

# Improved image registration in skull–face overlay using expert knowledge

Oscar Gómez<sup>1</sup>  · Oscar Ibáñez<sup>1</sup> · Oscar Cordon<sup>1</sup>

Received: 24 February 2017 / Accepted: 29 March 2017 / Published online: 18 April 2017  
© Springer-Verlag Berlin Heidelberg 2017

**Abstract** Craniofacial superimposition involves the process of overlaying a skull with a number of ante-mortem images of the face of an individual and the analysis of their morphological correspondence. This research focused on the skull–face overlay stage with the aim of modeling the expert knowledge that is related to the existing anthropometric differences among landmarks and incorporating it into this stage. Consequently, we have moved from a single-objective optimization problem to a multiobjective optimization one aimed to reduce the distances between pairs of landmarks from each group independently. To tackle it, two classic approaches from the area of multicriteria decision making were used: weighted sum and lexicographical order. The results, which were obtained over a ground truth dataset, are promising in those cases where the forensic expert has located a large number of landmarks, and worse results than the state-of-the-art method in cases with few landmarks.

**Keywords** Computer vision · Soft computing · Craniofacial superimposition · Skull–face overlay · Craniometric and cephalometric landmarks

## 1 Introduction

Craniofacial superimposition (CFS) [1] is one of the most relevant skeleton-based identification techniques. It involves the process of overlaying a skull image (or a skull 3D model) with a number of ante-mortem images of an individual and the analysis of their morphological correspondence [2] to try to establish whether they correspond to the same individual. Three consecutive stages for the whole CFS process have been distinguished in [3] (see Fig. 1):

- Acquisition and processing of the face and the skull photographs/models. In some approaches, this step also involves the location of anatomical landmarks on the face and the skull.
- Skull–face overlay (SFO). This step focuses on achieving the best possible superimposition of an image, video-frame or a 3D model of a physical skull and a single ante-mortem image of a missing person.
- Skull–face overlay assessment and decision making, in which the degree of support (strong support, moderate support, limited support and undetermined [4]) that the skull and the available photograph belong to the same person or not (exclusion) is determined.

From the computer vision (CV) point of view, there is a clear relation between the SFO procedure and an image registration (IR) problem [5]. SFO can be tackled following an IR approach in order to superimpose the skull onto the facial photograph. To do so, the most convenient procedure is to guide the IR process by matching the corresponding cranial and facial landmarks.

This matching process involves a really complex optimization task. On the one hand, there is incomplete and vague information guiding the process (matching of two different

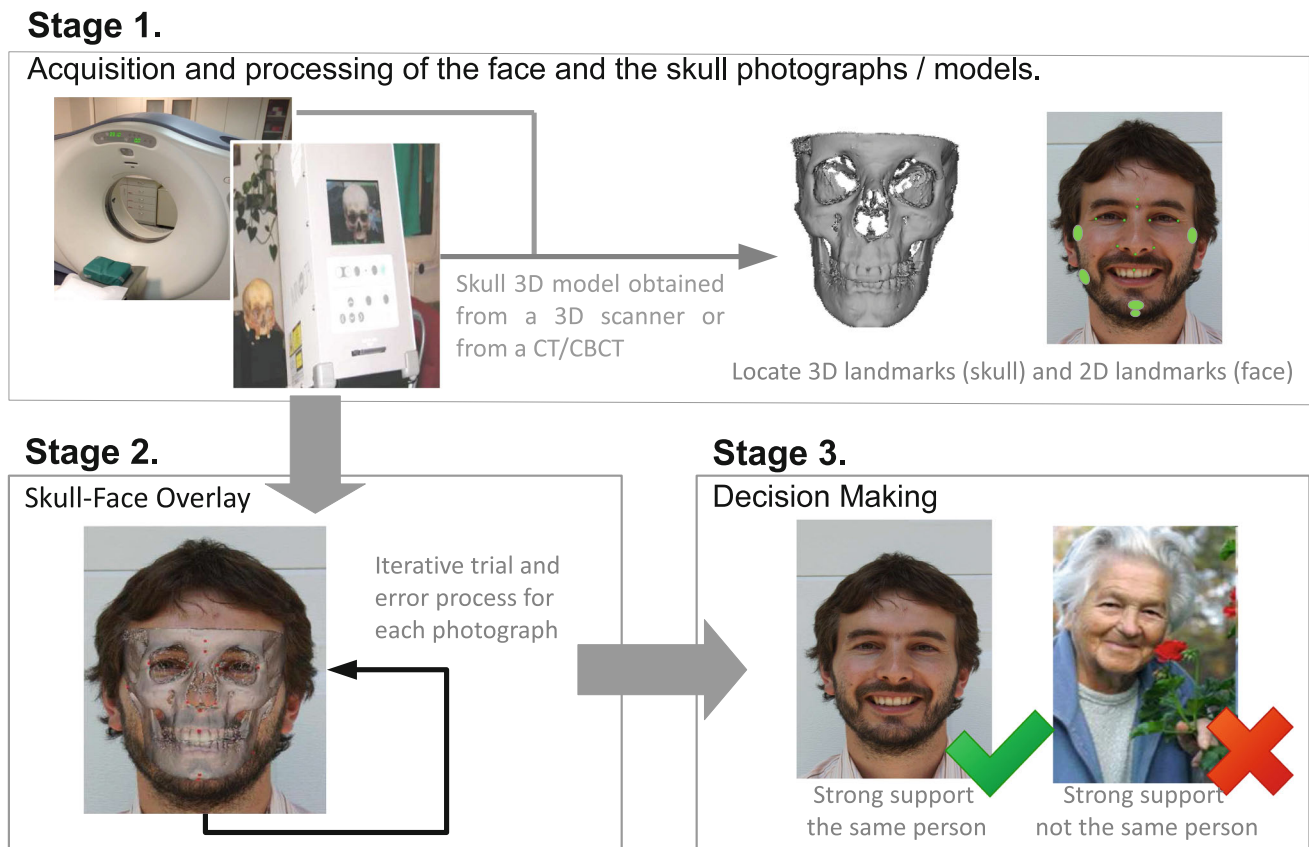
---

✉ Oscar Gómez  
ogomez@decsai.ugr.es

Oscar Ibáñez  
oscar.ibanez@decsai.ugr.es

Oscar Cordon  
ocordon@decsai.ugr.es

<sup>1</sup> Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain



**Fig. 1** Summary of the three craniofacial superimposition stages

objects, a skull and a face, which are different due to facial soft tissue). On the other hand, the resulting search space is huge and presents many local minima, especially when a skull 3D model is considered instead of a skull 2D image. Hence, exhaustive search methods are not useful. Furthermore, forensic experts demand highly robust and accurate results. Image registration approaches based on evolutionary algorithms (EAs) are a promising solution for facing this challenging optimization problem. Thanks to their global optimization nature, EAs are capable of performing a robust search in complex and ill-defined problems as IR [6, 7].

However, one drawback of the existing EA-based proposals dealing with the SFO problem is that they consider all landmarks equally important (the objective function is a mean distance of every pair of corresponding landmarks) when they are not. For instance, landmarks located in the teeth represent the most confident source of information since it is the only bony part visible in the face. Thus, there is a need to properly model the different relative importance of the pairs of landmarks used to perform SFO as a 3D–2D IR approach.

To overcome these problems, we modeled the expert knowledge related to the differences among landmarks based on anthropometric characteristics into the existing automatic

SFO method (second CFS stage). This has been modeled with two classic approaches from the area of multicriteria decision making [8]: weighted sum and lexicographical order. In addition, the obtained results have been compared with the state-of-the-art approach (rcga-mc45) [9] using a “ground truth” dataset [10].

