Universidad de Granada

Departamento de Ciencias de la Computación e Inteligencia Artificial

Programa de Doctorado en Tecnologías de la Información y la Comunicación

Automation of the assessment of craniofacial superimposition using soft computing and computer vision

Tesis Doctoral

Carmen Campomanes Álvarez

Granada, Mayo de 2017

Gf kxqt<'Wpkxgtukf cf "f g"I tcpcf c0'Vguku'F qevqtcngu

Cwqtc<'Ecto gp'Eco r qo cpgu"ƒ rxctg|

KUDP <'; 9: /: 6/; 385/562/4

WT K²'j wr <1lj f rfJ cpf rg0pgvl326: 3169792"

Universidad de Granada

Automation of the assessment of craniofacial superimposition using soft computing and computer vision

Memoria que presenta

Carmen Campomanes Álvarez

Para optar al grado de Doctor

Mayo de 2017

Directores

Oscar Cordón García
Oscar Ibáñez Panizo
Caroline Wilkinson

Departamento de Ciencias de la Computación e Inteligencia Artificial

Agradecimientos

En primer lugar, agradecer a mis directores de la Universidad de Granada, Óscar Cordón y Óscar Ibáñez, con los cuales ha sido un verdadero placer trabajar. Muchas gracias por la inestimable ayuda, la continua disponibilidad, la cercanía y, por supuesto, el buen trato, que en estos cuatro años ha sido siempre impecable. También a mi directora del área forense, Caroline Wilkinson, gracias a la cual he aprendido muchas cosas totalmente diferentes y nuevas para mí, que nunca me hubiese imaginado antes de empezar en este proyecto. Soy consciente de la suerte que he tenido por haberlos tenido a los tres como directores, y del honor que ello supone. Para mí siempre serán un ejemplo a seguir de profesionalidad en la investigación.

Mi más sincero agradecimiento también a todos los que han sido mis compañeros durante este tiempo, tanto de la Universidad de Granada, especialmente del CITIC y del departamento, como a los del European Center for Soft Computing de Asturias. Ha sido un placer haber compartido el día a día junto a todos ellos.

También tengo que agradecer a todo el grupo de trabajo de Caroline, que durante las dos estancias que hice con ellos, en la Universidad de Dundee y en la de Liverpool John Moores, me hicieron sentir como una más y me ayudaron en todo lo posible.

A mis padres que siempre me han apoyado, y por supuesto, a mi hermana Charo, que también me ha ayudado mucho a nivel profesional y que siempre va delante haciéndome el camino más fácil. Por último, a mis amigos y a toda la gente que he conocido durante estos años y que ha aportado su granito de arena de alguna forma u otra.

Finalmente, agradecer a la Universidad de Granada y al Ministerio de Educación, Cultura y Deportes que bajo la concesión de la beca para la Formación de Personal Universitario (referencia AP2012-4285) han financiado y hecho posible estos cuatro años de trabajo.

Muchas gracias.

El doctorando / The *doctoral candidate* **[Carmen Campomanes Álvarez]** y los directores de la tesis / and the thesis supervisor/s: **[Oscar Cordón García, Óscar Ibáñez Panizo y Caroline Wilkinson]**

Garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.
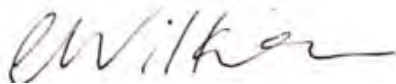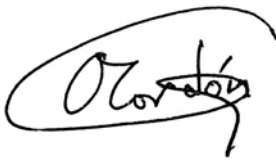
/

Guarantee, by signing this doctoral thesis, that the work has been done by the doctoral candidate under the direction of the thesis supervisor/s and, as far as our knowledge reaches, in the performance of the work, the rights of other authors to be cited (when their results or publications have been used) have been respected.

Lugar y fecha / Place and date:

Granada, 25 de mayo de 2017

Director/es de la Tesis / *Thesis supervisor/s;*   Doctorando / *Doctoral candidate*:

Firma / Signed   Firma / Signed

Contents

Resumen

Dentro de la identificación forense, una de las principales técnicas basadas en el estudio de los restos óseos es la superposición cráneo-facial. En ella se superpone un cráneo encontrado, del que se desconoce la identidad, sobre una o más fotografías de personas desaparecidas y se comparan analizando sus correspondencias morfológicas. Este método de identificación ofrece un gran potencial de aplicación debido a que en la actualidad la gran mayoría de gente dispone de fotografías que muestran con claridad sus caras (material *ante-mortem*). Además, el cráneo (material *post-mortem*) es un hueso que se conserva en perfecto estado con el paso de los años y no se degrada fácilmente bajo los efectos de elementos como el fuego, la humedad, las temperaturas extremas, etc. En este proceso se distinguen tres etapas [DCI+11]: 1) La adquisición y el procesado de los materiales, el cráneo (que en las técnicas más novedosas se escanea obteniendo un modelo 3D) y las fotografías faciales *ante-morten* de los posibles candidatos; 2) El proceso de solapamiento cráneo-cara, consistente en obtener la mejor superposición posible entre el cráneo y la cara de una fotografía concreta; 3) La toma de decisiones, en la cual se estudia la correspondencia morfológica y anatómica cráneo-cara con el objetivo de determinar si ámbos pertenecen a la misma persona o no.

