An Experimental Study on Fuzzy Distances for Skull-Face Overlay in Craniofacial Superimposition

Carmen Campomanes-Álvarez[a,*], B. Rosario Campomanes-Álvarez[b], Sergio Guadarrama[c], Oscar Ibáñez[a], Oscar Cordón[a,b]

[a]*Department of Computer Science and Artificial Intelligence, University of Granada, C/ Daniel Saucedo Aranda, s/n, 18071 Granada, Granada, Spain.*
[b]*European Centre for Soft Computing, C/ Gonzalo Gutierrez Quirós, s/n, 33600 Mieres, Asturias, Spain.*
[c]*Electrical Engineering and Computer Sciences Department, University of California at Berkeley, 94720 Berkeley, CA, USA.*

Abstract

Skull-face overlay is the most time-consuming and error-prone stage in craniofacial superimposition, an important skeleton-based forensic identification technique. This task focuses on achieving the best possible overlay of an unknown skull found and a single ante-mortem image of a candidate missing person. The process is influenced by some sources of uncertainty since two objects of different nature are involved, i.e. a skull and a face. In previous works we have developed a computer-aided craniofacial superimposition system aimed to assist forensic anthropologists in obtaining the best possible skull and face overlay. The system has successfully allowed us to reduce the processing time, simplify the forensic anthropologists work, and make the process more objective and reproducible. Our approach is based on automatically overlaying a skull three dimensional model onto a facial photograph by minimizing the distance between two subsets of corresponding cranial and facial landmarks. The proposed method properly deals with the inherent uncertainty sources to the skull-face overlay process by considering fuzzy sets to model imprecise landmark location, and imprecise cranial and facial landmarks spatial correspondence (resulting from the presence of soft tissues in the face). Accordingly, our methodology requires computing two kinds of distance metrics: between a point and a fuzzy set, and between two fuzzy sets. This contribution is devoted to study the performance and influence of the most significant and suitable fuzzy distances proposed in the specialized literature, as well as other new ones proposed, on our skull-face overlay system. In particular, we have tested the behavior of our automatic method when considering eight different distance measurements. The system performance has been objectively evaluated considering 18 case studies resulting from a ground truth dataset following a rigorous statistical experimental setup. The fact that the choice of a good distance metric is crucial to our method has been demonstrated since it significantly affects the quality of the final solutions. It has been shown that our skull-face overlay approach presents the best performance using the weighted mean distance in most of the cases and that the results are both more accurate and robust than the other studied metrics.

Keywords: Forensic identification, Craniofacial superimposition, Skull-face overlay, Fuzzy landmarks, Fuzzy distances, Evolutionary algorithms, Genetic fuzzy systems

1. Introduction

Craniofacial superimposition (CFS) [1], one of the approaches in craniofacial identification [2, 3], is a forensic process where a number of ante-mortem images of a missing person are superimposed with the

---

*Corresponding author
Email addresses:* `carmen.campomanes@decsai.ugr.es` (Carmen Campomanes-Álvarez), `rosario.campomanes@softcomputing.es` (B. Rosario Campomanes-Álvarez), `sergio.guadarrama@berkeley.edu` (Sergio Guadarrama), `oscar.ibanez@decsai.ugr.es` (Oscar Ibáñez), `ocordon@decsai.ugr.es` (Oscar Cordón)

skull that is found to determine if they belong to the same subject. The appropriate projection of the skull onto the facial photograph, known as skull face overlay (SFO) [4], is a very challenging and time-consuming part of the CFS technique [5]. There is a strong interest in designing automatic methods to support the forensic anthropologist to put it into effect [6]. In particular, the design of computer-aided CFS methods has experienced a boom over the past twenty years [7]. The most recent approaches consider the use of skull 3D models, as it is the case in the current contribution.

The SFO process is influenced by inherent uncertainty since two objects of different nature are involved (a skull and a face) [8]. Other limitations associated with the different sources of uncertainty in this problem are related with the difficult task to locate landmarks, both in the skull and in the face [9, 10]. Namely, the difficulty to precisely positioning facial landmarks in photographs, specially with a poor quality, and the inability to locate a large set of (noncoplanar) landmarks due to occlusions [8]. Hence, there is a need for an appropriate automatic SFO method able to model this imprecision.

Computer vision (CV) and soft computing (SC) methods can be extremely useful for this aim. Computer vision includes techniques for processing, analyzing, segmenting, and registering image data in an automatic way [11]. Within CV, image registration (IR) aims to find a geometric transformation that overlays two images taken under different conditions (at different times, from different viewpoints, and/or by different sensors) [12]. Soft computing is aimed to design intelligent systems to process uncertain, imprecise and incomplete information [13]. Two of the main SC techniques are fuzzy logic (FL) [14] and evolutionary algorithms (EAs) [15]. The former extends classical logic to provide a conceptual framework for knowledge representation under imprecision and the consequent uncertainty. Specifically, fuzzy sets have largely demonstrated their capability to deal with vagueness and imprecise information. The latter comprises powerful bio-inspired search and optimization tools to automate problem solving in areas such as modeling, simulation, or global optimization [15].

Our previous work tackle SFO in an automatic way using EAs and fuzzy sets [8, 16, 17]. These approaches are based on overlaying a skull 3D model on a facial photograph by minimizing the distance among pairs of landmarks as well as handling the imprecision due to the facial landmarks location [9, 10]. Fuzzy landmarks in photographs are used to jointly deal with the imprecise landmark location and the coplanarity problem [8]. Besides, since the correspondence between facial and cranial landmarks is not always symmetrical and perpendicular [18], cranial landmarks are also modeled by fuzzy sets in our previous approach [19], taking into account the available information concerning soft tissue depths [20, 21].

This methodology thus requires computing distances between pairs of cranial and facial landmarks. The registration transformation leading to the overlay corresponding to the minimum distance should be the best solution to the SFO problem. Since both kinds of landmarks can be represented as fuzzy sets, the need to compute distance measures between fuzzy sets arises. In fact, cranial landmarks are always represented by fuzzy sets in order to account for the soft tissue depths while facial landmarks can sometimes be modeled by crisp points and others by fuzzy landmarks, depending on the landmark location conditions. Hence, two kinds of fuzzy distances are involved in this problem: between a crisp point and a fuzzy set, when facial landmarks are located using a precise point (one pixel on the image); and between two fuzzy sets, when facial landmarks are located using an imprecise region (an ellipse on the image). Regardless the distance nature, the choice of a particular distance measure is expected to have an impact on the performance and robustness of our automatic SFO method.

In the literature, many fuzzy distance measures have been proposed. These distances have been classified depending on the type of information they convey and the application they attend. In this contribution, we aim to study the performance and influence of the most significant, and, *a priori*, most appropriate distance definitions on our SFO method. We will also propose and test a few new metrics. To do so, we have tested our 3D-2D automatic approach using all these distances on 18 skull-face overlay instances from a ground truth dataset [22].

The paper is organized as follows: Section 2 reviews the related work and the problem statement. In Section 3 we introduce and describe the existing distance definitions that we considered appropriate for our problem and we also propose some new definitions. In Section 4 we present the experiments and results obtained. Finally, the discussion and conclusions are detailed in Section 5.

2. Preliminaries

2.1. Skull-Face Overlay in Craniofacial Superimposition

Skull-face overlay is one of the stages of CFS [4]. This process requires positioning the skull in the same pose as the face in the photograph. From a CV point of view, the ante-mortem image is the result of the 2D projection of a real (3D) scene that was acquired by a particular (unknown) camera. In such a scene, the living person was somewhere inside the camera field of view with a given pose [23].

The most natural way to deal with the SFO problem is to replicate this original scenario. To do so, a 3D model of the skull must be employed. Current 3D scanners provide skull 3D models with a precision lower than one millimeter in a few minutes [24]. Once the skull 3D model has been obtained, the goal is to adjust its size and its orientation with respect to the head in the photograph [1]. In addition, the specific characteristics of the camera must also replicated to reproduce the original as much as possible [23].

First, the skull 3D model is positioned in the camera coordinate system through geometric transformations, i.e., translation, rotation and scaling which corresponds to the adjustment of the skull size and its orientation in the same angle as the face in the image [1, 5]. Then, a perspective projection of the skull 3D model is performed onto the facial photograph.

Hence, a 3D-2D IR process where these unknown parameters are estimated seems to be the most appropriate formulation to automate SFO. In fact, this process directly replicates the original scenario in which the photograph was taken [16, 23].

In our automatic SFO procedure, the 3D-2D IR approach is guided by a set of cranial and facial landmarks previously located by a forensic expert in both the skull 3D model and the facial photograph. Once the location of these landmarks is provided by the forensic anthropologist, the SFO procedure is based on searching for the skull orientation leading to the best matching of the two sets of landmarks. We aim to properly align the skull 3D model and the 2D facial photograph in a common coordinate frame system following a 3D-2D IR approach. The required perspective transformation to be applied on the skull was modelled in [16] as a set of geometric operations involving 12 parameters/unknowns which are encoded in a real-coded vector to represent a superimposition solution.

Hence, given two sets of cranial and facial landmarks, $C = \{cl^1, ..., cl^n\}$ and $F = \{fl^1, ..., fl^n\}$, the overlay procedure aims to solve a system of equations with the following 12 unknowns [16]: the direction of the rotation axis $\vec{d} = (d_x, d_y, d_z)$, the location of the rotation axis with respect to the center of coordinates $\vec{r} = (r_x, r_y, r_z)$, the rotation angle θ, the factor s that scales the skull 3D model as the face in the photograph, the translation $\vec{t} = (t_x, t_y, t_z)$ 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 ϕ. These 12 parameters determine the geometric transformation f which projects every cranial landmark cl^i in the skull 3D model onto its corresponding facial landmark fl^i of the photograph as follows:

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

where $\cdot$ is the matrix multiplication.

The rotation matrix R orients the skull in the same pose of the head in the photograph. S, T, and P are scaling, translation and perspective projection matrices, respectively [16]. A complete description of the matrices of Eq. 1 is detailed in [25].