This paper is structured as follows: Sect. 2 reviews the current state of the art and introduces the automatic CFS system. Section 3 describes our proposals for modeling and incorporating expert knowledge within the optimization process. Section 4 presents the experiments and results. The final conclusions are detailed in Sect. 5.

## 2 State of the art

### 2.1 Craniofacial superimposition

The diverse CFS approaches evolved as new technologies became available on foundations laid previously. Although several authors had made different classifications of the technique, all of them recognize three different categories: photographic superimposition (developed in the mid-1930s), video superimposition (widely used since the second half

of the 1970s) and computer-aided superimposition (introduced in the second half of the 1980s) [1, 11]. Another one was proposed by Damas et al. [3] classifying them into two groups: non-automatic computer-aided methods and automatic computer-aided methods. Those later approaches deal with the SFO task within CFS. They obtain unbiased results and drastically reduced the time taken for SFO. Those proposals are based either on photograph-to-photograph comparison [12] or on skull 3D model-to-photograph comparison [9, 13–16].

Most of the existing methods are guided by a set of anthropometric landmarks that can be located in both the skull and the photograph of the missing person. They usually consider the set of landmarks proposed by Martin and Saller [17] which has been used since then for the assessment of correspondence between the skull and the face. The selected landmarks are located in those parts where the thickness of the soft tissue is low. The goal is to facilitate their location when the anthropologist must deal with changes in age, weight and facial expressions. Figures 2 and 3 show the most used cranial and facial landmarks, respectively [18]:

### 2.2 Skull–face overlay as a computer vision problem

Skull–face overlay requires positioning the skull in the same pose as the face of the photograph. From a pure CV point of view, the ante-mortem photograph is the result of the 2D projection of a real (3D) scene that was acquired by a particular (unknown) camera [19]. In such a scene, the living was somewhere inside the camera field of view with a given pose.

The most natural way to face the SFO problem is to replicate the original scenario. To do so, a 3D model of the skull must be considered. The goal is thus to adjust its size and ori-

entation with respect to the head in the photograph through geometric transformations of the skull 3D model in the camera coordinate. The specific characteristics of the camera must also be replicated to reproduce the original as far as possible and hence the perspective projection of the skull 3D model onto the facial photograph. Figure 4 shows an superposition of a case in three different views.

### 2.3 Our automatic skull–face overlay procedure

The 3D–2D IR approach is guided by the cranial and facial landmarks previously assigned by a forensic expert in the skull 3D model and the facial photograph.

Hence, given two sets of cranial and facial landmarks,  $C = \{c^1, \dots, c^m\}$  and  $F = \{f^1, \dots, f^n\}$ , the process has to solve a system of equations with 12 unknowns [14]: the direction of the rotation axis  $\mathbf{d} = (d_x, d_y, d_z)$ , the location of the rotation axis with respect to the center of coordinates  $\mathbf{r} = (r_x, r_y, r_z)$ , the rotation angle  $\theta$ , the factor  $s$  that scales the size of the skull 3D model as the face in the photograph, the translation  $\mathbf{t} = (t_x, t_y, t_z)$  that places the origin of the skull 3D model in front of the camera to replicate the moment of the photograph and the camera angle of view  $\phi$ . Those parameters determine the geometric transformation  $f$  that projects every cranial landmark  $c^i$  of the skull 3D model onto its corresponding facial landmark  $f^i$  of the photograph:

$$F = C \cdot R \cdot S \cdot T \cdot P \tag{1}$$

In addition, it modeled two sources of uncertainty.

Firstly, the location of facial landmarks refers to the difficult task of placing landmarks on a photograph [20]. The definition of many anthropometric landmarks is imprecise in nature [21]. Using precise landmarks, forensic anthropol-

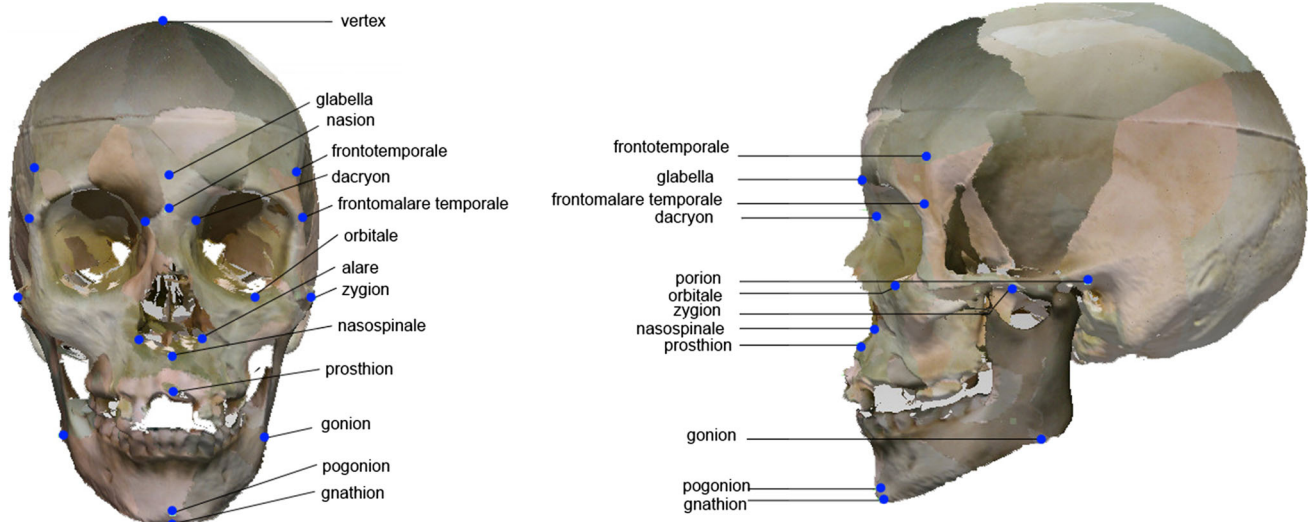
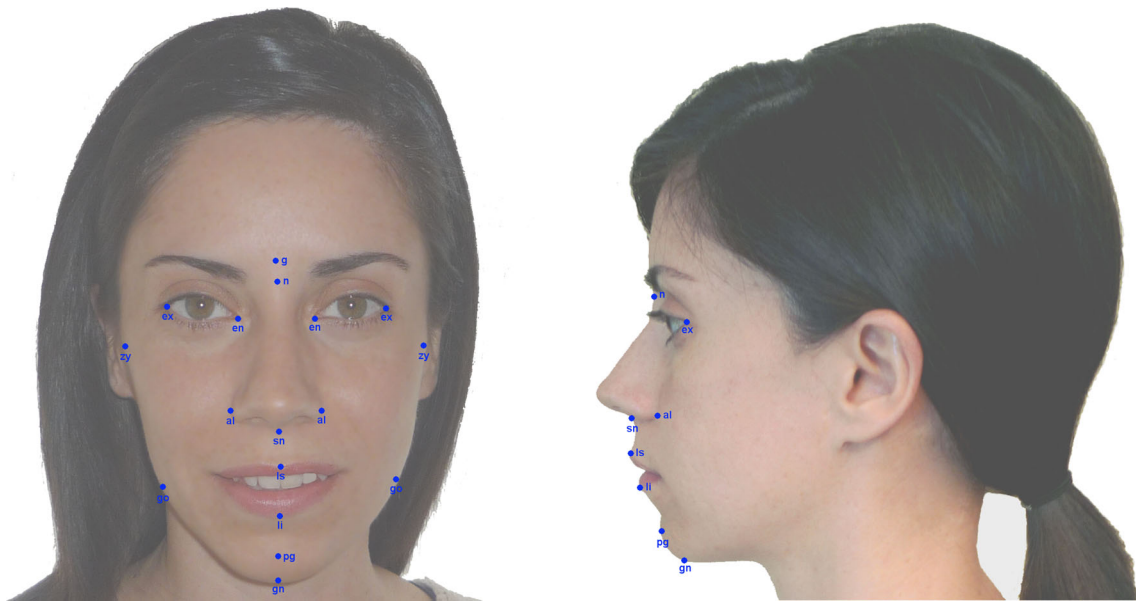


