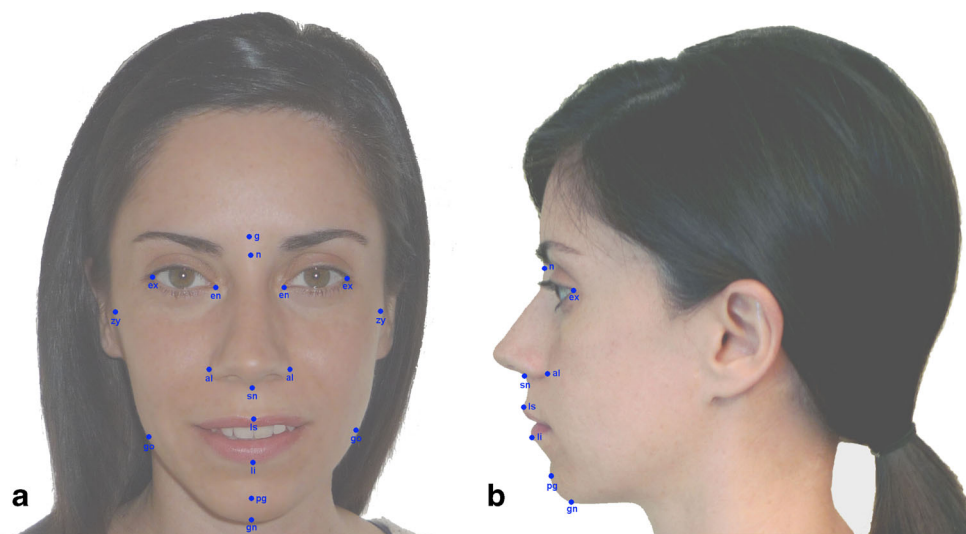


Table 2 The 13 landmarks used in the intra-observer study, together with their abbreviations

Landmarks
Glabella (g)
Nasion (n)
Endocanthion left, right (enl, enr)
Exocanthion left, right (exl, exr)
Subnasale (sn)
Zygion left, right (zyl, zyr)
Labiale superius (ls)
Gonion left, right (gol, gor)
Gnathion (gn)

were the landmark, the type of image, the type of expert, and the interaction between the landmark and the type of image. The quantitative dependent variable (variable to model) was the dispersion.

Regarding the intra-observer study, the dispersion was calculated for 13 landmarks and two types of images (frontal and oblique). For each of the three observers who participated in the experiment, a two-way ANOVA for unbalanced data was performed considering factors such as the landmark, the type of image, and the interaction between the landmark and the type of image, whose levels were the same as in the inter-observer study. Thus, the qualitative independent variables were the landmark, the type of image, and the interaction between the landmark and the type of image. The quantitative dependent variable (variable to model) was the dispersion, as well.

The considered level of statistical significance was $\alpha=0.05$ for inter- and intra-observer analyses. The specific aim was to test the null hypothesis (H_0) stating that the dispersion of landmark identification does not depend on the specific landmark, type of image, their interaction, and type of observer (the latter, only in the case of the inter-observer experiment). Our procedure has consisted of performing type III sum of squares (SS) ANOVA to test the null hypothesis. In the case that no significant interaction was present, we have continued the analysis by applying type II SS ANOVA

test to the data. Further details about ANOVA tests are available in the *The Analysis of the Variance ANOVA* Section of the [Online Resource](#).