Using the cranial and facial landmarks, an EA iteratively searches for the best geometric transformation f, i.e., the optimal combination of the 12 parameters that minimizes the following mean error (ME) fitness function [16]:

$$ME = \frac{\sum_{i=1}^{N} d(f(cl^i), fl^i)}{N}, \tag{2}$$

where cl^i is the 3D cranial landmark, fl^i is the 2D facial landmark, f is the geometric transformation, $f(cl^i)$ represents the 2D position of the 3D cranial landmark when projected on the photograph, d is the 2D Euclidean distance, and N is the number of landmarks placed by the expert. Figure 1 shows the visual meaning of the fitness function for one case.
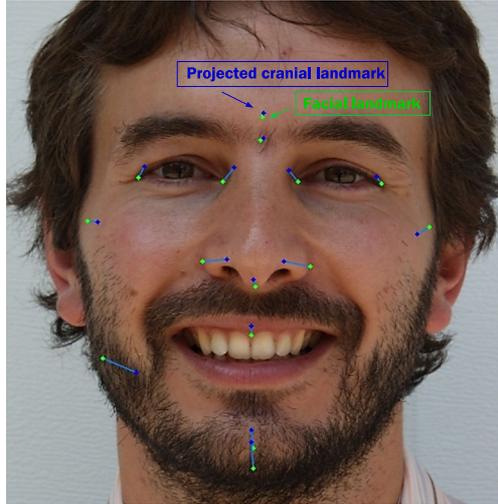
Figure 1: Scenario of the original genetic algorithm to search the best skull-face overlay: geometrical distance between crisp facial landmarks and projected crisp cranial landmarks.

Therefore, the use of geometrical distances in the $\Re^2$ space was the operational environment of our optimization algorithm.

2.2. Modeling the Facial Landmark Location Uncertainty

The landmark location uncertainty is related to the extremely difficult task to precisely and invariably locate the facial landmarks on a photograph. The ambiguity of placing points in a photograph may arise due to the imprecise definition of many anthropometric landmarks [9, 10] but also from additional reasons. These reasons include variation in shade distribution depending on light condition during photographing, unsuitable camera focusing, poor image quality, face pose and expression, partial or whole landmarks occlusion, etc. Therefore, forensic anthropologists are usually only able to unquestionably identify a reduced set of all the available facial landmarks.

We have developed a fuzzy approach to deal with the uncertainty related to the location of facial landmarks [8]. This proposal is based on imprecise landmarks, i.e., the forensic anthropologist can mark the approximate location of any landmarks using an ellipse. This ellipse delimitates a region where the experts can actually assure the anatomical location of the landmark. The size of the ellipse will be directly related with the imprecision in the landmark location: the broader the region, the higher the uncertainty in the location of that landmark. Of course, forensic anthropologists can both define precise and imprecise facial landmarks in a photograph. Those additional landmarks are essential to deal with the coplanarity problem in the automatic search of the best SFO [16].

Following the idea of fuzzy plane geometry in [26] and metric spaces in [27] we have defined a fuzzy landmark as a fuzzy convex set of points having a nonempty core and a bounded support. That is, all its α-levels are nonempty bounded and convex sets. In our case, since we are dealing with 2D photographs with an $x \times y$ resolution, we have defined the fuzzy landmarks as 2D masks represented as a matrix M with $m_x \times m_y$ points (i.e., a discrete fuzzy set of pixels). Each fuzzy landmark will have a different size depending on the imprecision on its localization but at least one pixel (i.e. crisp point related to a matrix cell) will have membership with degree one. These masks are easily built starting from two triangular fuzzy sets $\widetilde{V}$ and $\widetilde{H}$ modeling the approximate vertical and horizontal position of the ellipse representing the location of the landmark, thus becoming 2D fuzzy sets.

An example of these fuzzy facial landmarks is given in Figure 2, where the corresponding membership values (calculated using the product t-norm) of the pixels of one of these landmarks are depicted on the
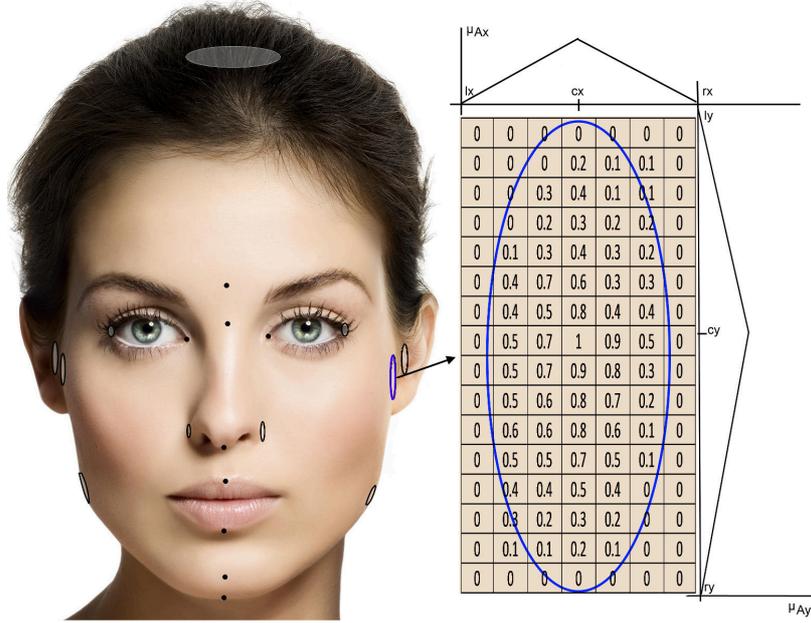
Figure 2: Example of the imprecise location of facial landmarks (left) and representation of an imprecise landmark using fuzzy sets (right).

right. Left and rights bounds of the $\widetilde{V}$ and $\widetilde{H}$ fuzzy sets correspond to the most left/right-upper/lower point of the ellipse, while the modal point corresponds to the center of the ellipse.

2.3. Modeling the Landmark Matching Uncertainty

The landmark matching uncertainty refers to the imprecision that is involved in the matching of two sets of landmarks corresponding to two different objects: a face and a skull. The correspondence between facial and cranial anthropometric landmarks is not always symmetrical and perpendicular; some pairs of landmarks show a very close relationship; meanwhile, others do not exactly overlap because of varying thicknesses of soft tissue between them [18]. Besides, the facial soft tissue depth varies for each facial landmark, as well as for different person groups (based on age, race, and gender). This variability has been studied in many populations considering different age and gender subgroups [20, 21].

The fuzzy sets-based approach developed in [19] models the imprecision related to the facial soft tissue depth between corresponding pairs of cranial and facial landmarks within our automatic SFO procedure. To do so, the population-based statistical minimum (min), mean ($mean$) and maximum (max) distances between a pair of cranial and facial landmarks are represented by fuzzy sets assuming a certain degree of perpendicularity between cranial and facial landmarks as most soft tissue studies do [20, 21].

To do this, we consider the normal vector $\vec{v}$ on the surface of the skull 3D model at each cranial landmark. In order to estimate the position of the facial landmarks, the unit corresponding vector $\vec{u}$ (same direction that normal vector $\vec{v}$ but magnitude of the unit) is multiplied by the specific distance (minimum, mean or maximum). In addition, different inclination angles can be applied to the unit vector $\vec{u}$ in order to define the volume where the facial landmark will be likely located. The landmark matching uncertainty is defined using 3D masks. Hence, a fuzzy set $\tilde{B}_p$ with $p \in \{x, y, z\}$ (see Figure 3 for a graphical representation) is determined by its center $c \in \{c_x, c_y, c_z\}$ (the 3D coordinates of the cranial landmark), the normal vector coordinates $u \in \{u_x, u_y, u_z\}$, and the min, $mean$, and max soft tissue distances.

Therefore, the original definition of our evolutionary SFO technique's fitness function in Eq. (2) was modified in [19] to take into account distances between two fuzzy sets, *Fuzzy Mean Error* (FME), as follows:
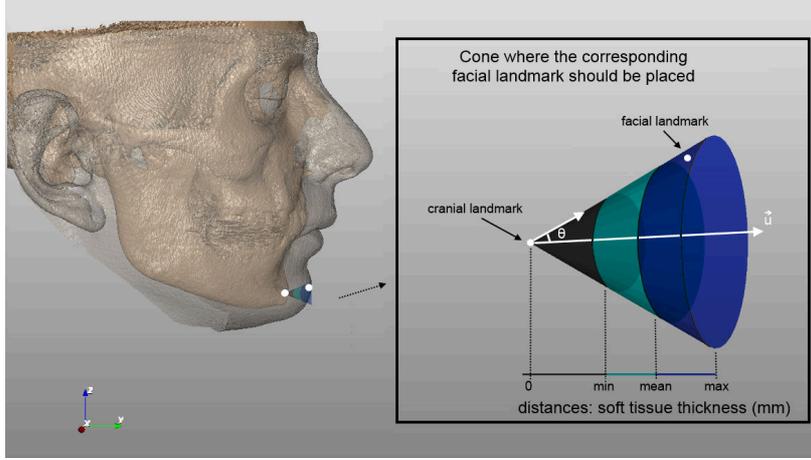
Figure 3: Facial landmark position from a cranial landmark using a cone. *min*, *mean*, and *max* are the soft tissue depths. $\vec{u}$ is the normal vector at the cranial landmark in the skull, θ is the rotation angle of $\vec{u}$.

$$FME = \frac{\sum_{i=1}^{Ncrisp}(d'(x_i, f(\tilde{C}^i)) + \sum_{j=1}^{Nfuzzy}(d''(\tilde{F}^j, f(\tilde{C}^j))}{N}, \qquad (3)$$

where *Ncrisp* is the number of 2D facial landmarks precisely located (crisp points), *Nfuzzy* is the number of 2D facial landmarks imprecisely located and defined as 2D fuzzy sets, N is the total number of landmarks ($N = Ncrisp + Nfuzzy$), 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. 1); $f(\tilde{C}^i)$ and $f(\tilde{C}^j)$ are two fuzzy sets corresponding to the result of applying the perspective transformation f to the 3D volume that 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.

The interested reader is referred to [19] for a detailed description of the procedure.