El *soft computing* y la visión por computador presentan ciertas características que los convierten en potentes herramientas para automatizar la superposición cráneo-facial, reduciendo drásticamente el tiempo que emplearía un experto forense en aplicar la técnica de forma manual. En concreto, en [SCD07a, SCD+07b, SCD+09, CÁCD13] se presentan técnicas evolutivas de reconstrucción 3D muy últiles para la primera etapa del proceso. Además, el sistema automático propuesto en [ICDS09, ICDS11, ICD12, CAIN+14] lleva a cabo el solapamiento cráneo-cara mediante la proyección del modelo 3D sobre la imagen 2D utilizando algoritmos evolutivos y conjuntos difusos para lograr la correspondencia directa entre *landmarks* faciales y craneales, modelando también la localización imprecisa de *landmarks* faciales en la fotografía. Si bien los logros obtenidos modelando las dos primeras etapas del proceso son prometedores, la tercera etapa sigue realizándose de forma manual por el antropólogo forense, quien toma la decisión final sobre la identificación a partir de las superposiciones obtenidas en el paso anterior.

En la presente tesis se propone extender la funcionalidad del método existente de superposición cráneo-facial para desarrollar un procedimiento guiado por computador, más fiable y robusto. Por un lado, en la segunda etapa del proceso, hemos mejorado el método automático actual de solapamiento, modelando la relación espacial existente entre el hueso y el tejido blando, relativa al emparejamiento de los *landmarks* craneales y faciales correspondientes. Esta información está disponible en múltiples estudios antropométricos pero su naturaleza imprecisa ha provocado que no se considerase en métodos automáticos de este tipo. Además, se ha realizado un profundo estudio para aplicar las métricas más adecuadas de cara a la obtención del mejor solapamiento posible entre el cráneo y la cara. En la tercera etapa, proponemos un marco de trabajo para un sistema de ayuda a la toma de decisiones, que tiene en cuenta todas las fuentes de información y de incertidumbre implicadas en el proceso. Este sistema de ayuda a la toma de decisiones se ha

desarrollado con la ayuda de técnicas de visión por computador y conjuntos difusos y se ha evaluado y validado utilizando casos reales positivos y negativos, obteniendo muy buenos resultados.

Abstract

Within the forensic identification techniques, craniofacial superimposition is one of the most relevant skeleton-based approaches. It involves the process of overlaying a skull with one or more photographs of missing persons and the analysis of their morphological correspondence. This identification technique has a great application potentiality since nowadays the wide majority of the people have photographs (*ante-mortem* material) where their faces are clearly visible. The counterpart, the skull (*post-mortem* material), is a bone that hardly degrades with the effect of fire, humidity, high or low temperatures, time lapse, etc. Three consecutive stages for the whole CFS process have been distinguished [DCI$^+$11]: 1) Acquisition and processing of the materials, the skull (the 3D model in the latest techniques) and the textitante-mortem facial photographs of the possible candidates; 2) Skull-face overlay, which focuses on achieving the best possible superimposition of the skull and a single *ante-mortem* image; 3) Decision making, where the degree of support that the skull and the face belong to the same person (positive identification) or not (exclusion) is determined.

   Soft computing and computer vision present certain characteristics that become it powerful tools to automate craniofacial superimposition, reducing the time taken by the expert and obtaining an unbiased overlay result. In particular, evolutionary techniques for 3D reconstruction of the skull are presented in [SCD07a, SCD$^+$07b, SCD$^+$09, CÁCD13], highly useful for the first stage. In addition, the automatic system proposed in [ICDS09, ICDS11, ICD12, CAIN$^+$14] performs the skull-face overlay projecting the skull 3D model on the facial 2D image through a direct correspondence between cranial and facial landmarks using evolutionary algorithms and fuzzy sets. It also models the imprecision location of the facial landmarks in the photograph. This method has achieved promising results, however, the final decision of the third stage is made manually by the forensic anthropologist in view of the superimposition obtained in the previous step.

   The aim of this PhD dissertation is to extend the functionality of the current automatic CFS procedure in order to develop a more reliable and robust computer-aided system. Concerning the second stage of the process, we have accomplished an extension of the existing automatic methods to superimpose a skull 3D model on a facial photograph by modeling the facial soft tissue depth between corresponding pairs of cranial and facial landmarks. This information is available in several anthropometric studies but its imprecise nature has caused it not to be considered in automatic skull-face overlay methods yet. Besides, a deep study for applying the most appropriate metrics in order to obtain the best possible superimposition has been performed. In the third stage, we propose a complete framework for a decision support system, which takes into account all the sources of information and uncertainty involved in the process. This decision support system has been developed using computer vision techniques and fuzzy logic, and it has been evaluated and validated with real positive and negative cases obtaining really good performance.

Part I. Report

1 Statement

The aim of this section is to describe the main aspects of the current doctoral dissertation. First, an introduction to the general topic is reported. Then, a list of open problems is provided as the justification of this work. Later, the objectives to be achieved during the development of the PhD dissertation are described. Finally, the global structure of the dissertation is shown.