Fig. 2 Cranial landmarks in frontal and lateral view



**Fig. 3** Facial landmarks in frontal and lateral view



**Fig. 4** Skull–face overlay in three different views (frontal, lateral and oblique)

ogists can only place the facial landmarks that they clearly identify in the facial photograph. The fuzzy approach developed in [15] allows experts to enclose a region where the facial landmark is placed without any doubt by using variable size ellipses (fuzzy landmarks) instead of locating a precise point as usual. The number of landmarks placed by the expert can thus increase when those landmarks are considered. This leads to a better description of the skull–face correspondence thanks to the new pairs of cranial points and fuzzy landmarks in the face. The performance of the automatic SFO method is thus improved.

Secondly, facial soft tissue depth varies for each landmark correspondence and for different groups of people. It produces a mismatch among cranial and facial landmarks. Thus, the correspondence of a particular landmark on the

surface of the skull and on the surface of the skin may not be symmetrical and perpendicular. This variability has been widely studied in many populations and considering different age and gender subgroups. The first and unique proposal tackling this uncertainty within an automatic SFO process has been recently published in [9]. This directly incorporates the corresponding landmark spatial relationships and distances within the automatic SFO procedure. To do this, they model the minimum (min), mean (mean) and maximum (max) distances between a pair of cranial and facial landmarks. These distances can be obtained from any anthropometric study looking at the specific population group considered. They used two alternative approaches to deal with the landmark matching imprecision in SFO (using spheres or cones). The main drawback of this approach is the computational com-

plexity, increased with the need of calculating distances between a point (or a fuzzy set) and a fuzzy set [22]. In the first case, distance between a point and a fuzzy set, the same group of authors proposed a cooperative coevolutionary algorithm able to achieve a similar performance with a dramatic execution time reduction [16].

Using the cranial and facial landmarks together with the previous consideration, an EA iteratively searches for the best geometric transformation  $f$ , i.e., the optimal combination of the 12 parameters that minimizes the mean error (ME) fitness function [9]:

$$FME = \frac{\sum_{i=1}^{N_{crisp}} (d'(x_i, f(\tilde{C}^i))) + \sum_{j=1}^{N_{fuzzy}} (d''(\tilde{F}^j, f(\tilde{C}^j)))}{N}, \tag{2}$$

where  $N_{crisp}$  is the number of 2D facial landmarks precisely located (crisp points),  $N_{fuzzy}$  is the number of 2D facial landmarks imprecisely located and defined as bidimensional fuzzy sets,  $N$  is the total number of landmarks considered ( $N = N_{crisp} + N_{fuzzy}$ ),  $x_i$  corresponds to a 2D facial landmark defined as a crisp point ( $x_i \in F$ ),  $\tilde{C}^i$  and  $\tilde{C}^j$  are fuzzy sets modeling each 3D cranial landmark and the soft tissue distance to the corresponding 3D facial landmark  $i$  or  $j$ ,  $f$  is the function that determines the 3D–2D perspective transformation that properly projects every 3D skull point onto the 2D photograph (Eq. 2),  $f(\tilde{C}^i)$  and  $f(\tilde{C}^j)$  are two fuzzy sets, corresponding to the result of applying the perspective transformation  $f$  to the 3D volume (either sphere or cone), which model the landmark matching uncertainty,  $\tilde{F}^j$  represents the fuzzy set of points of the imprecise 2D facial landmark,  $d'(x_i, f(\tilde{C}^i))$  is the distance between a point and a fuzzy set of points, and  $d''(\tilde{F}^j, f(\tilde{C}^j))$  is the distance between two fuzzy sets.

### 3 Modeling anthropometric landmarks relative importance within the automatic SFO process

The aim of this section is twofold: firstly, to introduce the anthropometric characteristics that distinguish groups of landmarks and secondly, to present two approaches to model the relative importance among landmark groups identified in the previous point and their incorporation to the state-of-the-art SFO automatic method.

#### 3.1 Anthropometric differences among landmarks for SFO purposes

The rationale behind differentiating or grouping landmarks could be multiple. However, since not all of the possible combinations can be tested to find the best way of grouping them for every particular scenario, there is a need to choose among

them. In this work, we focused on landmark anthropometric differences in order to give them a different importance as guiding elements within the SFO optimization process. In particular, we modeled the three following scenarios:

##### 3.1.1 Landmark classification I (LC-I): according to their anatomical definition

It has long been recognized that not all landmarks are equally identifiable. This way distinguishes three types of landmarks in [21] named types 1, 2 and 3 according to the decreasing precision of their anatomical location (see Table 1). Type 1 includes landmarks at which three different tissues meet. Type 2 defines points of maximum curvature or other local morphogenetic processes, usually with a biomechanical implication like a muscle attachment site. Finally, type 3 refers to external landmarks, which belong to a curve or surface. In addition, there is a good reason to suspect that the identification precision differs among landmarks [23]. Related to the previous classification, a recent study analyzing the spatial distribution/precision of forensic experts while locating landmarks in facial photographs concluded that there is a significant correlation between the type of landmark and the precision in their location [24].

##### 3.1.2 Landmark classification II (LC-II): according to the rigid or mobile nature of the region

The jaw is the only articulated part on the skull; hence, slightly or even large differences in the articulation of the mandible in the available facial photographs and in the 3D skull model are always expected (see Table 1). In fact, CFS practitioners call this region “terra incognita,” in the sense that they cannot precisely assess craniofacial correspondence in this region due to its mobile nature. Although jaw articulation has been widely studied and mathematically modeled, there is not a single CFS method or practitioner reproducing jaw articulation in ante-mortem images in a reliable and objective manner. Another alternative to address this source of error/uncertainty will be thus to introduce a mathematical modeling of the jaw articulation into the automatic SFO process, so it could be estimated for each particular ante-mortem image. However, even if the latter is successfully performed, there will always be a margin of error justifying the need of considering the landmarks within this region in a more suspicious way.