3. Distances between Points and Fuzzy Sets for Automatic Skull-Face Overlay

The approach introduced in Section 2.3 involves working with a fuzzy set for each cranial landmark. This fuzzy set models the projection of the 3D cranial landmark on the facial photograph, which is composed of the location of the precise 3D cranial landmark (cl^i) and the min, mean and max intervals modeling the soft tissue distance according to a particular population-based study. While cranial landmarks are always identified as a precise point, facial landmarks could be located either precisely or imprecisely, as introduced in Section 2.2. Thus, we can have crisp and fuzzy landmarks at the same time.

Our automatic SFO procedure tries to minimize all the distances between every pair of corresponding landmarks, i.e., cranial landmarks (always represented by fuzzy sets) and facial landmarks (sometimes represented by fuzzy sets and sometimes by crisp points). Accordingly, the need to compute two kind of distance measures arises: between a point and a fuzzy set and between two fuzzy sets.

To illustrate the latter concept, Figure 4 represents the 3D fuzzy sets modeling the spatial correspondence (cones) between cranial (crisp points) and facial landmarks (either crisp or fuzzy landmarks represented by an ellipse) as a consequence of the presence of soft tissues in the face.

6
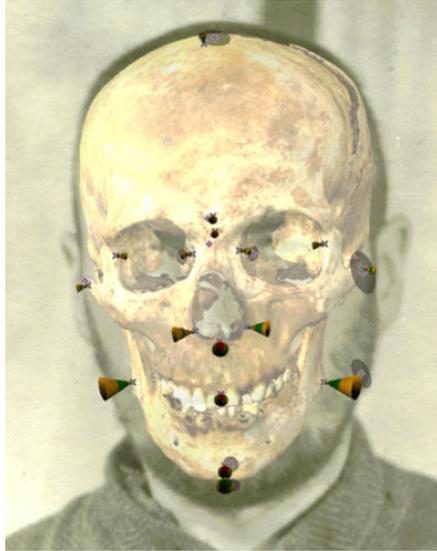
Figure 4: 3D cranial fuzzy landmarks (cones) projected onto a facial photograph with fuzzy landmarks (ellipses).

Once the facial-cranial distances (cones) are projected in the 2D image, we obtain a fuzzy set of 2D points. Now, we need to measure the spatial distance between the corresponding pairs of fuzzy sets. Graphically, we can have the scenario of Figure 5.

The choice of a particular distance is expected to have an influence on the performance and robustness of our automatic SFO method. Therefore, we aim to find the most appropriate distance calculation or a sub-set of them performing better in specific SFO scenarios (for example, a large or a small number of fuzzy landmarks).

For this purpose, we review the existing fuzzy distance measures in the literature and their features. Then, we select the most suitable distances for our approach and we propose some new ones that we consider appropriated for our specific application.

A large variety of distance measures have been proposed from an image processing point of view, depending on the requirements needed for each application field [28, 29]. Applications of such distances cover a very large range, including image registration, assessment of relationships between image components, comparison of imprecise image objects, structural pattern recognition, etc. A review of several definitions of fuzzy distances is presented in [30], including some generalizations and a classification with respect to the type of information they convey. In this work, two kinds of methods are distinguished. On the one hand, distances that compare only the membership functions representing the concerned fuzzy objects, and, on the other hand, distances that combine spatial distance between objects and membership functions. In the first group, no spatial information is taken into account, so the distance measures are more restricted for applications in image processing. The second allows a more general analysis of structures in images, for applications where topological and spatial arrangement of the structures of interest is important. The author proposed several criteria to deal with the problem of choosing a distance, like the type of application, the properties of the distances, and the computation time.

When we face the problem of choosing a distance, we consider the criterion that we need to evaluate distances between objects in the same image. As mentioned, distances that only compare membership functions are not suitable. It is true that in several applications in image processing, all properties of a metric are not needed or are not possible. That is the case of image retrieval and class representation when the type of data representation is complex. For instance, when the data are mapped to the nodes of a weighted graph, this type of data lacks geometrical notions [31]. For this reason, some recent works use measures complex and non-metric measures. However, for this particular study we have only chosen distances between fuzzy sets and between a point and a fuzzy set that always use the Euclidean distance
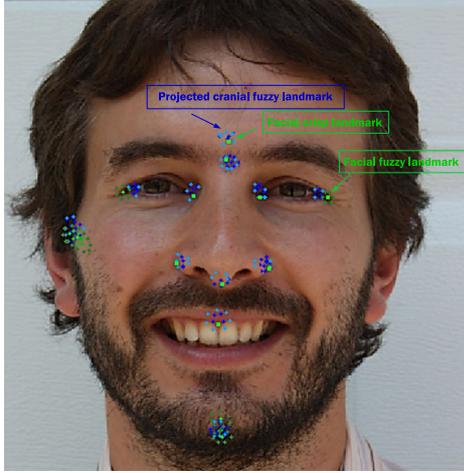
Figure 5: Scenario of the actual genetic algorithm to search the best skull-face overlay: geometrical distance between fuzzy or crisp facial landmarks (green tones) and projected fuzzy cranial landmarks (blue tones).

between two points in the $\Re^2$ space, since this is the main concept of our evolutionary optimization algorithm. We consider essential for this algorithm to be able to establish a fair measure between the pairs of landmarks of the same identification case and also between the final optimization function value (fitness) of the different identification cases. This could give us an idea of how good a skull-face overlay is. This is the reason why the current study only focuses on metric measures based on geometrical notion.

In the following sections, we review some important definitions and notations and we summarize the taxonomy made in [30]. Then, we define the adapted geometrical approach distances for SFO (according to this classification) and we propose some new ones. We distinguish between distance from a point to a fuzzy set and between two fuzzy sets.

3.1. Definitions and Properties of Distances and Fuzzy Distances

Let us denote by S the space on which the image is defined ($\Re^2$ in our case). We denote by x, y, etc., the spatial variables, i.e., points of S (pixels), where $x = (x_1, x_2)$ and $y = (y_1, y_2)$.

The spatial distance between two points x and y of S is denoted by $d_S(x, y)$ (related to the Cartesian space they belong to and independent of their membership to any possible fuzzy set). In this work, $d_S(x, y)$ is taken as the Euclidean distance on S. In $\Re^2$ it is defined by:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \tag{4}$$

A crisp object A is, as usual, a subset of S. Similarly, a fuzzy object $\tilde{A}$ is a fuzzy subset of S and defined bi-univoquely by its membership function. For each x in S, $\tilde{A}(x)$ is a value in $[0, 1]$ that represents the membership degree of the point x to the fuzzy set $\tilde{A}$.

We denote the distance between a point and a set of points as $d(x, A)$. In the same way, the distance between a point and a fuzzy set of points is stated by $d'(x, \tilde{A})$.

The distance between two crisp sets can be expressed as $d(A, B)$. Similarly, the distance between two fuzzy objects $\tilde{A}$ and $\tilde{B}$ is described by $d''(\tilde{A}, \tilde{B})$.

For our purpose we will use the following distances between two crisp sets [32]:

1. **Nearest point distance:**
$$d_{\mathrm{N}}(A, B) = \min_{(x, y)} d(x, y) \tag{5}$$

2. **Symmetrical mean distance:**
$$d_{\mathrm{M}}(A, B) = \frac{\sum_x \min_y d(x, y)}{|A|} + \frac{\sum_y \min_x d(x, y)}{|B|} \tag{6}$$

3. **Hausdorff distance:**

$$d_{\mathrm{H}}(A, B) = \max\{\sup_{x \in A} \inf_{y \in B} d(x, y), \sup_{y \in B} \inf_{x \in A} d(x, y)\} \tag{7}$$

Some definitions of measurements between sets or fuzzy sets do not always satisfy strictly the properties of a distance (or metric). The definition of a metric and its variants are recalled below [27, 30].

Definition 1. *A metric (or distance) is a positive function $d : \mathcal{F} \times \mathcal{F} \to \mathcal{R}^+$ such that:*

1. $\forall \tilde{A} \in \mathcal{F}, d(\tilde{A}, \tilde{A}) = 0$ *(reflexivity)*
2. $\forall (\tilde{A}, \tilde{B}) \in \mathcal{F}^2, d(\tilde{A}, \tilde{B}) = 0 \Rightarrow \tilde{A} = \tilde{B}$ *(separability)*
3. $\forall (\tilde{A}, \tilde{B}) \in \mathcal{F}^2, d(\tilde{A}, \tilde{B}) = d(\tilde{A}, \tilde{B})$ *(symmetry)*
4. $\forall (\tilde{A}, \tilde{B}, \tilde{C}) \in \mathcal{F}^3, d(\tilde{A}, \tilde{C}) \leq d(\tilde{A}, \tilde{B}) + d(\tilde{B}, \tilde{C})$ *(triangular inequality)*

A pseudometric is a function satisfying 1, 3 and 4 (separability does not necessarily hold), and a semi-metric satisfies 1,2, and 3 (and not the triangular inequality), and a semi-pseudometric satisfies only 1 and 3.

According to the particular problem these properties can be strictly necessary (for instance, $d(\tilde{A}, \tilde{A}) = 0$ could needed for the application at hand).

With regard to fuzzy objects, the author of [30] distinguishes two types of distances. On the one hand, distances that basically compare only the membership functions representing the concerned fuzzy objects. On the other hand, distances that combine spatial distance between objects and membership functions.

In particular, as we have mentioned, in our case, the first type of distances are not suitable since spatial information, while being crucial, is not modeled. Because of that, we will only focus on the second one. The distances in this family include the spatial relation d_S in the distance between two fuzzy objects $\tilde{A}$ and $\tilde{B}$. The membership values at different points of S are linked using some formal computation, making introduction of d_S possible. Four main subsets are distinguished within this type of measurements:

- The first subset is the geometrical approach, which consists of generalizing one of the distance metrics between crisp sets. This has been done for many distances following four main principles: considering fuzzy sets in a n-dimensional space [33], using the the fuzzification principle [34], weighting distances by membership values [35], and defining a fuzzy distance as a fuzzy set on $\Re^+$ instead of as a crisp number using the extension principle [35, 36]. The fuzzification principle extends the distance between two crisp sets in its fuzzy equivalent using the corresponding α-cuts set. This is a very common way to proceed and it has already been used for defining several operations on fuzzy sets [37], and especially on fuzzy distances [32, 30].