1.1 Introduction

Forensic anthropology applies the scientific knowledge of physical anthropology to the collection and analysis of legal evidence. This discipline includes recovery, description and identification of human skeletal remains [BW99, YS00] . Although there are other identification procedures more reliable than skeleton-based identification, i.e., comparison of fingerprints, foot and hand prints, external and internal autopsy, or DNA techniques, sometimes there is not enough (*ante* or *post-mortem*) information available to apply them. For example, in cases in which people are missing for a long time or under circumstances of war and mass disasters [Isc81]. Hence, anthropological identification based only on the skeletal information can be considered as the last chance for forensic identification when none of the previous methods can be applied.

Craniofacial superimposition (CFS) [Yos12] is one of the most relevant skeleton-based identification approaches. It involves the process of overlaying a skull (or a skull 3D model) with a number of *ante-mortem* (AM) images of an individual and the analysis of their morphological correspondence. By projecting the skull and the photographs of the missing person on top of each other, the practitioner can try to establish whether they correspond to the same individual.

Since the first documented use of CFS for identification purposes [GB37] the technique has been under continuous development. Although the foundations of the CFS method were laid by the end of the ninetieth century [Bro75], the associated procedures evolved as new technology was found available. Therefore, three main different approaches have been developed: photographic, video, and computer-aided superimposition [DCI+11, WR12, Yos12]. The first superimpositions involved acquiring the negative of the original facial photograph and marking the cephalometric landmarks on it. The same task was done with a photograph of the skull. Then, both negatives were overlapped and the positive was developed. This procedure was called photographic superimposition [Yos12]. Many authors further developed photographic superimposition techniques to improve the scale and the orientation of the skull and the facial images [BH89, Dor83, Maa89]. Video superimposition was introduced in 1976 [HG77]. Instead of marking photographs, tracings or drawings in order to properly superimpose the skull and the face, video cameras provide a "live image" of the object (skull, photograph) focused. These systems present an enormous advantage over the former photographic superimposition procedure by minimizing several problems associated to it. The video superimposition technique continued evolving [FHS08, LC93, SY93] and it became the most

broadly employed method. The popularization, huge development and larger amount of possibilities offered by computers turned them into the next generation of CFS systems. Two different system categories arise within this group [DCI$^+$11]. Non-automatic computer-aided methods use the computer for storing and/or visualizing the data [AAMG$^+$06, PVP$^+$93, RMA06, UBO92] but they do not consider the computational capacity to automate human tasks. Automatic computer-aided methods use a computer program to (semi) automatically accomplish any CFS sub-task itself [GS01, ICDS09, ICDS11, ICD12, NFKC91].

Regardless the technological means considered, three different stages for the whole CFS process were distinguished in [DCI$^+$11]:

1. The first stage involves the acquisition and processing of the skull (or skull 3D model) and the AM facial images. In some approaches this step also involves the cranial and facial landmark location.

2. The second stage is the skull-face overlay (SFO), which focuses on achieving the best possible superimposition of the skull and a single AM image of the suspect. This process is repeated for each available photograph, obtaining different overlays. SFO thus corresponds to what traditionally has been known as the replication of the skull size and its orientation with respect to the facial photograph [Yos12].

3. The third stage accomplishes decision making. Based on the superimpositions achieved in the latter SFO stage, the degree of support of being the same person or not (exclusion) is determined following two main approaches. On the one hand, in the landmark-based approach [GS12, YMK$^+$97, NM11] the skull-face relationships are evaluated according to the degree of matching of corresponding cephalometric and craniometric landmarks considering the soft tissue depth and the pose of the face in the photograph(s). On the other hand, in the morphological based approach [ASM94, JSA01, WR12] a visual assessment of how well the face matches the skull is made using a number of criteria purely based on a morphological comparison. There are also combinations of both approaches [GS12, Lan92, YIMS95] where distances between landmarks are evaluated together with the morphology.

Figure 1 depicts these three consecutive stages.