##### 3.1.3 Landmark classification III (LC-III): according to the presence or absence of soft tissue

Most landmarks do not have a exact match between their position in the skull and in the face due to the facial soft tissue thickness. Contrary to them, a few landmarks (located

**Table 1** Groups of landmarks according to the three classification criteria

Landmark's name	Groups in LC-I	Groups in LC-II	Groups in LC-III	Landmark's name	Groups in LC-I	Groups in LC-II	Groups in LC-III
Vertex	3	1	2	Zygion (left)	3	1	1
Glabella	3	1	2	Zygion (right)	3	1	2
Nasion	1	1	2	Subnasale	1	1	2
Frontotemporale (right)	3	1	2	Prosthion	1	1	2
Frontotemporale (left)	3	1	2	Labiale superius	2	1	2
Endocanthion (left)	1	1	2	Labiale inferius	2	2	2
Endocanthion (right)	1	1	2	Infradentale	1	2	2
Ectocanthion (left)	2	1	2	Gonion (left)	3	2	2
Ectocanthion (right)	2	1	2	Gonion (right)	3	2	2
Tragion (right)	3	1	2	Pogonion	3	2	2
Tragion (left)	3	1	2	Gnathion	3	2	2
Alare (left)	3	1	2	Menton	3	2	2
Alare (right)	3	1	2				

For landmarks location, please refer to Figs. 2 and 3

in the teeth) have a direct matching relation since they are located in a bony region (see Table 1). Thus, it seems quite obvious to consider this group of landmarks as the most representative to study craniofacial anatomical correspondence, something recently corroborated by an experimental study developed with the framework of the European project MEPROCS [25,26]. However, it has not been analytically modeled or tested this higher importance within an automatic SFO procedure, which in any case will need the guidance of other landmarks due to the mostly coplanar region represented by teeth (in many cases, only more frontal teeth are visible in ante-mortem images).

### 3.2 Modeling the differences among landmarks

As a result of distinguishing different groups of landmarks with a different relative importance, we have to move from a single-objective optimization problem to a multiobjective optimization one aimed to reduce the distances between pairs of landmarks from each group independently. In order to incorporate it into the EA optimization process, two classic approaches from the area of multicriteria decision making [8] were followed: weighted sum and lexicographical order.

#### 3.2.1 Weighted sum

In this approach, all landmarks will always contribute to the final fitness; however, not all of them will contribute equally. Depending on the relative importance of a particular group of landmarks and the number of marked landmarks per group in each case (to be able to fairly compare the results of different cases with a different number of marked landmarks per

group). More formally, the fitness of each individual of the genetic algorithm (GA) population will be calculated according to Eq. (3):

$$\text{Fitness} = \frac{\sum_{i=1}^n w_i * \text{fitnessLevel}_i * n\text{Level}_i}{\sum_{i=1}^n n\text{Level}_i} \quad (3)$$

where  $n\text{Level}_i$  is the number of landmarks of the group  $i$  present in a given SFO case; thus,  $\sum_{i=1}^n n\text{Level}_i$  is always equal to the number of landmarks in the case.

Once this proposal has been defined, the last point is to establish the value of the free parameters of this approach, i.e., the number of groups and landmarks included in them, and the weight  $W_i$  of each particular group (their relative importance). While the three different and independent landmarks grouping approaches have an anthropometric motivation, the values for weighting them could be any possible combination adding 1.

#### 3.2.2 Lexicographical order

This approach lexicographically minimizes the fitness of each individual of the GA population. The first group of landmarks is the most important, and it always contributes to the final fitness. However, the information of a following group is only used when two individual are “similar” in all the previous groups. Two individual are considered “similar” when the differences between their fitness is lower than an  $\epsilon$ . However, since the marked landmarks and the distance between them are different in each case, this epsilon has thus to be adaptive to each case, group and generation.

$$\epsilon_i^t = k * |\text{bestFitness}_i^t - \text{worstFitness}_i^t| \quad (4)$$

**Table 2** Experimental design for landmark classifications I, II and III

Modeling approach	Parametrization	Landmark classification	Acronym
Weighted sum	(Group 1: 0.7; Group 2: 0.2; Group 3: 0.1)	Bookstein anatomical definition	G1W1
Weighted sum	(Group 1: 0.6; Group 2: 0.25; Group 3: 0.15)	Bookstein anatomical definition	G1W2
Weighted sum	(Group 1: 0.5; Group 2: 0.3; Group 3: 0.2)	Bookstein anatomical definition	G1W3
Lexicographical order	$K = 0.05$	Bookstein anatomical definition	G1L1
Lexicographical order	$K = 0.1$	Bookstein anatomical definition	G1L2
Lexicographical order	$K = 0.2$	Bookstein anatomical definition	G1L3
Weighted sum	(Group 1: 0.9; Group 2: 0.1)	Mobile or rigid nature	G2W1
Weighted sum	(Group 1: 0.75; Group 2: 0.25)	Mobile or rigid nature	G2W2
Weighted sum	(Group 1: 0.6; Group 2: 0.4)	Mobile or rigid nature	G2W3
Lexicographical order	$K = 0.05$	Mobile or rigid nature	G2L1
Lexicographical order	$K = 0.1$	Mobile or rigid nature	G2L2
Lexicographical order	$K = 0.2$	Mobile or rigid nature	G2L3
Weighted sum	(Group 1: 0.9; Group 2: 0.1)	Bone or soft tissue	G3W1
Weighted sum	(Group 1: 0.75; Group 2: 0.25)	Bone or soft tissue	G3W2
Weighted sum	(Group 1: 0.6; Group 2: 0.4)	Bone or soft tissue	G3W3
Lexicographical order	$K = 0.05$	Bone or soft tissue	G3L1
Lexicographical order	$K = 0.1$	Bone or soft tissue	G3L2
Lexicographical order	$K = 0.2$	Bone or soft tissue	G3L3

where  $i$  is the group,  $it$  is the generation number,  $\epsilon_i^{it}$  is the adaptive  $\epsilon$  of the group  $i$  at generation  $it$ ,  $bestFitness_i^{it}$  is the best value of the fitness at generation  $it$  calculated only using the landmarks of the group  $i$ ,  $worstFitness_i^{it}$  is the worst value of the fitness at generation  $it$  calculated only using the landmarks of the group  $i$ , and  $K$  is a parameter that defines how severe is the epsilon.

The variable  $k$  modulates how easily two individuals are considered “similar.” A high value of  $k$  will produce more ties at each lexicographical level, and thus, the information of the less important group of landmarks will be considered more frequently.

### 4 Experiments

Once the distinct ways to differentiate landmarks and the two proposals to combine the error of the landmarks groups have been introduced, it is necessary to assess the performance of the resulting new optimization methods for SFO. Firstly, the dataset of cases used along all the experiments is presented. Secondly, the experimental design is introduced. Then, the performance of the two models proposed given different parametrizations is analyzed. After assessing the performance, the correlation between the final objective function error and the ground truth value of the corresponding solution is depicted. Finally, a statistical analysis of the performance of our proposals in comparison with the state-of-the-art optimization method (RCGA-MC45) has been performed.

### 4.1 Materials

The experimental design involves 19 SFO problem instances (9 frontal instances, 9 lateral instances and 1 oblique instance) corresponding to nine cases of live people (from Spain and Italy) with ground truth data that will allow an objective evaluation of the results.

These cases were created as follows [10]: The subjects were submitted to cone beam computed tomography (CBCT) for clinical purposes. It generates precise 3D models (0.3-mm slices) in an orthostatic position, and it thus avoids undesired gravitational effects on the soft tissue. The images resulting from the CBCT device were automatically processed to obtain the 3D models of both the skull and the face. After positioning homologous points on the facial 3D model and its corresponding photograph, the former was automatically projected onto the latter using a geometric transformation  $g$  so they perfectly matched, i.e., an actual 3D face–2D face overlay. Then, this geometric transformation  $g$  was applied to project the skull 3D model onto the photograph resulting in a perfect SFO. The latter is considered as the ground truth that can later be compared with the outcome of every SFO method. As a result, we obtained the ground truth data which are the 2D coordinates of the 3D cranial landmarks projected onto the photograph by means of  $g$ .