  For any fuzzy set $\tilde{A} \in [0, 1]^X$, let us denote by $\tilde{A}_\alpha$ the crisp set:

$$\tilde{A}_\alpha(x) = \begin{cases} 1, if \alpha \leq \tilde{A}(x), \\ 0, otherwise, \end{cases} \tag{8}$$

  for all $\alpha \in [0, 1]$, and by $\left[\tilde{A}_\alpha\right]$ the corresponding set $\left\{x \in X; \alpha \leq \tilde{A}(x)\right\}$. These sets are called the α-cuts of $\tilde{A}$ and it is always held that $[\tilde{A}_0] = X$.

- The second subset consists of a morphological approach [32]. This defines fuzzy distances taking into account spatial information based on fuzzy mathematical morphology. They are obtained by direct translation of crisp equations expressing distances in terms of mathematical morphology into fuzzy ones. This approach requires to run more operations than the former group, so it is not appropriate for systems that repeat the distance computation a high number of times, such as EAs. Since our application is based on these algorithms we will not look at these kinds of distances.

- The third approach is based on tolerance [38]. The basic idea is to combine spatial information and membership values by assuming a tolerance value τ, indicating the differences that can occur without saying that the objects are no more similar. These measurements are semi-pseudometrics and they are useful to know the similarity between two fuzzy objects but not for the attended case in the current manuscript.

- The last one is the graph theoretic approach. This is similar to the previous one but objects are not represented directly as fuzzy sets but as fuzzy graphs [39]. This approach is not considered for our purpose for the same reason.

3.2. Distance Between a Point and a Fuzzy Set

The computation of the distance between a crisp point and a classical set can be extended to distances between a crisp point and a fuzzy set in many different ways [30, 40]. The geometrical approach seems to be a priori the most suitable one since the aim is to have a geometric measurement. While the majority of the following distances were directly taken from [30], some metrics are our own contribution based on previously existing distances.

The following distances are defined between a crisp point x and a fuzzy set of points $\tilde{B}$:

1. **Weighted mean distance:** it calculates an average of the distances between x and all the points in $\tilde{B}$, weighted by the membership values [30]:

$$d'_1(x, \tilde{B}) = \frac{\sum_y d(x,y) \cdot \tilde{B}(y)}{\sum_y \tilde{B}(y)} \tag{9}$$

2. **Nearest point distance:** it calculates the distance to the nearest point taking into account the membership degrees [32]:

$$d'_2(x, \tilde{B}) = \min_y \left\{ \frac{d(x,y)}{\tilde{B}(y)} \right\} \tag{10}$$

3. **Nearest point extension 1 distance:** this approach is a new proposal based on the original nearest point distance, but taking into account the membership degrees also inside the fuzzy set:

$$d'_4(x, \tilde{B}) = \min_y \left\{ \frac{d(x,y) + 1}{\tilde{B}(y)} \right\} \tag{11}$$

4. **Nearest point extension 2 distance:** it is another new approach that calculates the distance to the nearest point tending to the fuzzy set point with higher membership degree.

$$d'_5(x, \tilde{B}) = \begin{cases} d'_4(x, \tilde{B}) & \text{if x } \notin \tilde{B} \\ 1 - \tilde{B}(x) & \text{if x } \in \tilde{B} \end{cases} \tag{12}$$

5. **Mean α-cuts distance:** it calculates the distance from the point to each α-cut and weights it by the level of the α-cut [32, 41].

$$d'_3(x, \tilde{B}) = \sum_{i=1}^{m} d_H(x, \tilde{B}_\alpha).(\alpha_i - \alpha_{i-1}) \tag{13}$$

where d_H is the Hausdorff distance between a point and a crisp set:

$$d_H(x, B) = \inf_{y \in B} d(x,y) \tag{14}$$

Note that denominators in Eqs. 9, 10, and 11 cannot be zero. In our case this situation never happens due to the fact that we do not consider part of the fuzzy set the points with membership degrees equal to zero.

3.3. Distance Between Two Fuzzy Sets

Distances between two fuzzy sets are the most widely addressed in the literature. This distance is usually represented by a real number, i.e., taking values in $\mathcal{R}^+$. However, since we work with objects imprecisely defined, the distance between them can also be imprecise. Then, sometimes the distance is better represented as a fuzzy set and more precisely as a fuzzy number [35].

A real fuzzy number $\tilde{n}$ is a fuzzy subset of the real line whose membership function $\mu_{\tilde{n}}$ is [42]:

(i) A continuous mapping from $\Re$ to the closed interval [0,1].
(ii) Constant on $(-\infty,c]$: $\mu_{\tilde{n}} = 0 \; \nabla x \in (-\infty,c]$.
(iii) Strictly increasing on $[c,a]$.
(iv) Constant on $[a,b]$: $\mu_{\tilde{n}}(x) = 1 \; \nabla x \in [a,b]$.
(v) Strictly decreasing on $[b,d]$.
(vi) Constant on $(d,+\infty)$: $\mu_{\tilde{n}}(x) = 0 \; \nabla x \in [d,+\infty)$.

a, b, c and d are real numbers. Eventually we can have $c = -\infty$, or $a = b$, or $c = a$, or $b = d$, or $d = +\infty$. The mean value of $\tilde{n}$ is an element of $[a,b]$, often $(a+b)/2$.

We extend the distances between a point and a fuzzy set introduced in Section 3.2 to model the corresponding distance between fuzzy sets. We denote them using $d''(\tilde{A}, \tilde{B})$ when the result is a crisp value, and $\tilde{D}(\tilde{A}, \tilde{B})(r)$ when it is a fuzzy number.

1. **Mean weighted distance:** it calculates the average of the distances between all the points belonging to both fuzzy sets $\tilde{A}$ and $\tilde{B}$, weighted by the membership values [30, 35].

$$d_1''(\tilde{A}, \tilde{B}) = \frac{\sum_x \sum_y d(x,y) t[\tilde{A}(x), \tilde{B}(y)]}{\sum_x \sum_y t[\tilde{A}(x), \tilde{B}(y)]} \tag{15}$$

   where t is a t-norm.

2. **Nearest point distance:** it calculates the distance to the nearest point taking into account the membership degrees. This approach is proposed as an extension of the nearest point distance between a point and a fuzzy set (10).

$$d_2''(\tilde{A}, \tilde{B}) = \min_{(x,y)} \left\{ \frac{d(x,y)}{\tilde{A}(x) \cdot \tilde{B}(y)} \right\} \tag{16}$$

3. **Nearest point extension 1 distance:** This approach is a new proposal based on the nearest point extension 1 distance between a point and a fuzzy set (11).

$$d_4''(\tilde{A}, \tilde{B}) = \min_{(x,y)} \left\{ \frac{d(x,y) + 1}{\tilde{A}(x) \cdot \tilde{B}(y)} \right\} \tag{17}$$

4. **Nearest point extension 2 distance:** This approach is proposed as an extension of the nearest point extension 2 distance between a point and a fuzzy set (12). It is based on the dissimilarity measure between two fuzzy sets taking into account only the intersection [43, 30].

$$d_5''(\tilde{A}, \tilde{B}) = \begin{cases} d_4''(\tilde{A}, \tilde{B}) & \text{if } \tilde{A} \cap \tilde{B} = \emptyset \\ 1 - \max \left\{ t[\tilde{A}(y), \tilde{B}(y)] \right\} & \text{if } \tilde{A} \cap \tilde{B} \neq \emptyset \end{cases} \tag{18}$$

   where t is a t-norm.

5. **Mean α-cuts distance:** it calculates the distances between crisp sets weighting them by the level of the α-cut [30, 41].

$$d_3''(\tilde{A}, \tilde{B}) = \sum_{i=1}^{m} d(\tilde{A}_\alpha, \tilde{B}_\alpha).(\alpha_i - \alpha_{i-1}) \tag{19}$$

6. **Rosenfeld distance:** Another approach for the distance between two fuzzy sets consists of defining a fuzzy distance between two fuzzy sets as a fuzzy set on $\mathcal{R}^+$ instead of a crisp number. Rosenfeld proposed the shortest distance between two fuzzy sets in [35] using the extension principle:

$$\tilde{D}(\tilde{A}, \tilde{B})(r) = \sup_{\substack{(x,y) \\ d(x,y)=r}} \inf \left[\tilde{A}(x), \tilde{B}(y) \right] \tag{20}$$

There are many other distances that provide a fuzzy number as a result, for example the distances based on the morphological approach [30]. In this work we will only analyze Rosenfeld distance because it is still a geometrical approach.

It was demonstrated in [30] that the choice of the t-norm, among the most common ones (minimun, product and Lukasiewicz), did not imply significant differences, since it changed the absolute values but not the ranking between distances. Since the absolute value is not important in our real-world application, we will use the minimum both for the mean weighted distance (Eq. 15) and for the nearest point extension 2 distance (Eq. 18).

As before, denominators in Eqs. 15, 16, and 17 cannot be zero. In our case this situation never happens due to the fact that we do not consider part of the fuzzy set the points with membership degrees equal to zero.

3.4. Fitness Function Adaptation Using Fuzzy Distances as Fuzzy Numbers in the Skull-Face Overlay Method

The fitness function showed in Eq. 3 is employed for distances 1 to 5. However, in the case of using fuzzy distances as fuzzy numbers (Rosenfeld distance, Eq. 20), the design of the fitness function of our evolutionary-based SFO method changes significantly. In this case, we have fuzzy numbers instead of crisp values (numbers) to represent the distance between each pair of landmarks. This can be understood as if we have an imprecise distance, for example around n, instead of exactly n. It can be proven that, if $\tilde{A}$ and $\tilde{B}$ are fuzzy numbers, and if $\star \in \{+, -, \cdot, \div\}$, then $\tilde{A} \star \tilde{B}$ is also a fuzzy number. So the average of all fuzzy distances is also a fuzzy number.