Designing automatic methods to address CFS and support the forensic anthropologist remains a challenge and dreamed milestone within the anthropology community. In fact, the development of computer-aided CFS methods has increased over the past twenty years [HIWK15]. Within artificial intelligence (AI), soft computing (SC) [Bon97, Zad01] is aimed for the design of intelligent systems to process uncertain, imprecise, and incomplete information. Methods based on SC applied to real-world problems often offer more robust and tractable solutions than those obtained by more conventional mathematical techniques. Among SC techniques, fuzzy logic (FL) [Zad65] 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 [Zad94]. On the other hand, evolutionary algorithms (EAs) are powerful bio-inspired search and optimization tools to automate problem solving in areas such as modeling, simulation, or global optimization [BFM97, ES03]. Besides, computer vision (CV) [SHB14] includes methods for acquiring, processing, analyzing, and understanding images and, in general, high-dimensional data from the real world in order to produce numerical or symbolic information, e.g., in the forms of decisions. Recent approaches use skull 3D models, CV and SC methods for the first two identification stages. These methods allow us to both automate the task and handle the inherent uncertainty [CÁCD13, ICDS09, ICDS11, ICD12, CÁIN$^+$15]. However, the thickness of the facial soft tissue differs for each corresponding pair of landmarks, it varies among individuals and produces a
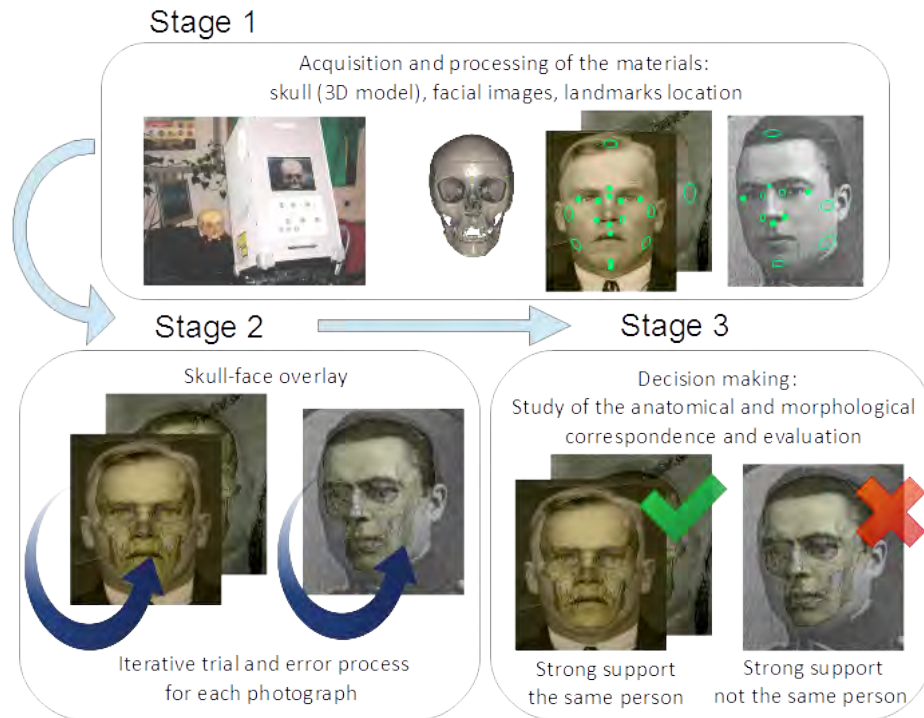
Figure 1: CFS procedure scheme.

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 [Ste09, GS12]. That fact represents a source of uncertainty that should be tackled during the whole CFS process.

In the third stage, once one or more SFOs are obtained, experts evaluate morphological and spatial skull-face relations, focusing on certain regions demonstrated to be more discriminative. The final decision is provided from an aggregation of the partial decisions. It is taken in terms of limited, moderate or strong support to the assumption that the skull and the facial image belong to the same individual or not [DWK+15]. This process is subjective and it relies on the skills of the forensic expert while influenced by the quantity and quality of the used materials. Therefore, there is a need to design a decision support system (DSS) to help practitioners to make their decision based on the fusion of the available information sources in a faster and more objective way. It would also lead to the application of CFS in identification scenarios with multiple comparisons, a possibility not explored yet due to the unaffordable time lapse needed to analyze all possible cross-comparisons.

1.2 Justification

Two different factors can be identified as the main limitations for recognizing CFS as an acceptable and reliable identification method. On the one hand, the absence of a systematic methodology accepted worldwide. Indeed, experts tends to apply their own approach to the problem based on their knowledge on human craniofacial anatomy, soft tissues, and on the available technologies. In particular, most forensic anthropologists manually apply a repetitive method in order to adjust the skull size and orient it with respect to the facial photograph until they achieve a good enough superimposition. That process is a very challenging and time-consuming part of the CFS process. In addition, every expert achieves different overlays, which have a definitive influence on the subsequent morphological correspondence analysis [IVN+15]. Thus, a systematic, unbiased,

accurate, and automatic method is necessary [Ube00]. On the other hand, CFS is usually performed as a subjective method that does not properly consider the different sources of uncertainty while in addition it is carried out as a time consuming, trial-and-error process.

Concerning the first drawback, the University of Granada and the PhD student, who is the author of this dissertation, collaborated as participants in the European project "New Methodologies and Protocols of Forensic Identification by Craniofacial Superimposition" (MEPROCS). This project proposed a common framework, avoiding particular assumptions that could bias the process and allowing the extensive application of the CFS technique in practical forensic identification scenarios. To achieve this purpose, the project joined the most representative experts in the field in a series of discussions, intended to identify and agree on the criteria to be considered in the proper implementation of the CFS technique. The common methodology designed includes a document with the good and bad practices to be considered through all the CFS process, sources of error and uncertainties, technological requirements and desirable features, and a common scale for the craniofacial matching evaluation [DWK$^+$15]. Two multiple-lab studies on craniofacial superimposition were also carried out in order to analyze the performance of a diverse set of CFS methodologies and the corresponding technical approaches when dealing with a common dataset of real-world cases. As a result, a novel methodology for understanding and standardizing identification methods based on the observation of the most convenient characteristics of every method was proposed [IVN$^+$15, IVC$^+$16]. In addition, a validation study of the proposed CFS framework was performed [IVN$^+$16]. Finally, a handbook with a complete study of CFS methodology will be published in the next few months [DIC17].