Results

Inter-observer study

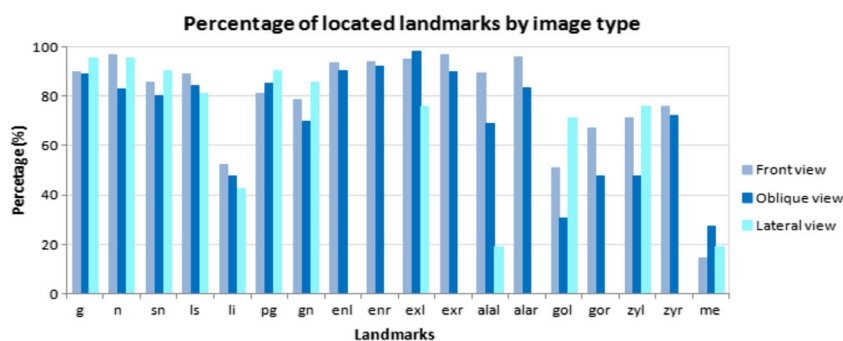
From the results obtained, we observed that not all the operators marked every landmark in every image. Figure 3 depicts the frequency of landmark location (as a percentage) depending on the image type. Glabella, nasion, subnasale, labiale superius, and pogonion obtain the highest location frequency in the three categories, at least marked in 80 % of the cases. By contrast, the lowest rate corresponds to labiale inferius and menton, 50 and 25 %, respectively. The two endocanthion, exocanthion, and alare achieve a frequency larger than 70 % in the frontal and the oblique views. Besides, both gonion present more differences in the location frequency among image types. Note that some landmarks (i.e., enl, enr, exr, and alar) have not been located due to the left orientation of the lateral photograph included in the analysis, which constitutes an appropriate behavior in these cases.

Next, we will focus on the results of the dispersion statistical analyses performed with the inter-observer data collection. The coefficient of determination R^2 was 0.81. In that case, 81 % of the variability is explained. The remaining 19 % are hidden in other variables, which the model classifies as “random effects”. The figures of the type III SS test are shown in Table 4. The analysis reveals a significant interaction (the p value=0.003 is lower than the considered level of statistical significance of 0.05) between the landmark and the image type variables. An interaction between factors with statistical significance implies that the effect of each factor varies with the level of the other. Thus, the effect of the landmark variable on the dispersion depends on the level (frontal, oblique, or lateral) of the image type variable. Under these circumstances, a main effect represents an average over the heterogeneous effects of one factor over the levels of the other [9, 15]. Hence, despite the landmark variable obtains a statistical significance (p value <0.0001), its effect on the dispersion is related to the image type where the landmark has been marked.

Table 3 Parameters and their corresponding formulas for calculating the dispersion

Observation	Landmark	x_i	y_i	DRM_{xi}	DRM_{yi}	TDM_i	Dispersion _{<i>i</i>}
1	Glabella	x_1	y_1	$ x_1 - \bar{x} $	$ x_1 - \bar{y} $	$\frac{DRM_{x1} + DRM_{y1}}{2}$	Norm(TDM_1)
2	Glabella	x_2	y_2	$ x_2 - \bar{x} $	$ x_2 - \bar{y} $	$\frac{DRM_{x2} + DRM_{y2}}{2}$	Norm(TDM_2)
...
n	Glabella	x_n	y_n	$ x_n - \bar{x} $	$ x_n - \bar{y} $	$\frac{DRM_{xn} + DRM_{yn}}{2}$	Norm(TDM_n)

Fig. 3 Frequency of landmark location (in %) depending on the image type



Due to the fact that the interaction between the landmark and the type of image variables was significant, we have considered the type III SS test because this approach is valid in the presence of significant interactions [9].

No significant effect on the dispersion was found for the rest of variables.

Table 5 shows the values of the estimated model (see Equation 2 in the Online Resource) calculated by the ANOVA test, which correspond to the landmark variable. The ANOVA estimated model can be used to analyze the impact of the independent variables or factors on the dependent variable, specifically what levels of each factor present a higher influence on the modeled variable.

As we can see, the gnathion, the two gonia, and the two zygia present the highest impact on the dispersion (the higher the value of the parameters estimators, the greater the influence on the dependent variable). They also obtain statistical significance with p values lower than 0.05. As previously explained, the effect of those landmarks on the dispersion is conditioned to the image type due to the interaction between landmark and image type variables (Table 4).

Table 6 depicts the values of the estimated model (see Equation 3 in the Online Resource) calculated by the ANOVA test, which correspond to the interaction between landmark and image type at the p value lower than 0.05. The gnathion, the left gonia in oblique and lateral view photographs, and the two zygia in oblique view present the highest impact on the dispersion (the higher the value of the parameters estimators, the greater the influence on the dependent variable).