The common method to operate with fuzzy numbers applies the extension principle [36]:

$$(\tilde{A} \star \tilde{B})(t) = \sup_{t = x \star y} \inf \left[\tilde{A}(x), \tilde{B}(y) \right] \tag{21}$$

Another commonly employed method is based on α-cuts [44]. It defines a fuzzy set in the α-cut, $(\tilde{A} \star \tilde{B})_\alpha$, as

$$(\tilde{A} \star \tilde{B})_\alpha = \tilde{A}_\alpha \star \tilde{B}_\alpha, \tag{22}$$

for any $\alpha \in (0, 1]$, where $\tilde{A}_\alpha$ denotes the α-cut of $\tilde{A}$.

It can be proved that $\tilde{A} \star \tilde{B}$ is:

$$\mu \star \nu = \bigcup_{\alpha \in (0,1]} (\tilde{A} \star \tilde{B})_\alpha, \tag{23}$$

where $(\tilde{A} \star \tilde{B})_\alpha$ is a closed interval for each $\alpha \in (0, 1]$ and $\tilde{A} \star \tilde{B}$ is a fuzzy number.

So, we can apply the two methods to calculate the FME when fuzzy numbers are used as distances. In the cases where crisp points are marked in photographs, the distance is not a fuzzy number but a crisp number. We assume that a crisp number n is a fuzzy number with $\tilde{A}(n) = 1$.

Therefore, we can also obtain the mean distance as a fuzzy number applying any of these two methods. However, only the α-cuts-based method was implemented for our purpose due to its lower computational complexity, an interesting requirement for an iterative process as ours. We call this average distance as *fuzzy mean distance* (FMD), and we denote it by $\tilde{D}_{FMD}$:

$$\tilde{D}_{FMD} = \frac{\bigcup_{\alpha \in (0,1]} \sum\limits_{i=1}^{N} \tilde{D}(\tilde{F}^i, f(\tilde{C}^i))_\alpha}{N} \qquad (24)$$

This distance is a fuzzy number where the α-cuts are the membership values.

A final crisp value (number) for this distance can be calculated by diverse defuzzification methods [45]. Then, the obtained number is applied to the FME fitness function. To do so, we have considered the *center of gravity*. It is a basic general defuzzification approach that computes the center of gravity of the area under the membership function. So, the FME is calculated as follows:

$$FME = \frac{\sum\limits_{\alpha_{min}}^{\alpha_{max}} \alpha \cdot \tilde{D}_{FMD}(\alpha)}{\sum\limits_{\alpha_{min}}^{\alpha_{max}} \tilde{D}_{FMD}(\alpha)} \qquad (25)$$

4. Experiments and Analysis of Results

Some experiments have been developed to analyze the influence of using different distances between fuzzy sets and between a point and a fuzzy set in our automatic SFO system [19].

4.1. Experimental Setup

The experimental design involves 18 SFO problem instances corresponding to nine cases of living people (from Spain and Italy) for each of which two different photographs are available in lateral and frontal poses. Details on how these ground truth data was generated are provided in [22].

The skull 3D models and the facial photographs were stored using the Face2Skull^{TM} software [46], which has been developed by our team. This software allows forensic experts to precisely position the cranial landmarks as well as placing the facial landmarks in the photographs in a precise and imprecise (using ellipses) manner. Face2Skull^{TM} also integrates and runs the proposed automatic SFO algorithm. All the experiments have been performed on an Intel Corel^{TM} 2 Quad CPU Q8400 2.6.

We have run the EA-based method to perform the SFO using the distance definitions given in Section 3. The parameter configuration used is that which obtained the best results in [19]. These proposals calculate the distance between a cranial and its corresponding facial landmark by means of the fitness function defined in Eqs. 3 and 25. Since these methods are based on stochastic processes, 30 independent runs were performed for each problem instance to compare the robustness and to avoid any possible bias.

4.2. Results Obtained

Table 1 presents the mean error and standard deviation achieved by each case and pose (f = frontal view and l = lateral view) in 30 runs for the analyzed approaches as well as the total average error of the algorithm using each distance.

The distance error has been calculated for all the landmarks marked in each case study following the same methodology introduced in [19], i.e., distance in millimeters from the final 2D position of cranial landmarks to the their optimal position given by the ground truth. The mean error of a case study refers to the average of all their single landmark errors. The total average error corresponds to the average of all mean errors obtained by the algorithm using a particular distance. Meanwhile, Table 2 shows the corresponding error to the minimum fitness value by each case and pose. Notice that this value is not the minimum error achieved but the error calculated for the best run (while the error is given by the ground truth the fitness function is different for each distance considered).

The best performance regarding to the ground truth is obtained by using the weighted mean distance in most of the cases. This distance measure is the best performing in 10 out of 18 cases for the mean error of 30 runs (see Table 1), five in frontal view and five in lateral view. With respect to the minimum fitness error (see Table 2), it is the best in eight cases, five in frontal view and three in lateral view. The total average

Case, pose	pl	il	Weighted Mean	Nearest Point	Nearest Point E1	Nearest Point E2	Mean α-cuts (Nearest point)	Mean α-cuts (Sym. mean)	Mean α-cuts (Hausdorff)	Rosenfeld
1,f	7	7	3.17±0.31	3.28±0.65	2.97±0.81	2.07±0.87	2.51±1.00	**1.98±0.89**	3.03±0.53	3.04±0.91
1,l	5	4	**5.83±2.59**	6.68±2.84	8.13±3.25	7.63±3.24	7.73±3.17	7.78±3.14	7.42±3.85	9.39±1.47
2,f	8	5	3.33±0.45	3.18±1.42	3.14±1.88	**2.15±0.42**	2.45±1.12	3.27±1.78	4.24±1.18	3.90±0.36
2,l	3	2	**3.69±0.23**	9.25±2.14	8.24±1.90	8.19±2.76	7.64±2.22	6.52±2.71	6.48±2.62	8.57±2.78
3,f	8	7	**2.94±0.23**	4.81±0.69	4.55±0.56	4.46±0.56	4.35±0.65	4.17±0.67	3.66±0.65	4.64±0.77
3,l	4	4	7.00±0.09	7.20±0.56	7.35±0.81	**6.94±0.39**	7.05±0.51	6.98±0.37	7.54±0.59	9.59±1.50
4,f	7	6	**3.05±0.06**	4.24±0.18	3.90±0.27	3.90±0.34	4.17±0.40	3.67±0.30	4.10±0.13	4.46±0.67
4,l	4	3	12.17±2.83	11.64±2.71	11.95±2.90	11.93±2.94	**10.83±2.51**	11.38±2.66	12.01±2.08	12.65±2.92
5,f	10	6	**2.60±0.13**	5.21±0.65	5.14±0.84	5.34±0.71	5.26±0.66	4.12±0.33	4.42±0.46	7.77±1.66
5,l	5	4	**2.83±0.15**	5.15±2.33	5.05±2.07	5.69±2.14	6.14±2.15	4.49±1.71	6.85±2.60	8.57±2.28
6,f	8	7	**2.92±0.12**	4.98±0.37	3.78±0.74	3.98±0.88	3.70±0.54	3.23±0.53	3.97±0.57	2.93±0.88
6,l	3	4	10.63±3.02	12.64±3.69	11.19±4.10	11.88±3.24	**9.89±2.74**	10.41±3.35	12.05±2.46	11.34±2.38
7,f	10	5	3.71±0.17	3.93±1.64	4.42±2.12	4.12±2.06	3.50±1.77	**3.19±0.81**	5.05±0.64	5.33±1.09
7,l	5	4	10.35±1.97	10.96±1.21	10.48±1.15	10.36±0.88	**9.91±0.93**	10.20±0.89	10.95±0.90	10.85±1.33
8,f	9	6	**2.88±0.24**	4.35±0.49	4.21±0.46	4.39±0.40	4.36±0.21	4.43±0.33	3.56±0.16	5.85±1.35
8,l	4	4	**5.41±0.15**	8.64±1.96	8.19±1.76	9.15±1.55	8.10±1.77	6.87±1.70	7.40±1.85	10.60±1.93
9,f	10	4	4.89±0.15	5.72±0.44	5.79±0.33	5.74±0.36	5.61±0.40	5.19±0.34	**4.54±0.25**	5.25±0.72
9,l	3	5	**9.51±0.69**	11.68±2.84	11.02±2.19	11.22±1.98	11.41±2.06	10.22±1.15	10.53±1.23	13.26±2.31
Average			**5.38**	6.86	6.64	6.62	6.37	6.01	6.54	7.67

Table 1: Mean error in mm and standard deviation regarding the ground truth obtained in 30 runs for each case. f = frontal and l = lateral poses of the face in the photograph, pl = number of precise landmarks, and pi = number of imprecise landmarks located by the experts in each case. In bold number we have highlighted the best result for each case.

Case, pose	pl	il	Weighted Mean	Nearest Point	Nearest Point E1	Nearest Point E2	Mean α-cuts (Nearest point)	Mean α-cuts (Sym. mean)	Mean α-cuts (Hausdorff)	Rosenfeld
1,f	7	7	2.97	4.16	3.22	4.01	3.78	2.73	**2.37**	3.87
1,l	5	4	4.62	**4.33**	4.43	4.85	5.17	5.17	4.51	6.63
2,f	8	5	3.10	2.64	**1.68**	2.04	1.93	2.29	3.49	3.65
2,l	3	2	**3.52**	9.05	8.99	4.89	6.30	3.56	4.46	8.68
3,f	8	7	**2.95**	3.01	4.95	4.68	4.47	4.19	4.02	4.37
3,l	4	4	6.88	6.78	7.04	6.68	7.10	7.02	**6.32**	10.39
4,f	7	6	**3.04**	4.33	4.12	4.40	4.19	3.89	3.99	3.32
4,l	4	3	7.05	6.87	6.53	6.66	6.53	**6.47**	6.79	14.50
5,f	10	6	**2.26**	5.71	5.68	5.58	5.66	4.04	4.17	8.83
5,l	5	4	**2.82**	3.73	4.09	4.05	3.73	3.38	3.80	4.01
6,f	8	7	3.11	5.02	4.14	3.72	3.42	3.18	3.86	**2.07**
6,l	3	4	6.66	**5.86**	6.10	6.22	6.77	6.23	8.70	10.04
7,f	10	5	**3.51**	7.79	7.71	7.44	7.44	4.44	4.89	5.17
7,l	5	4	**7.77**	9.53	9.49	9.12	9.18	9.40	9.64	9.99
8,f	9	6	**2.48**	5.05	4.41	4.37	4.60	3.12	3.59	6.37
8,l	4	4	5.40	7.40	7.09	6.94	5.74	5.71	**5.20**	11.58
9,f	10	4	5.02	5.69	5.47	5.70	5.69	5.51	5.59	**4.81**
9,l	3	5	9.53	**8.29**	9.66	9.37	9.47	9.50	9.95	11.66
Average			**4.60**	5.85	5.82	5.60	5.62	4.99	5.30	7.22

Table 2: Error in mm regarding the ground truth corresponding to the minimum fitness of 30 runs for each case. f = frontal and l = lateral poses of the face in the photograph, pl = number of precise landmarks, and pi = number of imprecise landmarks located by the experts in each case. In bold number we have highlighted the best result for each case.

value is also the lowest with respect to the other distances in both the mean error and minimum error of the fitness, 5.38 mm and 4.60 mm, respectively.