Regarding to the second limitation, some automatic methods for CFS have been developed until now [ICDS09, ICDS11, ICD12]. However, they are based on overlaying a skull 3D model and a facial photograph by minimizing the distance among pairs of landmarks as well as handling the imprecision due to the facial landmarks location [CGM$^+$13]. That minimization process involves the search for the specific projection of the skull 3D model that reduces all the distances between every pair of corresponding landmarks as much as possible. That is a good approximation to face the problem which provided reasonable results but it is not anatomically correct. In fact, the anatomical distance between a cranial landmark and its corresponding facial point (soft tissue depth) is not considered. Instead, the methods wrongly minimize the theoretical distance among corresponding landmarks. Solving that problem involves dealing with all the sources of uncertainty inherent to the CFS problem, which must be properly studied and identified, and proposing a more reliable automatic procedure to measure the quality of the resulting SFOs.

Besides, these existing methods just focuses on the SFO stage and do not provide any solution for the third stage, the decision making. In fact, decisions about the skull and face relationships are still made in a subjective way where there is a complete absence of measurements and countable aggregation of positive and negative factors. The experience and knowledge of the practitioner result to be the key aspect for a correct decision. Thus, there is a strong interest on designing and implementing a DSS for the decision making stage of the CFS technique.

In order to issue an identification result from the analysis of a SFO, it is necessary to establish clear and objective evaluation criteria to avoid possible bias in the interpretation of the result. Fortunately, a compendium of the anthropometric and morphological criteria that meet these characteristics and their evaluation was one of the main results of the MEPROCS European project [IVN$^+$16]. However, the obtained results did not show strong associations due to the reduced number of cases and their quality [IVC$^+$16]. Likewise, there is a need to design a complete framework for the SFO evaluation and the final decision making which takes into consideration all the necessary data and sources of uncertainty involved. Following this framework, the set of criteria should be integrated into a semi-automatic environment to practically allow CFS to be applied in multiple comparison scenarios. The semi-automatic system design should properly model the

expert knowledge in morphological comparisons expressed as prepositions of place and similarity of different but analogous anatomical structures (in both the face and the skull). Finally, taking into account both anthropometric measures and the morphological analysis the system should be able to properly aggregate all this information considering the sources of uncertainty. As a consequence, the DSS must be able to give a numerical value and/or verbal result on the correspondence between a skull and a face photograph. This result will always be validated by the expert at the time of issuing an identification report.

As one important conclusion, there is a reasonable and appropriate justification for the employment of SC and CV tools to tackle the challenging tasks proposed for the complete automation of the CFS procedure. Regarding to the SFO stage, the existing automatic method presents some aspects still needed to be refined to obtain more accurate results. Skull-face overlay is tackled following an image registration (IR) approach [ZF03] 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. Concerning SC, 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 objects, a skull and a face). 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. IR approaches based on EAs are a good solution for facing this challenging optimization problem. Thanks to their global optimization nature, EAs are capable to perform a robust search in complex and ill-defined problems as IR [SCD10, DCS11]. Besides, fuzzy sets have been considered in order to handle uncertainty related to the location of facial landmarks [ICDS11, ICD12]. In the same line, fuzzy sets can also be suitable tools for modeling the imprecision related to the facial soft tissue expecting to reduce the mismatch among pairs of cranial and facial landmarks. Concerning about the automation of the decision making stage, CV methods include a unique set of tools to deal with the analysis of the craniofacial correspondence as well as the delineation or segmentation of specific regions in both the skull and face for the study of their relationship. Nevertheless, SC are quite important in this stage. The morphological correspondence of face and skull structures is never precise due to the presence of soft tissue. Forensic anthropologists usually express the identification decision according to several confidence levels depending on the quality and quantity of the materials (AM photographs, mandible, and cranium) and the existence of discriminatory characteristics: strong support, moderate support, limited support, and undetermined [DWK+15]. As a result, the final identification decision is characterized by some uncertainty and partial truth. Thus, there is a need of using a tool capable of aggregating multiple criteria together considering their corresponding uncertainty for decision making. In this way, the final identification result must take into consideration all the uncertainty and the degrees of confidence inherent to the whole process.

Dealing with the previous issues when designing a new, better performing computer-aided CFS system is the main motivation for the current PhD dissertation.

1.3 Objectives

The main objective of this PhD dissertation is thus to extend and improve the functionality of the current automatic CFS system based on CV and SC in order to develop a semi-automatic method suitable for both individual and multiple comparisons. The resulting system will support forensic anthropologists in the human identification task by means of the CFS technique. It would become an invaluable tool for police departments, law enforcement, and forensic laboratories worldwide to address real world scenarios with a large number of missing people cases and mass disasters.

This overall aim will be achieved through the following specific objectives:

- To model the imprecision related to the matching of cranial and facial landmarks and to incorporate it to the existing automatic SFO method (second CFS stage). The thickness of the facial soft tissue differs for each corresponding pair of landmarks, varies among individuals, and produces a mismatch among landmarks of the skull and the face [Ste09, GS12]. That variability has been widely studied in many populations considering different age and gender subgroups [SS08a, SS08b, VC07]. By including the soft tissue depth measurements of the latter studies in the SFO method, we aim to model the distances between a pair of cranial and facial landmarks, known as landmark matching uncertainty [ICDS11].