In the case of the frontal view, the landmarks that have significantly influenced the dispersion variable are gonia and

zygia. The location variability is higher for zygia than for other facial points.

Regarding the oblique view, significantly greater dispersion was found for the two gonia, two zygia, and gnathion. In particular, zygia and gonia are less precise to locate (p values < 0.0001) than gnathion (p value = 0.001).

In the lateral view, the dispersion is statistically higher for placing gnathion, gonion, and zygion. The results suggest that zygion (p value = 0.030) is easier to identify than gnathion and gonion (p values < 0.0001). Note that some landmarks (i.e., zyr, gor) have not been marked due to the left orientation of the lateral photograph included in the analysis.

As mentioned in [5, 7] the high variability of gonia and zygia can be due to the fact that these landmarks are difficult to

Table 5 Inter-observer dispersion results. Model values obtained by the ANOVA test, corresponding to the landmark variable (significant landmarks and p values < 0.05 are shown in bold)

Parameter	Value	Standard deviation	t	Pr > t
alal	0.013	0.008	1.729	0.084
alar	0.004	0.010	0.420	0.674
exol	0.006	0.010	0.625	0.532
exor	0.005	0.010	0.487	0.626
endol	0.005	0.010	0.550	0.583
endor	0.007	0.010	0.713	0.476
gb	0.006	0.011	0.531	0.595
gn	0.034	0.011	3.259	0.001
gol	0.030	0.011	2.622	0.009
gor	0.031	0.010	3.029	0.002
li	0.0001	0.012	0.036	0.971
ls	0.006	0.010	0.576	0.564
me	0.006	0.027	0.225	0.822
n	0.009	0.010	0.924	0.356
pg	0.017	0.010	1.731	0.084
sn	0.005	0.010	0.540	0.589
zyl	0.134	0.010	3.100	<0.0001
zyr	0.035	0.010	3.395	0.001

Table 4 Type III SS ANOVA results for the inter-observer analysis. Statistical significance (p values < 0.05) in bold

Source	Df	Sum of squares	F value	Pr (> F)
Landmark	17	1.252	17.414	<0.0001
Image type	2	0.014	1.666	0.189
Observer type	1	0.0001	0.039	0.844
Landmark*image type	28	0.225	1.897	0.003

Table 6 Inter-observer dispersion results. Estimated model values obtaining by the ANOVA test, corresponding to the interaction between landmark and image type parameter with a p value <0.05 (significant interactions and p values <0.05 are shown in bold)

Parameter	Value	Standard deviation	t	$Pr> t $
gn*oblique view	0.034	0.011	3.259	0.001
gn*lateral view	0.079	0.170	4.671	<0.0001
gol*frontal view	0.030	0.011	2.621	0.009
gol*oblique view	0.086	0.017	5.122	<0.0001
gol* lateral view	0.078	0.018	4.280	<0.0001
gor*frontal view	0.031	0.010	3.028	0.002
gor*oblique view	0.074	0.013	5.723	<0.0001
zyl*frontal view	0.134	0.010	13.003	<0.0001
zyl*oblique view	0.118	0.012	9.839	<0.0001
zyl*lateral view	0.023	0.038	2.165	0.030
zyr*frontal view	0.035	0.010	3.396	0.001
zyr*oblique view	0.080	0.010	7.625	<0.0001

locate without palpation when the soft tissue covering the bony landmark is very thick.

Intra-observer study

Three ANOVA tests have been computed to analyze the data provided by each observer. We have obtained a coefficient of determination R^2 equal to 0.81, 0.77, and 0.82 for the three ANOVA tests, respectively. Hence, 81, 77, and 82 % of the variability is explained for each intra-observer analysis. The results of the type III SS analyses are described in Table 7. No significant interactions are achieved for any of the three performed analyses; the p values corresponding to the interactions between landmark and image type are 0.943, 0.698, and 0.682 for each observer experiment, respectively. In the absence of a significant interaction (the p value is higher than the considered level of statistical significance of 0.05), the type II SS is statistically more

powerful than the type III SS (see [15] and [Online Resource](#) for further details). Therefore, we will analyze the results of this test depicted in Table 8. The type II SS reports that the landmark variable produces a significant effect on the dispersion for the three studied observers (p values <0.0001). Besides, the image type factor has significantly influenced the dispersion in the experiments of observers 1 and 2 (p values=0.0004 and 0.009, respectively).