The skull 3D models and the facial photographs were stored using the Face2Skull™ software, which has been developed by the University of Granada and the European Centre for Soft Computing. This software allows forensic

**Table 3** Mean error in mm regarding the ground truth obtained running 10 times the weighted sum approach for each SFO case (19 in total), the three different landmark classifications (G1, G2 and G3) and three different set of weights (W1, W2 and w3)

Case	Pose	G1W1	G1W2	G1W3	G2W1	G2W2	G2W3	G3W1	G3W2	G3W3
Case 1a	Frontal	<b>2.0982</b>	<b>1.9387</b>	2.1991865	2.5794161	2.32864264	2.391726968	3.0621495	2.663665585	2.789666002
Case 2a	Frontal	<b>2.61412</b>	2.8895	2.8413616	3.17349349	3.17234548	3.092393959			
Case 3a	Frontal	4.84151	4.4755	4.161314	3.59093088	3.45785936	<b>3.347205977</b>			
Case 4a	Frontal	<b>2.96804</b>	3.0485	3.2276504	3.31841135	3.5156861	3.441481626			
Case 5a	Frontal	17.6186	9.8933	7.7452033	6.72594911	<b>6.28129899</b>	6.495759229	6.692036219	6.37502052	6.381191708
Case 6a	Frontal	3.24156	2.9046	2.7214046	<b>2.26981634</b>	2.41906931	2.766641053	2.987625282	2.823931906	2.891385547
Case 6c	Oblique	4.90847	4.7159	4.5277551	4.16608181	4.14160983	3.99195052	3.852904623	<b>3.705103382</b>	3.902782071
Case 7a	Frontal	3.55149	3.6307	3.7759317	<b>2.57529483</b>	2.83140217	3.292978476			
Case 8a	Frontal	<b>3.1226</b>	3.1831	3.3219199	3.63595407	3.54910536	3.523277963			
Case 9a	Frontal	6.39827	5.7742	5.4166415	5.53798268	<b>5.27296718</b>	5.306503895			
Case 1b	Lateral	8.86808	8.3574	8.1172327	7.64152903	<b>7.45735765</b>	7.675532087			
Case 2b	Lateral	11.3443	9.4682	9.3823871	14.736431	10.1834014	<b>8.983237685</b>			
Case 3b	Lateral	11.2753	10.513	10.770529	<b>10.0267103</b>	10.5063759	10.74607162			
Case 4b	Lateral	14.6834	13.59	<b>13.42905</b>	17.2406612	14.2202328	14.37807032			
Case 5b	Lateral	7.13066	6.3248	<b>6.0570856</b>	8.14421048	6.86385057	6.282798484			
Case 6b	Lateral	13.0089	14.804	12.997884	17.6126071	16.5270229	<b>12.20297883</b>			
Case 7b	Lateral	12.5316	12.146	11.374875	12.9127693	<b>11.0654859</b>	11.51613117			
Case 8b	Lateral	9.88603	8.7571	8.2834507	<b>6.98729195</b>	7.32732816	7.597412462			
Case 9b	Lateral	14.0994	12.872	12.339226	11.5777571	<b>11.5362924</b>	11.66952256			



**Table 4** Mean error in mm regarding the ground truth obtained running 10 times the lexicographical order approach for each SFO case (19 in total), the three different landmark classifications (G1, G2 and G3) and three different values for the similarity function parameter  $K$  (L1, L2 and L3)

Case	Pose	G1L1	G1L2	G1L3	G2L1	G2L2	G2L3	G3L1	G3L2	G3L3
Case 1a	Frontal	2.85762	2.7819	2.6809285	2.95551521	2.82659175	2.870259507	<b>2.579079944</b>	2.949873853	2.861587545
Case 2a	Frontal	3.33483	3.4494	3.2139842	2.78225123	<b>2.59360451</b>	2.972023812			
Case 3a	Frontal	4.66705	5.0254	4.7882074	<b>3.56941952</b>	3.86816825	3.580473392			
Case 4a	Frontal	<b>3.06948</b>	3.3243	3.3093015	3.75504176	3.69787192	3.588985181			
Case 5a	Frontal	15.6936	19.064	15.955574	8.83623565	8.90026503	8.867025075	7.021457574	<b>6.59178567</b>	7.337006425
Case 6a	Frontal	3.4234	3.1745	3.4105788	4.77126618	4.36686756	4.588230848	2.802925885	<b>2.729809489</b>	2.826526168
Case 6c	Oblique	4.74832	4.6918	4.6302527	4.32940734	4.48552873	4.373325739	3.756859687	<b>3.675611988</b>	3.690666829
Case 7a	Frontal	<b>3.69166</b>	4.3155	3.9995258	4.84732525	5.01220562	5.123782045			
Case 8a	Frontal	3.50715	3.7705	<b>2.9626292</b>	3.86349789	3.7167015	3.603455627			
Case 9a	Frontal	6.90057	6.7905	7.135047	5.70847276	5.46179388	<b>5.430260545</b>			
Case 1b	Lateral	11.8035	11.345	10.110766	<b>7.71661942</b>	8.59098896	7.942121218			
Case 2b	Lateral	17.174	18.047	17.423522	10.3784426	10.6266424	<b>8.990620152</b>			
Case 3b	Lateral	12.0381	12.438	11.723646	<b>10.6585358</b>	10.7767153	10.80050603			
Case 4b	Lateral	14.489	13.999	15.023362	14.4981552	13.9558517	14.0879031			
Case 5b	Lateral	8.14737	8.8951	7.5316653	6.11762908	6.33814498	<b>6.077587424</b>			
Case 6b	Lateral	15.6918	16.438	15.673819	<b>12.1514556</b>	14.384532	13.24937687			
Case 7b	Lateral	15.4834	14.452	13.04869	12.2013118	11.4521857	<b>10.75906529</b>			
Case 8b	Lateral	11.897	11.621	12.33365	<b>7.7564844</b>	8.01639912	7.837399042			
Case 9b	Lateral	17.7356	14.671	15.121583	11.5688885	12.0095877	<b>11.97019381</b>			

experts to precisely position the cranial landmarks as well as place the facial landmarks on the photographs in a precise and imprecise (using ellipses) manner. Face2Skull<sup>TM</sup> also integrates and runs the proposed automatic SFO algorithms. All the experiments have been performed on an Intel Core<sup>TM</sup> i3 CPU M370 2.40 GHz, with 4GB RAM, running Windows 8.1 Professional<sup>TM</sup>.

## 4.2 Experimental design

A total of 126 different experiments were carried out. These include 3 different parametrizations for both weighted sum and lexicographical order approaches, and the three different landmark classifications for the 19 SFO problems. Within the first and second way of classifying landmarks, the experiments have been performed over the entire dataset (19 SFO cases); however, the third group has been only tested over the 4 cases where the teeth are visible. Since all the approaches tested are based on stochastic processes, 10 independent runs were performed for each problem instance to compare the robustness of the methods and to avoid any possible bias. Table 2 shows a summary of the experiments that have been carried out, along with the configuration of their parameters.

## 4.3 Performance analysis of weighted sum and lexicographical order approaches

The results represent the distance of the obtained SFO to its corresponding ground truth counterpart. We have estimated the total error of each case in mm. By backprojecting the facial points located in the photograph, we can calculate a backprojection line for a given geometric transformation  $f$ . Thus, we apply the inverse,  $f^{-1}$ , of the same transformation if we want to validate and then we choose one point of this line as the 3D position of the 2D point. In particular, we select the point that makes minimum the Euclidean distance between the line and the facial point in the 3D model.