- To study the performance and influence of different fuzzy distance definitions on the previous SFO method. Since our approach is based on automatically overlaying a skull three dimensional model onto a facial photograph by minimizing the distance between two fuzzy subsets (corresponding cranial and facial landmarks), the choice of a good distance metric can be crucial to our method. The system will be objectively evaluated considering the unique ground truth data-set of this kind.

- To develop an automatic assessment of the skull-face relationships based on landmark matching and morphological spatial relationship in CFS. In particular, to design and model a semi-automatic algorithm able to evaluate the spatial morphological correspondence of the superimposed skull and face. The inherent uncertainty will be included as part of the resulting matching degree.

- To design and implement a semi-automatic CFS decision support system for scenarios of multiple comparisons. The aim is to design and implement a fuzzy decision support system integrating quantitative and qualitative morphological and morphometric information. This system should be able to properly model, propagate and aggregate the inherent uncertainty. Given a set of candidates it should be able to properly rank them based on a set of criteria defined by the forensic expert.

- To develop a system validation and a performance study. Using some data-sets with skull 3D models and digital photographs of different persons, the new automatic procedure should be validated. This set of cases will allow us to test the reliability of the proposed (semi-) automatic identification system using a large study with many positive and negative cases.

1.4 Structure

This PhD dissertation is divided into two parts. The first one is devoted to the statement of the problem, the revision of the current state of the art, the description of the proposed improvements to the existing automatic CFS system, the discussion of the results obtained, and the presentation of the lines for future works. The second one consists of the main scientific publications obtained as a result of the work developed.

In particular, Part I is organized as follows. After the introduction to the problem, we present a set of open issues justifying this PhD dissertation, as well as its main objectives. Then, in Section 2 we review the basic concepts about CFS, SFO, and the automatic CFS system that will be used as a base for our proposal. Section 3 introduces our new proposals for the latter three-stage CFS system, describing each method in detail. Section 4 discusses the results obtained. Finally, Section 5 shows the future research lines arising from our developments.

The work developed to achieve the stated objectives is described in the five scientific publications composing Part II of this PhD dissertation:

- B. R. Campomanes-Álvarez, O. Ibáñez, C. Campomanes-Álvarez, S. Damas, and O. Cordón. Modeling the Soft Tissue Thickness for Automatic Skull-Face Overlay, IEEE Transactions

on Information Forensics and Security, vol. 10, no 10, pp. 2057 - 2070, 2015. DOI: 10.1109/TIFS.2015.2441000. Impact factor: 2.441 Category: COMPUTER NETWORKS AND COMMUNICATIONS. Order: 11/203. Q1.

- C. Campomanes-Álvarez, B. R. Campomanes-Álvarez, S. Guadarrama, O. Ibáñez and O. Cordón. An Experimental Study on Fuzzy Distances for Skull-Face Overlay in Craniofacial Superimposition, Fuzzy Sets and Systems, vol. 318, pp. 100-119, 2017. DOI: http://dx.doi.org/10.1016/j.fss.2016.06.015. Impact factor 2015: 2.098 Category: COMPUTER SCIENCE, THEORY AND METHODS. Order: 13/105. Q1.

- C. Campomanes-Álvarez, O. Ibáñez and O. Cordón. Design of Criteria to Asses Craniofacial Correspondence in Forensic Identification based on Computer Vision and Fuzzy Integrals, Applied Soft Computing, vol. 46, pp. 596 - 612, 2016. DOI: http://dx.doi.org/10.1016/j.asoc.2015.11.006. Impact factor 2015: 2.857 Category: COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE Order: 21/130. Q1.

- C. Campomanes-Álvarez, O. Ibáñez, O. Cordón and C. Wilkinson. Hierarchical Information Fusion for Decision Making in Craniofacial Superimposition, Information Fusion. In press. DOI: http://dx.doi.org/10.1016/j.inffus.2017.03.004. Impact factor 2015: 4.353 Category: COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE Order: 9/130. Q1.

2 State of the Art

This section reviews the state of the art of forensic identification by CFS, focusing on the existing computer-aided approaches. We analyze the role of CV and SC in the three CFS stages more deeply.

2.1 Craniofacial Superimposition

Craniofacial superimposition [Yos12] is one of the approaches in craniofacial identification [AISB95, Ste09]. It is considered one of the most relevant skeleton-based identification techniques. The approach involves the superimposition of a skull with a number of *ante-mortem* images of an individual and the analysis of their morphological correspondence. Its basis were established more than 100 years ago by Broca and Bertillon [Bro75, Ber96]. They respectively focused on studying the correspondence of the cranial structures with the soft tissue covering them and on the basis to collect physiognomic data of the accused of a crime. Those data are still used nowadays. Since then, CFS evolved as new technology was available from its previously laid foundations. Three main families of CFS methods arose through this technology-based evolution: photographic, video, and computer-aided superimposition methods (see Sec. 2.2).