Table 9 shows the results of each model in order to know what landmarks and/or type of images have a significant influence on the dispersion. As previously said, the higher the value of the resulting parameter estimators, the greater the influence on the dependent variable.

The analysis reveals that the dispersion obtained by observer 1 is significantly higher in the two zygia (zyl p value=0.001, zyr p value <0.0001), gonion (gol p value=0.001, gor p value <0.0001), gnathion (p value=0.011), and glabella (p value=0.023) than in other landmarks.

Regarding the second observer, significant differences are also found for zygia (zyl p value <0.0001 , zyr p value=0.022), gonion (gol p value <0.0001 , gor p value=0.012) as well as subnasale (p value=0.013). The behavior of operators 1 and 2 is meaningful as they are less precise when locating landmarks in the oblique view images (p values <0.05) than in the frontal ones.

In the case of observer 3, the location of left zygon and gonion has significantly influenced the dispersion at a p value below 0.0001.

Figure 4 depicts the landmark dispersion, for each observer, depending on the orientation of the face in the photographs.

Glabella, gnathion, zygon, and right gonion, located by observer 1, present the highest differences between the frontal and oblique views. The two zygia and gonion are the less precise landmarks. In particular, the first observer had difficulties to place right zygon and gonion in oblique view photographs.

Table 7 Type III SS ANOVA results for the intra-observer analysis (p values <0.05 are shown in bold)

Observer	Source	Df	Sum of squares	F value	$Pr (>F)$
1	Landmark	12	0.805	3.479	<0.0001
	Image type	1	0.175	9.069	0.003
	Landmark*image type	12	0.103	0.447	0.943
2	Landmark	12	0.218	1.357	0.186
	Image type	1	0.077	5.737	0.017
	Landmark*image type	12	0.121	0.754	0.698
3	Landmark	12	0.078	0.926	0.521
	Image type	1	0.019	2.695	0.102
	Landmark*image type	12	0.065	0.770	0.682

Table 8 Type II SS ANOVA results for each intra-observer analysis (p values <0.05 are shown in bold)

Observer	Source	Df	Sum of squares	F value	Pr ($>F$)
1	Landmark	12	2.421	10.463	<0.0001
	Image type	1	0.256	13,298	0.0004
	Landmark*image type	12	0.013	0.656	0.793
2	Landmark	12	0.737	4.578	<0.0001
	Image type	1	0.094	7.015	0.009
	Landmark*image type	12	0.148	0.922	0.525
3	Landmark	12	0.305	3.614	<0.0001
	Image type	1	0.020	2.817	0.094
	Landmark*image type	12	0.068	0.804	0.646

In the case of the dispersion results of the frontal and oblique view images for the second observer, the greatest dispersion differences are found for subnasale and left zygion. Observer 2 has been less precise locating landmarks in oblique images than in frontal ones. Even though no significant differences between types of images were found by the ANOVA test, observer 3

uniformly located all landmarks but endocanthion, left zygion, and gonion in the oblique view.

Discussion and conclusions

The localization of facial landmarks is an important operation for many subsequent face processing tasks. The study of facial morphology from photographs using landmarks is considered in many approaches within forensic anthropology, from age estimation to CFS, as well as to identify a suspect on a video surveillance system or for general personal identification. Within those research lines, important studies are also based on biometrics and facial morphological comparisons instead on using anthropometric indices or proportions from certain facial landmarks [13, 20].

However, the correct position of facial landmarks is fundamental for further analyses of faces from photographs, both from a morphological and metrical point of view; being a complex task which depends upon the variability in positioning facial landmarks [5]. Quantifying the degree of error in the location of facial landmarks from photographs arise as a necessary task when those landmarks are considered in anthropometric studies and facial assessment from 2D images.