Tables 3 and 4 show the average distance error (in mm) in comparison with the ground truth data of the 10 runs performed using the weighted sum and lexicographical order approaches, respectively. Results are sorted according to the pose of the face of the person within the photograph of each particular case, i.e., we always show first the frontal pose cases and then the lateral pose ones. We choose this particular order according to the performance differences encountered when the pose of the face is considered. Indeed, it can be easily see that the performance of both approaches dramatically decreases facing lateral pose cases.

Table 5 shows the average error distance of our approaches for all the SFO cases in order to allow a more in-depth analysis aimed to determine the best performing approach, considering the parametrization and landmark classification

**Table 5** Mean error in mm regarding the ground truth of all the SFO cases (19 in total) for each particular approach, landmarks classification and parametrization

Experiment	Average	Experiment	Average
G1W1	0.804229963	G1L1	0.909523484
G1W2	0.749546683	G1L2	0.926004211
G1W3	<b>0.72936235</b>	G1L3	<b>0.88167599</b>
G2W1	0.762988683	G2L1	0.786697953
G2W2	0.72261966	G2L2	0.790510879
G2W3	<b>0.71343231</b>	G2L3	<b>0.7776824</b>
G3W1	0.690537259	G3L1	<b>0.64084892</b>
G3W2	<b>0.637742092</b>	G3L2	0.657517491
G3W3	0.66171256	G3L3	0.665916628

employed. Weighted sum performs slightly better when the differences of weights are small (W3). Similarly, lexicographical order performs better when the similarity function is relaxed (larger  $k$  values, L3). This similar behavior is more evident in the first and second group of landmarks, and it does not apply for the third group (bone or soft tissue landmarks) probably because of the limited number of bony landmarks. In fact, both approaches also reach very similar average errors when considering the latter group of landmarks. However, weighted sum performs better in the remaining two cases. The best parametrization for each particular approach and landmark group is marked in bold. G3W2 and G3L1 resulted to be the best performing approaches with similar average distance error. Notice that, as explained before, the third group of landmarks (G3) could be only tested on four cases. Then, within those approaches that could have been applied over the entire dataset, the weighted sum with parametrization W3 is the best approach (G1W3 and G2W3).

## 4.4 Analysis of the correlation between the fitness and the distance to the ground truth

One important issue in every problem that uses evolutionary algorithms is the correlation between the fitness and the goodness of a solution. In this problem and thanks to the existence of a ground truth dataset [10], we can analyze this correlation by measuring the goodness of the obtained solution of the cases of our datasets with their distance to the ground truth. Figures 5 and 6 show a quite good correlation, although not perfect, in both poses for the weighed sum approach, and the lexicographical approach also has the same correlation. The figures show the correlation for the cases 6a (frontal) and 6b (lateral). These two cases depict a similar correlation to the rest.

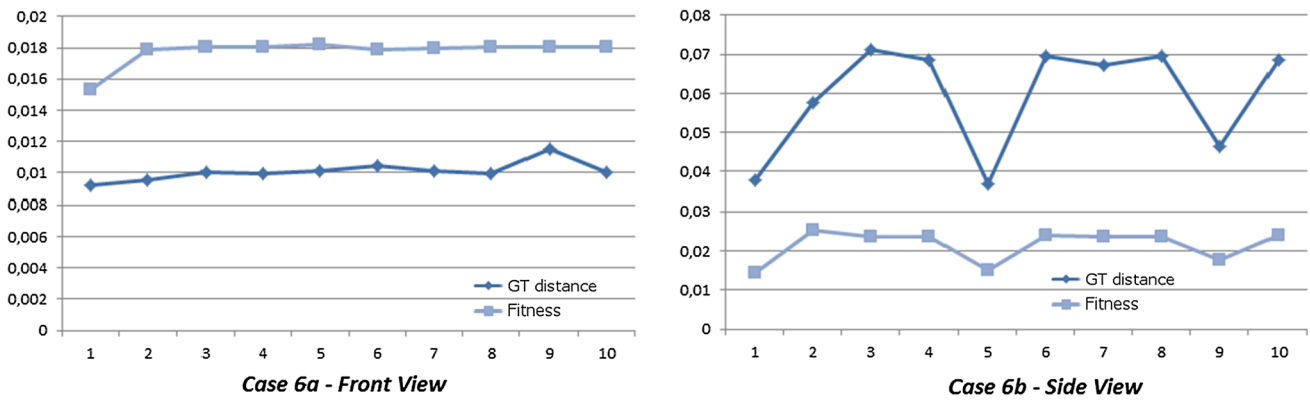


Fig. 5 Correlation between the fitness and the distance to the ground truth. G2W3 for case 6a (left) and case 6b (right)

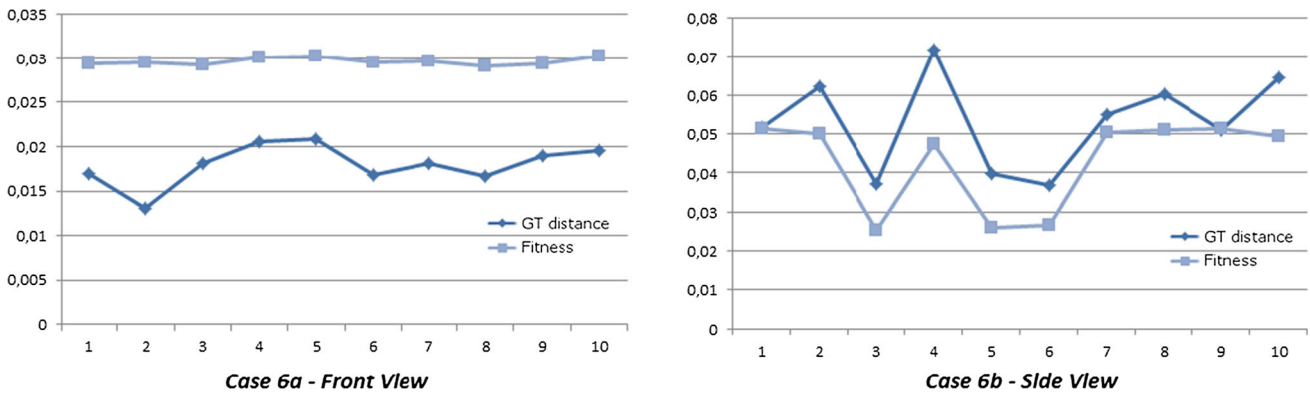


Fig. 6 Correlation between the fitness and the distance to the ground truth. G2L3 for case 6a (left) and case 6b (right)

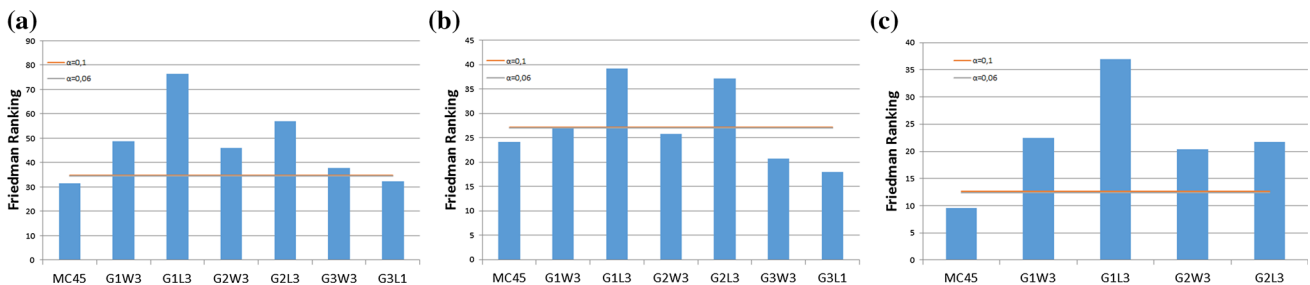


Fig. 7 Friedman and Bonferroni tests. a All cases, b frontal cases, c lateral cases

### 4.5 Statistical analysis and comparison with the state of the art

We performed the same statistical analysis as in [9], i.e., a Friedman test to analyze whether significant differences exist among the performance of our approaches and the state-of-the-art approach, RCGA-c-45 [9], a Bonferroni–Dunn test to detect significant differences among a control approach and the rest, and a paired t test with a Bonferroni and a Holm correction, as well as an unadjusted  $p$  value to know the differences within approaches. We firstly compared using all cases together to see the global performance of both approaches. Afterward, we studied frontal and lateral pose cases inde-

pendently due to their results having shown a significantly different behavior (Fig. 7).