The first documented cases of CFS date to the early decades of the 20th century. In these instances, AM images were compared to cranial remains in order to support evidence to be presented in court rather than as the principal means of identification of the victim. A significant example is the Ruxton case in 1937, in which the skulls of two females were compared with some photographs of missing women using the superimposition technique [GB37]. In other instances, the AM images were created by a draw line of the head of the missing person to be superimposed on the skull [GD48].

The first positive identification case based on CFS was in India in 1962 and it was accepted in a court of law [Sen62, Ube00]. During the following 50 years, the technique became more extended and several CFS identification cases were performed [Sek71, BDW86, Ube00]. In fact, the CFS procedure was highly helpful in the identification of criminals as Adolf Hitler's chief medical officer Dr. Josef Mengele at Sao Paulo, Brazil in 1985 [Hel86]. Besides, CFS is currently used in different

scenarios where other identification techniques are not available, such as the recent Indian Ocean tsunami [AAMG$^+$06] and terrorism [Ind09].

Recently, Damas et al. [DCI$^+$11] have distinguished three consecutive stages in the whole CFS process:

- The first stage involves the acquisition and processing of the materials, i.e., the skull and the AM facial photographs, as well as the location of anatomical landmarks on both by the forensic expert. This step is not necessary in some systems. The oldest approach but still used consists of acquiring photograph(s) and/or a series of video frames of the skull. However, the most advanced systems use a digital 3D model of the skull. Regarding the facial photograph, most recent systems use 2D digital images. In these cases, this stage also involves the application of image processing techniques to enhance the quality of the photograph [GW08].

- The second stage is SFO. It consists of searching for the best superimposition of a skull (or a skull 3D model) and a single AM image of a missing person. This process is iteratively repeated for each available AM image, obtaining different overlays. Skull-face overlay thus refers to what traditionally has been known as the adjustment of the skull size and its orientation with respect to the facial photograph [Yos12].

- Finally, the third stage corresponds to the decision making. Based on the overlays achieved in the previous stage, the identification decision is made by judging 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 depths, and the consistency between asymmetries [JSA01]. As a result, the degree of support that the skull and the available photograph belong to the same person or not (exclusion) is determined. Given the subjective and qualitative nature of the latter issues, the identification decision provided by the experts is usually characterized by several confidence levels as strong, moderate, or limit support of (not) being the same person or undetermined when it is not possible to establish either a positive or negative relationship between a skull and a candidate.

Despite the technical procedure followed by the experts, cranial and facial landmarks are often used in one o more of the previous stages. 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. Among the different defined landmarks, forensic anthropologists usually consider the set of landmarks proposed by Martin and Saller in 1956 [MS56].
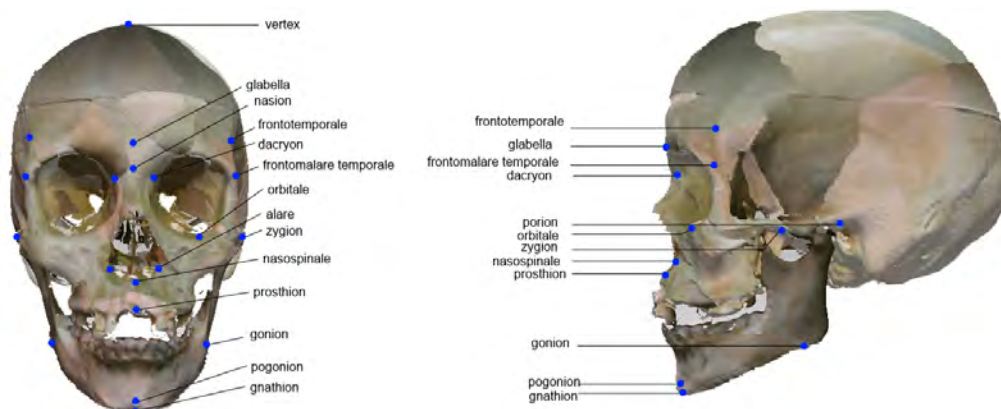
Figure 2: Cranial landmarks in frontal and lateral view.

The skull landmarks [Geo93] that are typically used (see Figure 2) in CFS are the following:

Glabella (g): The most prominent point between the supraorbital ridges in the midsagittal plane.

Nasion (n): The midpoint of the suture between the frontal and the two nasal bones.

Frontomalare temporale left, right (fmtl, fmtr): The most lateral point of junction of the frontal and zigomatic bones.

Dacryon left, right (dal, dar): The point of junction of the frontal, maxillary, and lacrimal bones on the lateral wall of the orbit.

Alare left, right (alal, alar): The most laterally positioned point on the anterior margin of the nasal aperture. This point should be marked on both the left and right sides of the nasal aperture.

Nasospinale (ns): The point where a line drawn between the lower margins of the right and left nasal apertures is intersected by the midsagittal plane.

Zygion left, right (zyl, zyr): The most lateral point on the zygomatic arch.

Prosthion (pr): The apex of the alveolus in the midline between the maxillary central incisor.

Gonion left, right (gol, gor): The most lateral point of the jaw line at the mandibular angle.