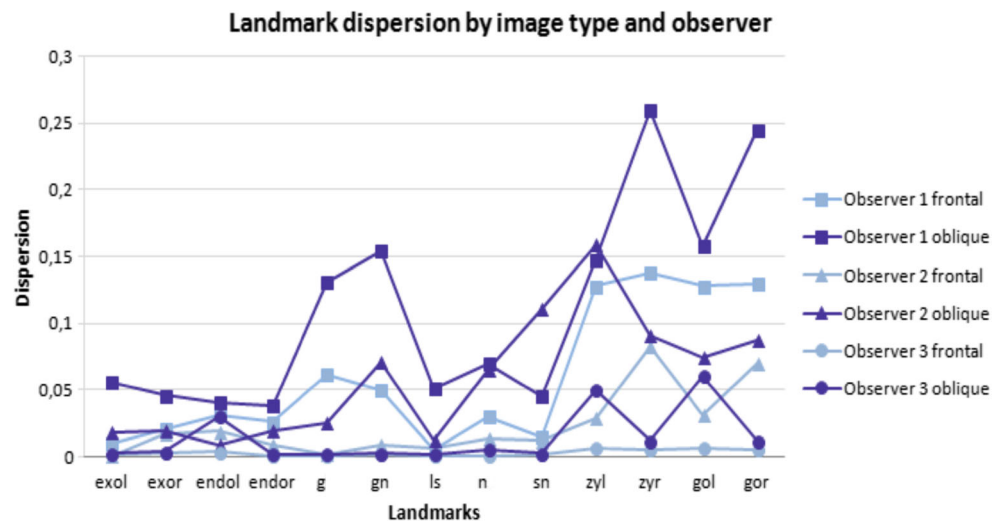
Cummaudo et al. provided the first quantitative results concerning the accuracy in positioning landmarks on photographs [5]. In view of their results, only very few landmarks seem to be reliable for facial assessment, and this is an important limitation of the information which may be extrapolated for further analysis.

Our work is the second study to assess the dispersion related to the location of facial landmarks on 2D images. We have analyzed how the precision of the landmark positioning is affected by the type of landmark, the orientation of the face in the photograph, and the observer experience. Inter- and intra-observer statistical analyses were performed to evaluate that variability.

Table 9 Intra-observer dispersion results. Estimated model values obtained by the ANOVA test, for each intra-observer analysis, corresponding to the landmark and image type parameters, which get a p value <0.05 (p values <0.05 and their observers are shown in bold)

Parameter	Observer	Value	Standard deviation	t	Pr> t
g	1	0.070	0.031	2.281	0.023
	2	0.003	0.026	0.098	0.922
	3	0.010	0.019	0.559	0.577
gn	1	0.078	0.031	2.549	0.011
	2	0.032	0.026	1.230	0.220
	3	0.009	0.019	0.511	0.610
sn	1	0.008	0.031	0.273	0.785
	2	0.064	0.026	2.492	0.013
	3	0.010	0.019	0.530	0.597
zyl	1	0.104	0.031	3.423	0.001
	2	0.06	0.026	4.126	<0.0001
	3	0.075	0.019	4.048	<0.0001
zyr	1	0.193	0.031	6.312	<0.0001
	2	0.059	0.026	2.306	0.022
	3	0.001	0.019	0.078	0.938
gol	1	0.104	0.031	3.423	0.001
	2	0.106	0.026	4.126	<0.0001
	3	0.065	0.019	4.548	<0.0001
gor	1	0.193	0.031	6.312	<0.0001
	2	0.049	0.025	2.206	0.012
	3	0.001	0.019	0.076	0.908
oblique view	1	0.059	0.017	3.550	0.0004
	2	0.034	0.014	2.447	0.015
	3	0.015	0.010	1.502	0.134

Fig. 4 Landmark dispersion according to the type of image and observer



Within the inter-observer experiment, we also studied the frequency of the located landmarks depending on the image type.