For all cases, there is no statistically significant difference between the MC45 and the two proposals when they model the differences according to the presence or absence of soft tissue (G3W2, G3L1). However, this way of classifying is only formed in our dataset for frontal poses and it could be misleading. With the rest of approaches, the MC45 is significantly better than the obtained results.

It can be easily seen that the performance of both approaches dramatically changes between frontal and lateral cases. For lateral cases, our results are always significantly worse than the MC45. However, frontal cases, G2W3, G3W2

**Table 6** Distribution of the landmarks of the cases into types for the three groups

Case	Pose	Group I: No Type 1	Group I: No Type 2	Group I: No Type 3	Group II: No Type 1	Group II: No Type 2	Group III: No Type 1	Group III: No Type 2
<b>Case 1a</b>	Frontal	3	2	9	10	4	1 (+3 teeth)	13
<b>Case 2a</b>	Frontal	4	2	7	9	4		
<b>Case 3a</b>	Frontal	4	3	8	11	4		
<b>Case 4a</b>	Frontal	4	3	6	10	3		
<b>Case 5a</b>	Frontal	4	3	9	12	4	0 (+2 teeth)	16
<b>Case 6a</b>	Frontal	5	3	7	12	3	1 (+4 teeth)	14
<b>Case 6c</b>	Oblique	4	1	6	8	3	1 (+3 teeth)	10
<b>Case 7a</b>	Frontal	4	3	8	12	3		
<b>Case 8a</b>	Frontal	4	3	8	11	4		
<b>Case 9a</b>	Frontal	4	3	7	11	3		
<b>Case 1b</b>	Lateral	2	1	6	6	3		
<b>Case 2b</b>	Lateral	2	0	3	3	2		
<b>Case 3b</b>	Lateral	2	1	5	5	3		
<b>Case 4b</b>	Lateral	2	1	4	4	3		
<b>Case 5b</b>	Lateral	2	1	6	6	3		
<b>Case 6b</b>	Lateral	2	1	4	4	3		
<b>Case 7b</b>	Lateral	2	1	6	6	3		
<b>Case 8b</b>	Lateral	2	1	5	5	3		
<b>Case 9b</b>	Lateral	2	1	5	5	3		

**Table 7** Normalized mean error (mm) regarding the ground truth obtained running 10 times with G2W3, sorted according to the number of landmarks

Case	Pose	No landmarks	No landmarks group 1	No landmarks group 2	G2W3
Case 5a	Frontal	16	12	4	<b>0.40711535</b>
Case 6a	Frontal	15	12	3	<b>0.60298645</b>
Case 7a	Frontal	15	12	3	<b>0.64268512</b>
Case 3a	Frontal	15	11	4	<b>0.699052</b>
Case 8a	Frontal	15	11	4	<b>0.97774978</b>
Case 9a	Frontal	14	11	3	<b>0.74372375</b>
Case 1a	Frontal	14	10	4	<b>0.75591876</b>
Case 4a	Frontal	13	10	3	<b>0.95890104</b>
Case 2a	Frontal	13	9	4	<b>0.92836805</b>
Case 6c	Frontal	11	8	3	<b>0.86214528</b>
Case 1b	Lateral	9	6	3	<b>0.75914446</b>
Case 5b	Lateral	9	6	3	<b>0.83418451</b>
Case 7b	Lateral	9	6	3	<b>0.88255074</b>
Case 3b	Lateral	8	5	3	<b>0.91661514</b>
Case 8b	Lateral	8	5	3	<b>0.61599062</b>
Case 9b	Lateral	8	5	3	<b>0.77171302</b>
Case 4b	Lateral	7	4	3	<b>0.95704746</b>
Case 6b	Lateral	7	4	3	<b>0.77855812</b>
Case 2b	Lateral	5	3	2	<b>0.51558105</b>

and G3L1 have shown a performance as good as the MC45 and sometimes it is better although no significant differences have been found.

Once it is clear that there is a completely different behavior of the two proposals in frontal and lateral poses, the following step is to study in depth the reason behind it.

Table 6 shows the number of landmarks in every group for the three ways of classifying them under study. In general, frontal cases usually have more landmarks in the first groups than the lateral poses, which used to have around the half.

In particular for the G2W3, Table 7 shows the normalized average distance error (in mm) in comparison with the ground truth data of the 10 runs performed of G2W3 sorted by the number of landmarks. The best performances are mostly located at the top of the table; in other words, they have more landmarks. However, there are a few exceptions (as the case 2b) that despite of having few landmarks have obtained a good performance.

We performed a Pearson test in order to measure the correlation between the number of landmarks and its final performance. This showed that the performance is related to the number of landmarks and also with which landmarks are located. This is because it is crucial to locate a significant number of landmarks, but also these landmarks must be in different planes in order to properly determine the geometric transformation. Thus, the performance of those cases that did not have enough landmarks (as usually happen with lateral cases) in each group was unsatisfactory.

In summary, it is not possible to define a winning approach for all scenarios. For lateral cases, the best approach is the MC45. For frontal cases where the teeth are visible, the best is the G3L1, whereas if these are not visible the best approach is the G2W3.

## 5 Conclusion

This paper addressed the SFO stage and the problem of the relative importance of landmarks based on their anthropometric differences. As a result, we moved from a single-objective optimization problem to a multiobjective one in order to reduce the distances between pairs of landmarks from each group, independently. Therefore, the problem was modeled using two classic approaches (weighted sum and lexicographical order) based on multicriteria decision making in the current SFO stage with three different ways of classifying landmarks.

The weighted sum obtained better results than the lexicographical order in almost all the experiments. At first sight, the performance of both approaches was significantly worse than the state-of-the-art method. However, analyzing the performance of the cases separately depending on the facial pose of the subject in the ante-mortem photographs, they showed

very different behaviors. On the one hand, in lateral pose cases the performance was significantly worse than the state of the art. This poor performance appears to be closely related to the small number of landmarks in the first groups. On the other hand, in frontal pose cases, the performance was better than the state-of-the-art proposal although no significant differences were achieved. In summary, although more future testing seems necessary, promising results were obtained in those cases where the forensic expert located a large number of landmarks, and worse results in those cases with few landmarks.

Promising research lines for future work include the study of other ways of classification of landmarks as well as modeling other relationships between landmarks such as their correlation with face symmetry. Our next study will focus on progressively reducing the uncertainty in fuzzy landmarks. Lastly, another interesting research line is to use the idea of using memetic algorithms in the current proposal as a means of local refinement of the chromosomes.