The following three best distances in total average error are the mean α-cuts, using the symmetrical mean distance; the nearest point distance, and the Hausdorff distance between crisp sets. They achieve
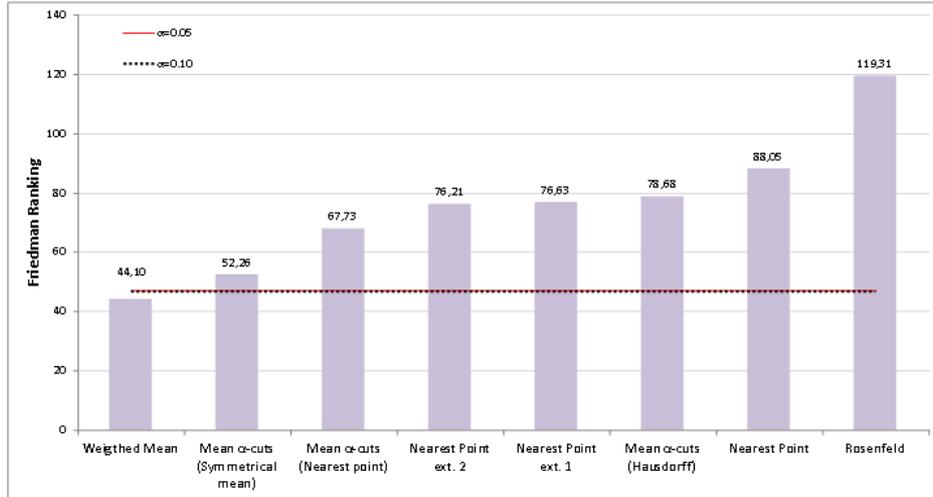
Figure 6: Friedman ranking and Bonferoni-Dunn lines for classification rate for the mean error in 30 runs.

values of 6.01 mm, 6.37 mm, and 6.54 mm respectively. Regarding the minimum fitness error, the mean α-cuts using the symmetrical mean distance achieves the second best results with 4.99 mm of average error.

The worst performance is achieved by the Rosenfeld distance. It provides the higher total average error for the mean error and for the minimum fitness, with 7.67 mm and 7.22 mm respectively. Furthermore, these values are significantly higher than the previous worst, that is the nearest point distance with an error of 6.86 mm in the case of the mean, and 5.85 mm for the best fitness.

As we can see in Table 1, standard deviation values of the weighted mean distance are low. This means that our algorithm has a stable and robust performance using this fuzzy distance. This fact does not occur in the rest of the distances, which show higher values of standard deviation.

4.3. Statistical Tests

We perform a Friedman test [47] to analyze whether significant differences exist among the performance of the different fuzzy distances. The aim is to test a null hypothesis stating that the mean total errors of all the distances are the same. We have set the experiment level of significance at $\alpha = 0.05$. The statistic results of this test are a Friedman χ^2 equal to 49.56, seven degrees of freedom, and a p value of 1.76e-08. This data reveals significant differences among the behavior of the automatic SFO method using the different distances with a p value < 0.0001, thus rejecting the null hypothesis. Due to the rejection of the null hypothesis, a post-hoc statistical analysis is needed. A Bonferroni-Dunn test [48] is carried out to detect significant differences among a control approach and the rest. The use of the weighted mean distance is taken as the control algorithm because it outperforms the remaining distances, i.e., it obtains the lowest value in the Friedman ranking (Figure 6). In the Bonferroni-Dunn test, we have obtained 3.15 and 2.93 as critical values using levels of significance $\alpha = 0.05$ and $\alpha = 0.10$, respectively.

The same test has been performed for the minimum fitness error. The statistic results in this case are a Friedman χ^2 equal to 26.81, seven degrees of freedom, and a p value of 3.59e-04. Again, weighted mean distance is the control algorithm (Figure 7).

Figures 6 and 7 summarize the ranking obtained by the Friedman test for mean error and minimum fitness error respectively. The bar height indicates the average ranking of each alternative. We have drawn a line through the chart whose value is the sum of the smallest bar height (the best approach) and the critical value achieved by the Bonferroni-Dunn test. Bars which are higher than the line are the methods whose performance is significantly worse than the control approach [49].

We have also applied a paired t test with a Bonferroni and a Holm correction, as well as an unadjusted p value in order to learn the differences within approaches [50].

15

Approach	Unadjusted p	p Bonf	p Holm
(Weigthed Mean is the control)			
Nearest Point	**0.00193**	0.05408	**0.04249**
Nearest Point ext. 1	**0.01109**	0.31044	0.23283
Nearest Point ext. 2	**0.01951**	0.54637	0.35124
Mean α-cuts (Nearest Point)	**0.03555**	0.99544	0.56882
Mean α-cuts (Sym. Mean)	0.33533	1.00000	1.00000
Mean α-cuts (Hausdorff)	**0.01539**	0.43081	0.30772
Fuzzy	**<0.00001**	**0.00023**	**0.00022**

Table 3: P values for the comparison between the control and the rest of approaches for the mean error in 30 runs
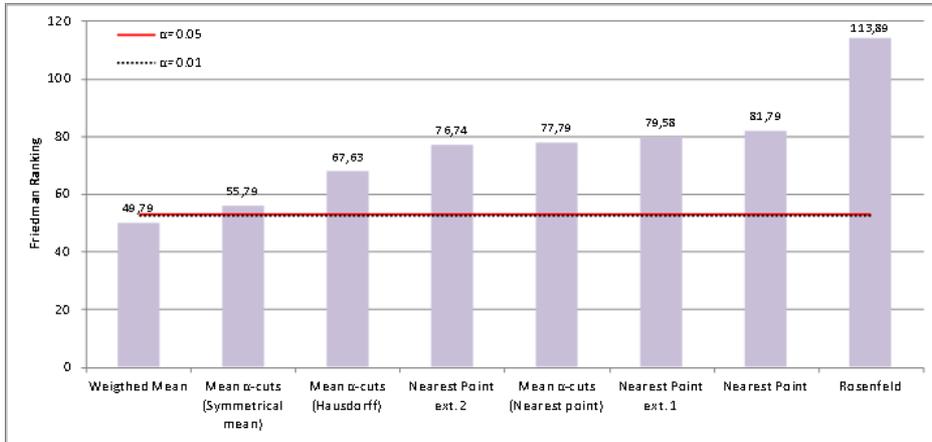
Figure 7: Friedman ranking and Bonferoni-Dunn lines for classification rate for the minimum fitness error in 30 runs.

Table 3 details the pairwise comparisons considering weighted mean distance as the control approach for the mean error. The p value is indicated in each comparison and we have marked in bold the approaches which are worse than the control, considering a level of significance $\alpha = 0.05$.

The weighted mean distance statistically outperforms the rest of the distances with a confidence level of 95 % except for the mean α-cuts using the symmetrical mean distance (Table 3). For the minimum fitness error, this distance is better than the remainder with the same confidence level except for the mean α-cuts using symmetrical mean and Hausdorff distances (Table 4). Those results corroborate the data obtained by the Bonferroni-Dunn test applied to the Friedman ranking where weighted mean distance present a better behavior in our system than the other distances (Figures 6 and 7).

4.4. Visual Analysis

To better illustrate the different distance measures' performance, Figure 8 shows the best superimpositions obtained in four different cases, two in frontal view (corresponding to the fifth and the eighth subjects) and two in lateral view (the second and the fifth) of the best and worst method performance. Blue points (dark grey in the black and white version) refer to the cranial landmarks after overlaying the skull 3D model on the photograph. Yellow points (light grey in the black and white version) are the actual landmarks achieved by the ground truth geometric transformation. Green points (grey in the black and white version) are the facial landmarks marked by the expert in the photograph. We should remind that these facial points have been placed as either precise or imprecise (ellipses) landmarks. Each ellipse contains a grey point inside corresponding to its center.

Analyzing the fifth case in frontal view (Figures 8 (a) and (b)), we can see how the best performance regarding to the minimum fitness is achieved by the mean weighted distance and the worst one by the

Approach	Unadjusted p	p Bonf	p Holm
(Weigthed Mean is the control)			
Nearest Point	**0.01997**	0.55912	0.43931
Nearest Point ext. 1	**0.02788**	0.78053	0.52965
Nearest Point ext. 2	**0.04798**	1.00000	0.81567
Mean α-cuts (Nearest Point)	**0.02005**	0.56136	0.43931
Mean α-cuts (Sym. Mean)	0.45717	1.00000	1.00000
Mean α-cuts (Hausdorff)	0.14124	1.00000	1.00000
Rosenfeld	**<0.00001**	**0.00063**	**0.00063**

Table 4: P values for the comparison between the control and the rest of approaches for the minimum fitness error in 30 runs

Rosenfeld distance, with an error of 2.26 mm and 8.83 mm, respectively. This difference can be easily appreciated in the visuals results. The superimposition using the Rosenfeld distance is clearly wrong as the skull is oriented downward and the teeth and the chin do not match.