The open-license Landmarker™ application has been created to handle and incorporate photographs of forensic cases for subsequent analyses. Hence, that tool can be considered for further studies related to the repeatability and accuracy in positioning landmarks on photographs.

The inter-observer dispersion was calculated for 18 landmarks, three images poses (frontal, oblique, and lateral views), and two types of observers (expert and master student). The results of this analysis indicated that gnathion, the two gonia, and zygia, all of them classified as type 3 landmarks by [2], present statistical significance, but their effect on the dispersion is conditioned to the image type where they have been placed. Gonia and zygia have significantly influenced the dispersion in frontal view images. Significantly, greater variability was also found for the two gonia, zygia, and gnathion in the oblique view. The dispersion was statistically higher for placing gonion, zygon, and gnathion in the lateral photograph. Besides, we obtained no significant differences between students and forensic experts when positioning facial landmarks.

Regarding the frequency analysis, glabella, nasion, subnasale, labiale superius, and pogonion achieved the highest rate in the three image poses, at least marked in 80 % of the cases. On the contrary, the lowest frequency corresponded to labiale inferius and menton, 50 and 25 %, respectively. The two endocanthion, exocanthion, and alare obtained a frequency larger than 70 % in the frontal and the oblique views. Meanwhile, both gonia presented more differences in the location frequency among image types. In general, the frequency of location was higher in landmarks classified as types 1 and 2 than in type 3 ones.

In the case of the intra-observer experiment, the dispersion was calculated for 13 landmarks in five

photographs classified in the frontal and oblique views. Three observers located the facial points five times at intervals of at least 24 h. The intra-observer tests revealed that zygia and gonia are significantly more difficult to locate than other facial landmarks. The results also showed that the first observer had difficulties to place gnathion and glabella, and observer 2 achieved significant differences among the five attempts to locate subnasale. In general, the dispersion was higher in the oblique view than in the frontal one for the three observers, while being statistically significant for the first and the second.

The high variability of gonia and zygia can be due to the fact that these landmarks are difficult to locate without palpation when the soft tissue covering the bony landmark is very thick. This conclusion was also drawn in [5].

Hence, our findings suggest that some type 3 landmarks tend to be highly variable when determining their exact anatomical location. These facial points are associated with a statistically significant degree of error both for inter- and intra-observers. Thus, it would be advisable to consider the variability of these landmarks for further morphological studies from photographs.

Acknowledgments The authors would like to acknowledge the team of the Physical Anthropology Laboratory, University of Granada, Granada, Spain, for their support during the data acquisition. We also thank the international forensic experts group and master students who kindly participated in the online poll by marking the landmarks.

This work has been supported by the Spanish MINECO under the SOCOVIF2 project (refs. TIN2012-38525-C02-01/02, <http://www.softcomputing.es/socovifi/>), and the Juan de la Cierva Fellowship JCI-2012-15359. The Andalusian Department of *Innovación, Ciencia y Empresa* under project TIC2011-7745, the Principality of Asturias Government under the project with reference CT13-55, and the European Union's Seventh Framework Programme for research technological development and demonstration under the MEPROCS project (Grant Agreement No. 285624), including European Development Regional Funds (EDRF).