**Acknowledgements** This research was supported by the Spanish *Ministerio de Economía y Competividad* under NEWSOCO Project TIN2015-67661-P, and the Andalusian Dept. of *Innovación, Ciencia y Empresa* under Project TIC2011-7745, including European Development Regional Funds (EDRF). Mr. O. Gomez's work was supported by Spanish MECD FPU Grant FPU14/02380. Dr. Ibáñez's work has been supported by Spanish MINECO Juan de la Cierva-Incorporación Fellowship JCI- 2014-22433.

## References

1. Yoshino, M.: Craniofacial superimposition. In: Wilkinson, C., Rynn, C. (eds.) *Craniofacial Identification*, pp. 238–253. Cambridge University Press, Cambridge (2012)
2. Campomanes-Álvarez, C., Ibáñez, O., Córdón, O.: Design of criteria to assess craniofacial correspondence in forensic identification based on computer vision and fuzzy integrals. *Appl. Soft Comput.* **46**, 596–612 (2016)
3. Damas, S., Córdón, O., Ibáñez, O., Santamaría, J., Alemán, I., Botella, M.: Forensic identification by computer-aided craniofacial superimposition: a survey. *ACM Comput. Surv.* **43**, 27 (2011)
4. Damas, S., Wilkinson, C., Kahana, T., Veselovskaya, E., Abramov, A., Jankauskas, R., Jayaprakash, P.T., Ruiz, E., Navarro, F., Huete, M.I., Cunha, E., Cavalli, F., Clement, J., Lestón, P., Molinero, F., Briers, T., Viegas, F., Imaizumi, K., Humpire, D., Ibáñez, O.: Study on the performance of different craniofacial superimposition approaches (ii): best practices proposal. *Forensic Sci. Int.* **257**, 504–508 (2015)
5. Zitová, B., Flusser, J.: Image registration methods: a survey. *Image Vis. Comput.* **21**, 977–1000 (2003)
6. Santamaría, J., Córdón, O., Damas, S.: A comparative study of state-of-the-art evolutionary image registration methods for 3D modeling. *Comput. Vis. Image Underst.* **115**, 1340–1354 (2010)
7. Damas, S., Córdón, O., Santamaría, J.: Medical image registration using evolutionary computation: an experimental study. *IEEE Comput. Intell. Mag.* **6**, 26–42 (2011)
8. Chankong, V., Haimes, Y.: *Multiobjective Decision Making Theory and Methodology*. Elsevier Science Ltd, North-Holland (1983)

9. Campomanes-Alvarez, B., Ibanez, O., Campomanes-Alvarez, C., Damas, S., Cordon, O.: Modeling facial soft tissue thickness for automatic skull–face overlay. *IEEE Trans. Inf. Forensics Secur.* **10**(10), 2057–2070 (2015)
10. Ibáñez, O., Cavalli, F., Campomanes-Álvarez, B.R., Campomanes-Álvarez, C., Valsecchi, A., Huete, M.I.: Ground truth data generation for skull-face overlay. *Int. J. Legal Med.* **129**, 569–581 (2015)
11. Huete, M.I., Ibáñez, O., Wilkinson, C., Kahana, T.: Past, present, and future of craniofacial superimposition: literature and international surveys. *Legal Med.* **17**(4), 267–278 (2015)
12. Ghosh, A.K., Sinha, P.: An economised craniofacial identification system. *Forensic Sci. Int.* **117**, 109–119 (2001)
13. Nickerson, B.A., Fitzhorn, P.A., Koch, S.K., Charney, M.: A methodology for near-optimal computational superimposition of two-dimensional digital facial photographs and three-dimensional cranial surface meshes. *J. Forensic Sci.* **36**, 480–500 (1991)
14. Ibáñez, O., Ballerini, L., Cordon, O., Damas, S., Santamaría, J.: An experimental study on the applicability of evolutionary algorithms to craniofacial superimposition in forensic identification. *Inf. Sci.* **79**, 3998–4028 (2009)
15. Ibáñez, O., Cordon, O., Damas, S., Santamaría, J.: Modeling the skull-face overlay uncertainty using fuzzy sets. *IEEE Trans. Fuzzy Syst.* **16**, 946–959 (2011)
16. Ibáñez, O., Cordon, O., Damas, S.: A cooperative coevolutionary approach dealing with the skull–face overlay uncertainty in forensic identification by craniofacial superimposition. *Soft Comput.* **18**, 797–808 (2012)
17. Martin, R.: *Lehrbuch der Anthropologie in systematischer Darstellung mit besonderer Berücksichtigung der anthropologischen Methoden*, vol. 4. G. Fischer (1966)
18. George, R.M.: Anatomical and artistic guidelines for forensic facial reconstruction. In: Iscan, M.Y., Helmer, R. (eds.) *Forensic Analysis of the Skull*, pp. 215–227. Wiley Liss, New York (1993)
19. Faugeras, O.: *Three-Dimensional Computer Vision. A Geometric Viewpoint*, 1st edn. The MIT Press, Cambridge (1993)
20. Cummaudo, M., Guerzoni, M., Marasciuolo, L., Gibelli, D., Cigada, A., Obertová, Z., Ratnayake, M., Poppa, P., Gabriel, P., Ritz-Timme, S., Cattaneo, C.: 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 (2013)
21. Bookstein, F.L.: *Morphometric Tools for Landmark Data: Geometry and Biology*, 1st edn. Cambridge University Press, New York (1991)
22. Campomanes-Álvarez, C., Campomanes-Álvarez, B.R., Ibáñez, O., Cordon, O., Guadarrama, S.: An experimental study on fuzzy distances for skull-face overlay in craniofacial superimposition. *Fuzzy Sets Syst.* **318**, 100–119 (2017)
23. Cramon-Taubadel, N.V., Frazier, B.C., Lahr, M.M.: The problem of assessing landmark error in geometric morphometrics: theory, methods, and modifications. *Am. J. Phys. Anthropol.* **134**(1), 24–35 (2007)
24. Campomanes-Álvarez, B., Ibáñez, O., Navarro, F., Alemán, I., Cordon, O., Damas, S.: Dispersion assessment in the location of facial landmarks on photographs. *Int. J. Legal Med.* **129**(1), 227–236 (2015)
25. Ibáñez, O., Valsecchi, A., Cavalli, F., Huete, M.I., Campomanes-Alvarez, B.R., Campomanes-Alvarez, C., Vicente, R., Navega, D.S., Ross, A., Wilkinson, C., Jankauskas, R., Imaizumi, K., Hardiman, R., Jayaprakash, P.T., Ruiz, E., Molinero, F., Lestón, P., Veselovskaya, E., Abramov, A., Steyn, M., Cardoso, J., Humpire, D., Lusnig, L., Gibelli, D.M., Mazzarelli, D., Gaudio, D., Collini, F., Damas, S.: Study on the criteria for assessing skull-face correspondence in craniofacial superimposition. *Legal Med.* **23**, 59–70 (2016)
26. Ibáñez, O., Vicente, R., Navega, D.S., Campomanes-Alvarez, B.R., Cattaneo, C., Jankauskas, R., Huete, M.I., Navarro, F., Hardiman, R., Ruiz, E., Imaizumi, K., Cavalli, F., Veselovskaya, E., Humpire, D., Cardoso, J., Collini, F., Mazzarelli, D., Gibelli, D., Damas, S.: Meproc framework for craniofacial superimposition: validation study. *Legal Med.* **23**, 99–108 (2016)