The second example in frontal view is shown in Figures 8 (c) and (d) and it corresponds to the eighth case. The best performance is again obtained with the mean weighted distance and the worst with the Rosenfeld one. The errors are 2.48 mm and 6.37 mm, respectively. As we can see in the right image, the skull is too big regarding the face, so the final projected points present a larger distance with respect to their counterpart ground truth points.

Figures 8 (e) and (f) show the best and worst cases for a lateral view of the fifth case. The overlay using the weighted mean distance corresponding to the minimum fitness presents an error equal to 2.82 mm. The worst performance in this case is obtained with the nearest point extension 1, achieving 4.09 mm of error. In the visual results we can appreciate that the worst result is not a correct superimposition since the skull protrudes the face in the nasal region, a fact that is anatomically inconsistent.

Finally, Figures 8 (g) and (h) show the second case in lateral view. Again, the best performance is achieved using the mean weighted distance with an error of 3.52 mm. The highest error (9.05 mm) is provided with the nearest point distance. Here, the result is clearly wrong since the projected skull is located out of the face and with an incorrect orientation.

As we can see, the most accurate results are visually better than the least, and high errors provide bad superimpositions. In general, the use of the weighted mean distance in our SFO system provides the best results.

4.5. Discussion

We have seen that the best performance for our SFO approach is achieved using the weighted mean distance. However, the most appropriate distance for a different application with other features could be another.

In our case, as we explain in Section 2.3, forensic experts draw an ellipse in the photograph when they cannot locate precisely a facial landmark. Our approach models this ellipse as a 2D fuzzy sets of points, whose center has the higher membership value, i.e., one. This fact means that the center of the ellipse is the point where the forensic thinks the facial landmark is placed with more confidence.

Figure 9 shows the performance of the studied distances between a point and a fuzzy set in a simple case of an image (the 2D points are pixels). The euclidean distance is calculated from the point to the center of the fuzzy set (point with membership value of one). When the point is outside the fuzzy set, all the distances tend linearly to the set but nearest point and its extensions, which decrease a little faster.

When the point is inside the set, the nearest point distance is always zero. This distance provides more flexibility to the algorithm since all fuzzy set points are equally processed. Our approach tries to minimize all the distances between every pair of corresponding landmarks. Thus, the use of this distance can cause bad solutions to be considered as good quality ones (i.e. assigned with a low fitness value) due to the larger number of freedom degrees for the location of the cranial landmarks in the photograph. This is a
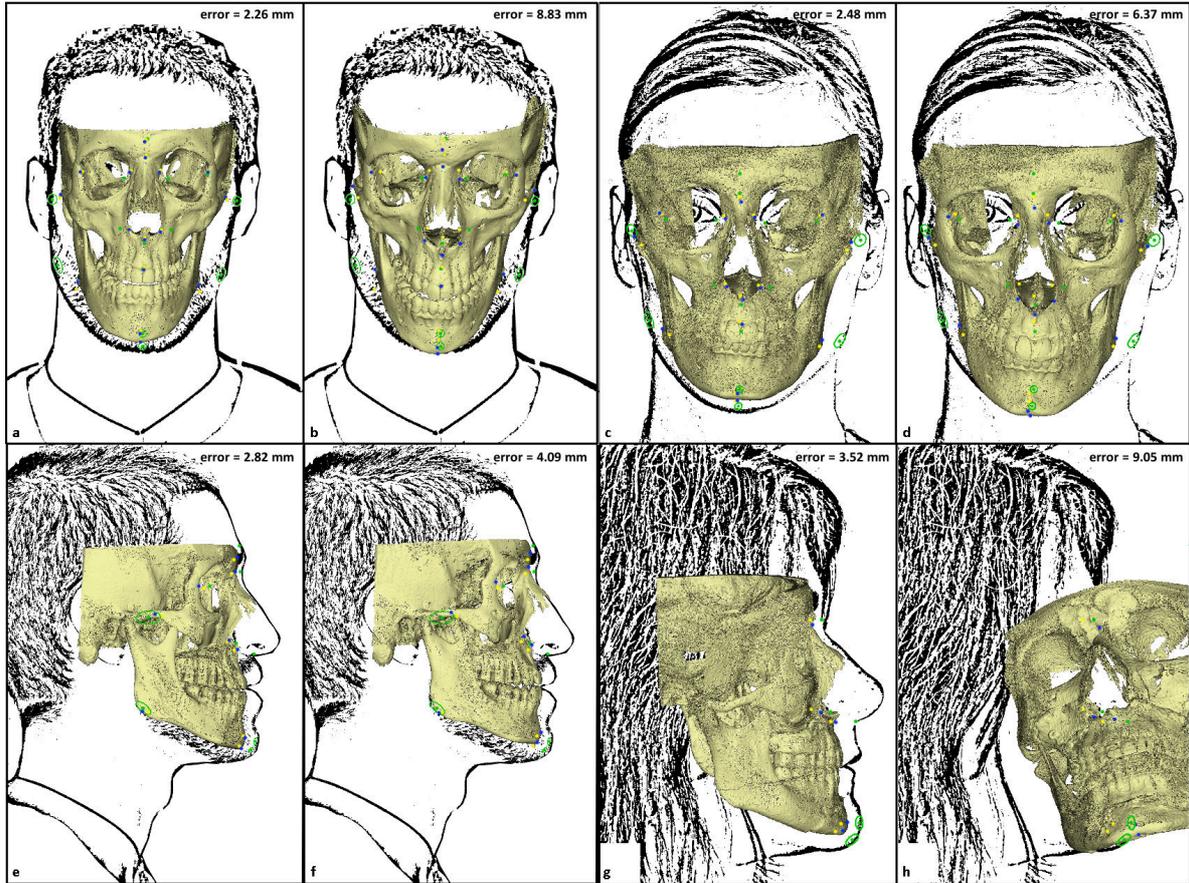
Figure 8: Visual results. Case 5, frontal view: (a) best result corresponds to the weighted mean distance, (b) worst result to the Rosenfeld distance. Case 8, frontal view: (c) best result with the weighted mean distance, (d) worst result with the Rosenfeld distance. Case 5, lateral view: (e) best result using the weighted mean distance, (f) worst result using the nearest point extension 1. Case 2, lateral view: (g) best result corresponds to the weighted mean distance, (h) worst result using the nearest point distance. Notice that the photographs have been processed in order to overcome legal and ethical issues.

consequence of the low discriminative power this distance shows inside the fuzzy set, assigning the same distance value to any of the fuzzy set points. This makes the system of equations to be solved to obtain the registration transformation (see Sec. 2.1) become undetermined, thus having a set of possible solutions that are ranked equally by the fitness function, even if their quality in the SFO problem solving is actually different.

The nearest point extension 2 distance has a similar performance but it is only null in the center of the fuzzy set. Nevertheless, it still offers a lot of freedom inside the set.

The mean α-cuts distance tends to the center of the set (with a null value) but it takes higher values in the extremes. In the case of the nearest point distance extension 1 the performance is uniform around the center (but not null) and it also provides higher values at borders. Those distances consider equal the points around the center.

The weighted mean distance tends to the center but it does not reach a null value. This performance is like the Euclidean distance smoothed. This distance is the most appropriate for our automatic algorithm, because it tends to locate the projected cranial landmarks where the forensic experts think they could be placed.
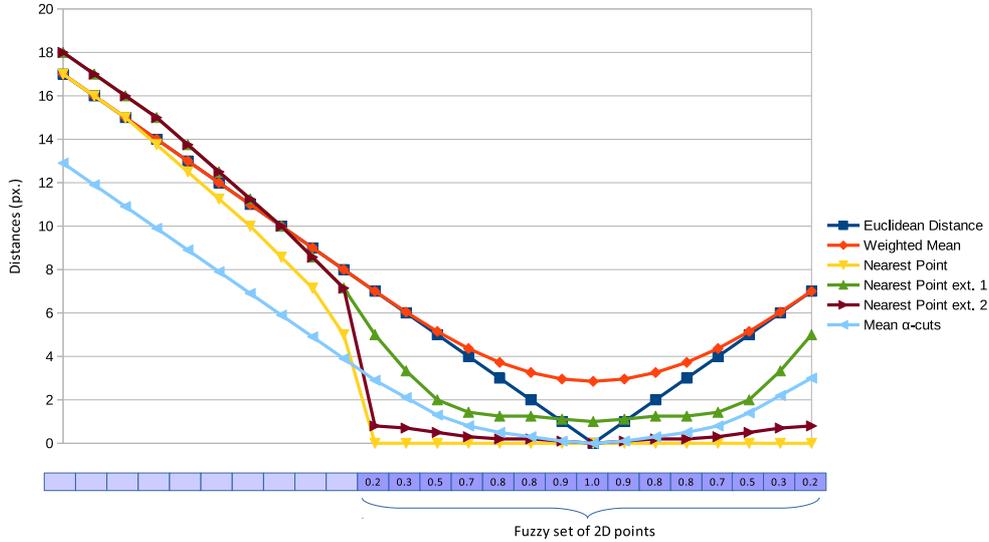
Figure 9: Performance of the different distances between a point and a fuzzy set. This example shows a simple case of an image, so the 2D points are pixels. X axis corresponds to the point from which the distance to the fuzzy set is calculated. Y axis represents the distance in pixels. The Euclidean distance is calculated from the point to the center of the set (point with membership value of one).

5. Conclusions and Future Works

One of the most tedious task in CFS is the SFO process. It requires several hours for overlaying a skull on a facial photograph. The design of automatic and quantifiable methods to perform SFO is a certain need in forensic anthropology [6].

The main goal of our methodology is to assist forensic anthropologists in obtaining the best possible skull and face overlay, reducing the SFO processing time and simplifying their work. Our automatic SFO system uses EAs to search for the best geometric transformation of the 3D skull on the facial photograph minimizing all the distances between every pair of cranial and facial landmarks.

In our previous work we accomplished modeling the two main sources of uncertainty, which are inherent to the SFO problem: the landmark location [8] and the landmark matching [19]. The approach involves computing two kinds of distances: between a point and a fuzzy set and between two fuzzy sets. A large variety of these distance measures have been studied in the literature. We have chosen the most suitable ones for our application and proposed new ones in order to analyze their performance and impact in our method. In particular, we have tested eight distance measurements in 18 case studies and they have been objectively evaluated considering a ground truth dataset [22].