References

1. Aulsebrook WY, Iscan M, Slabbert J, Beckert P (1995) Superimposition and reconstruction in forensic facial identification: a survey. *Forensic Sci Int* 75(2–3):101–120
2. Bookstein FL (1991) *Morphometric tools for landmark data. Geometry and biology.* Cambridge University Press, Cambridge
3. Cattaneo C (2009) Anthropology: aging the living. In: Moenssens A, Jamieson A (eds) *Wiley encyclopedia of forensic science.* Wiley, Chichester, pp 188–191
4. Cramon-Taubadel N, Frazier BC, Lahr MM (2007) The problem of assessing landmark error in geometric morphometrics: theory, methods, and modifications. *Am J Phys Anthropol* 134(1):24–35
5. Cummaudo M, Guerzoni M, Marasciuolo L, Gibelli D, Cigada A, Obertová Z, Ratnayake M, Poppa P, Gabriel P, Rizt-Timme S, Cattaneo C (2013) Pitfalls at the root of facial assessment on photographs: a quantitative study of accuracy in positioning facial landmarks. *Int J Legal Med* 127:699–706
6. Damas S, Cordon O, Ibanez O, Santamaria J, Aleman I, Botella M et al (2011) Forensic identification by computer-aided craniofacial superimposition: a survey. *ACM Comput Surv* 43(4):27
7. Douglas TS (2004) Image processing for craniofacial landmark identification and measurement: a review of photogrammetry and cephalometry. *Comp Med Imag Graph* 28:401–409
8. Farkas LG, Bryson W, Klotz J (1980) Is photogrammetry of the face reliable? *Plast Reconstr Surg* 66:346–355
9. Fox J (2008) *Applied regression analysis and generalized linear models*, 2nd edn. SAGE Publications, Inc, Thousand Oaks, California
10. Fu Y, Guo GD, Huang TS (2010) Age synthesis and estimation via faces: a survey. *IEEE Trans Pattern Anal Mach Intell* 32(11):1955–1976
11. Halberstein RA (2001) The application of anthropometric indices in forensic photography: three case studies. *J Forensic Sci* 46(6):1438–1441
12. <http://www.xlstat.com/> [Accessed: 2nd March 2014]
13. Knussmann R (1988) Die morphologische Identitätsprüfung (in German). In: Knussmann R (ed) *Anthropologische: Handbuch der vergleichenden Biologie des Menschen.* Gustav Fischer Verlag, Stuttgart, pp 386–405
14. Knussmann R (1988) Somatometrie (in German). In: Knussmann R (ed) *Anthropologie.* Gustav Fischer Verlag, Stuttgart, pp 232–285
15. Langsrud O (2003) ANOVA for unbalanced data: use type II instead of type III sums of squares. *Stat Comput* 13:163–167
16. Little RJA, Rubin DB (1987) *Statistical analysis with missing data.* Wiley, New York
17. Ludlow JB, Gubler M, Cevidanes L, Mol A (2009) Precision of cephalometric landmark identification: cone-beam computed tomography vs conventional cephalometric views. *Am J Orthod Dentofacial Orthop* 136(3):312.e1–e10
18. Ozkul T, Ozkul MH, Akhtar R, Al-Kaabi F, Jumaia T (2009) A software tool for measurement of facial parameters. *Open Chem Biomed Meth J* 2:69–74
19. Rogers SL (1986) *The personal identification of living individuals.* Charles C Thomas Pub Ltd. Springfield, Illinois
20. Rösing FW (2008) *Morphologische Identifikation von Personen. Grundlagen, Merkmale, Häufigkeiten (in German).* In: Buck J, Krumbholz H (eds) *Sachverständigenbeweis im Verkehrsrecht.* Nomos, Baden-Baden, pp 201–319
21. Ross AH, Williams S (2008) Testing repeatability and error of coordinate landmark data acquired from crania. *J Forensic Sci* 53(4):782–785
22. Sholts SB, Flores L, Walker PL, Wärmländer SKTS (2011) Comparison of coordinate measurement precision of different landmark types on human crania using a 3D laser scanner and a 3D digitiser: implications for applications of digital morphometrics. *Int J Osteoarchaeol* 21:535–543
23. Stephan CN (2009) Craniofacial identification: techniques of facial approximation and craniofacial superimposition. In: Blau S, Ubelaker DH (eds) *Handbook of forensic anthropology and archaeology.* Left Coast, California, pp 304–321
24. Thukral P, Mitra K, Chellappa R (2012) A hierarchical approach for human age estimation. In: *Proc IEEE ICASSP, Kyoto*, pp 1529 – 1532
25. Wu T, Turaga P, Chellappa R (2012) Age estimation and face verification across aging using landmarks. *IEEE T Inf Foren Sec* 7(6): 1780–1788
26. Yoshino M (2012) Craniofacial superimposition. In: Wilkinson C, Rynn C (eds) *Craniofacial identification.* University Press, Cambridge, pp 238–253