The choice of a good distance metric is crucial to our method since it can significantly affect the quality of the final solutions. Our SFO approach presents the best performance using the weighted mean distance in most of the cases. The results are more accurate and robust than the other studied implementations. In fact, this distance clearly outperforms the remainder regarding to the mean error in 30 runs. It also achieves the minimum fitness error. It is important to note that these studies are necessary for each application in order to find the most suitable distance.

This distance function between pairs of landmarks only look at spatial information related to distance but it does not take into account the directional relative position between them. This aspect can be very interesting for our approach because the relationship between cranial and facial landmarks have to satisfy a specific position criteria depending on the pose. We plan to carry out this extension as a future work.

We also plan to make some developments regarding to the third stage of the CFS process, i.e., the decision making. In this stage, the degree of support that the skull and the available photograph belong

to the same person or not (exclusion) is determined. This decision is guided by different criteria studying the relationship between the skull and the face: the morphological correlation, the matching between the corresponding landmarks according to the soft tissue depth and the consistency between asymmetries. We will aim to study the spatial relationship between some cranial and facial regions influenced by the distance of the soft tissue depth.

Acknowledgments

References

[1] M. Yoshino, Craniofacial superimposition, in: C. Wilkinson, C. Rynn (Eds.), Craniofacial Identification, University Press, Cambridge, 2012, pp. 238–253.

[2] W. A. Aulsebrook, M. Y. Iscan, J. M. Slabbert, P. Beckert, Superimposition and reconstruction in forensic facial identification: a survey, Forensic Sci Int 75 (1995) 101–120.

[3] C. N. Stephan, Craniofacial identification: Techniques of facial approximation and craniofacial superimposition, in: S. Blau, D. Ubelaker (Eds.), Handbook of Forensic Anthropology and Archaeology, volume 25, Left Coast Press, Walnut Creek, 2009, pp. 304–321.

[4] S. Damas, O. Cordón, O. Ibáñez, J. Santamaría, I. Alemán, M. Botella, F. Navarro, Forensic identification by computer-aided craniofacial superimposition: a survey, ACM Computing Surveys 43 (2011) 27.

[5] T. W. Fenton, A. N. Heard, N. J. Sauer, Skull-photo superimposition and border deaths: Identification through exclusion and the failure to exclude, J Forensic Sci 53 (2008) 34–40.

[6] D. H. Ubelaker, A history of smithsonian-FBI collaboration in forensic anthropology, especially in regard to facial imagery, Forensic Sci Communications 2 (2000).

[7] M. I. Huete, O. Ibáñez, C. Wilkinson, T. Kahana, Past, present, and future of craniofacial superimposition: Literature and international surveys, Legal Med 17 (2015) 267–278.

[8] O. Ibáñez, O. Cordón, S. Damas, J. Santamaría, Modeling the skull-face overlay uncertainty using fuzzy sets, IEEE Trans Fuzzy Syst 16 (2011) 946–959.

[9] M. Cummaudo, M. Guerzoni, L. Marasciuolo, D. Gibelli, A. Cigada, Z. Obertová, M. Ratnayake, P. Poppa, P. Gabriel, S. Ritz-Timme, C. Cattaneo, Pitfalls at the root of facial assessment on photographs: a quantitative study of accuracy in positioning facial landmarks, Int J Legal Med 127 (2013) 699–706.

[10] B. R. Campomanes-Álvarez, O. Ibáñez, F. Navarro, I. Alemán, O. Cordón, S. Damas, Dispersion assessment in the location of facial landmarks on photographs, Int J Legal Med 129 (2015) 227–36.

[11] M. Sonka, V. Hlavac, R. Boyle, Image processing, analysis, and machine vision, Cengage Learning, 2014.

[12] B. Zitova, J. Flusser, Image registration methods: a survey, Image Vision Comput 21 (2003) 977–1000.

[13] P. P. Bonissone, Soft computing: the convergence of emerging reasoning technologies, Soft Comput 1 (1997) 6–18.

[14] L. A. Zadeh, Soft computing and fuzzy logic, IEEE Softw 11 (1994) 48–56.

[15] A. E. Eiben, J. E. Smith, Introduction to Evolutionary Computing, SpringerVerlag, Heidelberg, 2003.

[16] O. Ibáñez, O. Cordón, S. Damas, J. Santamaría, An experimental study on the applicability of evolutionary algorithms to craniofacial superimposition in forensic identification, Inf Sci 79 (2009) 3998–4028.

[17] O. Ibáñez, O. Cordón, S. Damas, A cooperative coevolutionary approach dealing with the skull-face overlay uncertainty in forensic identification by craniofacial superimposition, Soft Comput 18 (2012) 797–808.

[18] G. M. Gordon, M. Steyn, An investigation into the accuracy and reliability of skull-photo superimposition in a south african sample, Forensic Sci Int 216 (2012) 198–e1.

[19] B. R. Campomanes-Álvarez, O. Ibáñez, C. Campomanes-Álvarez, S. Damas, O. Cordón, Modeling the facial soft tissue thickness for automatic skull-face overlay, IEEE T Inf Foren Sec. 10 (2015) 2057–2070.

[20] C. N. Stephan, E. K. Simpson, Facial soft tissue depths in craniofacial identification (part i): an analytical review of the published adult data, J Forensic Sci 53 (2008) 1257–1272.

[21] C. N. Stephan, E. K. Simpson, Facial soft tissue depths in craniofacial identification (part ii): an analytical review of the published sub-adult data, J Forensic Sci 53 (2008) 1273–1279.

[22] O. Ibáñez, F. Cavalli, B. R. Campomanes-Álvarez, C. Campomanes-Álvarez, A. Valsecchi, M. I. Huete, Ground truth data generation for skull–face overlay, Int J Legal Med 129 (2015) 569–81.

20

[23] B. R. Campomanes-Álvarez, O. Ibáñez, F. Navarro, M. Botella, S. Damas, O. Cordón, Computer vision and soft computing for automatic skull–face overlay in craniofacial superimposition, Forensic Sci Int 245 (2014) 77–86.

[24] H. K. Park, J. W. Chung, H. S. Kho, Use of hand-held laser scanning in the assessment of craniometry, Forensic Sci Int 160 (2006) 200–206.

[25] D. Hearn, M. P. Baker, Computer graphics. C version, Prentice-Hall, Upper Saddle River, 1997.

[26] J. J. Buckley, E. Eslami, Fuzzy plane geometry i: Points and lines, Fuzzy Set Syst 86 (1997) 179–187.

[27] P. Diamond, P. Kloeden, Metric topology of fuzzy numbers and fuzzy analysis, in: Fundamentals of Fuzzy Sets, Springer, 2000, pp. 583–641.

[28] J. Freeman, The modelling of spatial relations, Comput Vision Graph 4 (1975) 156–171.

[29] S. Dutta, Approximate spatial reasoning: integrating qualitative and quantitative constraints, Int J Approx Reason 5 (1991) 307–330.

[30] I. Bloch, On fuzzy distances and their use in image processing under imprecision, Pattern Recognit 32 (1999) 1873–1895.

[31] D. W. Jacobs, D. Weinshall, Y. Gdalyahu, Classification with nonmetric distances: Image retrieval and class representation, IEEE T Pattern Anal 22 (2000) 583–600.

[32] I. Bloch, Distances in fuzzy sets for image processing derived from fuzzy mathematical morphology, in: Information Processing and Management of Uncertainty in Knowledge-Based Systems (1996), volume 3, pp. 1307–1312.

[33] J. R. Goetschel, W. Voxman, Topological properties of fuzzy numbers, Fuzzy Set Syst 10 (1983) 87–99.

[34] I. Bloch, Spatial reasoning under imprecision using fuzzy set theory, formal logics and mathematical morphology, Int J Approx Reason 41 (2006) 77–95.

[35] A. Rosenfeld, Distances between fuzzy sets, Pattern Recogn Lett 3 (1985) 229–233.

[36] L. A. Zadeh, The concept of a linguistic variable and its application to approximate reasoningi, Inf Sci 8 (1975) 199–249.

[37] D. Dubois, M.-C. Jaulent, A general approach to parameter evaluation in fuzzy digital pictures, Pattern Recogn Lett 6 (1987) 251–259.

[38] R. Lowen, W. Peeters, Distances between fuzzy sets representing grey level images, Fuzzy Set Syst 99 (1998) 135–149.

[39] M. Gary, J. C. Poon, A fuzzy-attributed graph approach to handwritten character recognition, in: Fuzzy Systems, Second IEEE International Conference on (1993), IEEE, pp. 570–575.

[40] D. Dubois, H. Prade, On distances between fuzzy points and their use for plausible reasoning, in: Int Conf Systems, Man, and Cybernetics (1983), pp. 300–303.

[41] G. Gerla, R. Volpe, The definition of distance and diameter in fuzzy set theory, Stutia Univ Babes–Bolyai Math 31 (1986) 21–26.

[42] D. Dubois, H. Prade, Operations on fuzzy numbers, Int J Syst Sci 9 (1978) 613–626.

[43] R. Zwick, E. Carlstein, D. V. Budescu, Measures of similarity among fuzzy concepts: A comparative analysis, Int J Approx Reason 1 (1987) 221–242.

[44] G. Klir, B. Yuan, Fuzzy sets and fuzzy logic, volume 4, Prentice Hall, New Jersey, USA, 1995.

[45] W. V. Leekwijck, E. E. Kerre, Defuzzification: criteria and classification, Fuzzy Set Syst 108 (1999) 159–178.

[46] Face2skull^{TM} software, http://www.face2skull.com/, 2014. [Online; accessed 15-Jul-2014].

[47] M. Friedman, A comparison of alternative tests of significance for the problem of m rankings, Ann Math Stat 11 (1940) 86–92.

[48] O. J. Dunn, Multiple comparisons among means, J Am Stat Assoc 56 (1961) 52–64.

[49] S. García, A. Fernández, J. Luengo, F. Herrera, A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability, Soft Comput 13 (2009) 959–977.

[50] S. P. Wright, Adjusted p-values for simultaneous inference, Biometrics (1992) 1005–1013.