Pogonion (pg): The most anterior point in the midline on the mental protuberance.

Gnathion (gn): A constructed point midway between the most anterior and most inferior points on the chin.
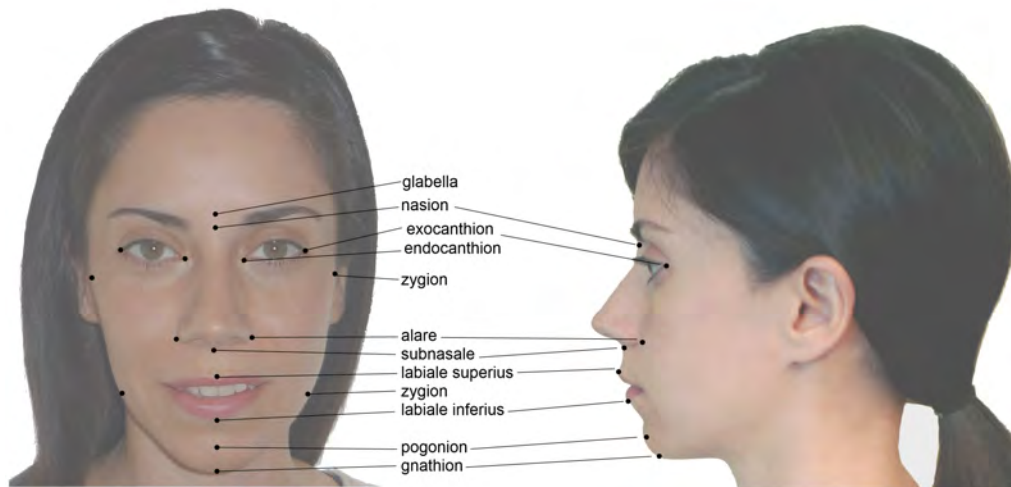
Figure 3: Facial landmarks in frontal and lateral view.

Meanwhile, the corresponding used facial landmarks (see Figure 3) are:

Glabella (g): In the midline, the most prominent point between the eyebrows.

Nasion (n): The midpoint on the soft tissue contour of the base of the nasal root at the level of the frontonasal suture.

Endocanthion left, right (enl, enr): The point at the inner commisure (medial canthus) of each palpebral fissure.

Exocanthion left, right (exl, exr): The soft tissue point located at the outer commissure of each eye fissure.

Alare left, right (alal, alar): The most lateral point on each alar contour where nose meets the skin of the philtrum and cheek.

Subnasale (sn): The midpoint on the nasolabial soft tissue contour between the columella crest and the upper lip.

Zygion left, right (zyl, zyr): The most lateral point on the soft tissue contour of each zygomatic arch.

Labiale superius (ls): The midpoint on the vermilion line of the upper lip.

Labiale inferius (li): The midpoint of the vermilion line of the lower lip.

Gonion left, right (gol, gor): The most lateral point of the jaw line at the mandibular angle.

Pogonion (pg): The most anterior midpoint on the soft tissue chin.

Gnathion (gn): The most inferior point on the soft tissue contour of the chin.

2.2   Technical Overview of Craniofacial Superimposition Methods

The different CFS approaches have been evolving as new technologies have became available [Isc93, TB98]. Although there are several classifications of the technique, three different categories are clearly identified: 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) [YS00, AISB95]. Yoshino et al. [YMK+97] classified some of the computer-aided CFS methods into two categories based on the identification strategy, i.e., morphological and morphometrical analysis. The former bases on the morphology matching between the skull and the face [JSA01]. The latter relies on the measurement of distances between pairs of landmarks and their comparisons to average facial tissue depths. Damas et al. [DCI+11] proposed a classification for computer-aided CFS techniques based on the role of computers that take part in the process: non-automatic and automatic methods. Recently, Huete et al. [HIWK15] extended this computer-aided category into three new subcategories: a) Computer-aided craniofacial photo superimposition, b) Computer-aided craniofacial video superimposition, and c) Computer-aided craniofacial 3D-2D superimposition.

The first CFS methods involved obtaining the negative of the original facial photograph and marking contours and/or facial landmarks on it. The same task was done with a photograph of the skull. Both negatives were overlapped and the positive was developed. This procedure was known as photographic superimposition [Yos12]. Many authors further developed photographic superimposition techniques to improve the scale and the orientation of the skull and the facial images [Dor83, BH89, Maa89].

Video superimposition was introduced in 1976 [HG77]. Video cameras provided a "live image" of the object (skull, photograph) focused, which presented an enormous advantage over the former photographic superimposition procedure. The video superimposition technique continued evolving [SY93, LC93, JSA01, FHS08] and it became the most broadly employed method.

The next generation of CFS system arose with the popularization, huge development and larger amount of possibilities offered by computers. Two kinds of methods can be differenciated: Non-automatic computer-aided and automatic computer-aided methods. The former use the computer only for storing and/or visualizing the data [UBO92, DCV+86, RMA06, AAMG+06] but without considering the computational capacity to automate human tasks. The latter employ a computer program to accomplish one or more CFS tasks in an automatic way supporting the forensic anthropologist's work. Thus, the time is drastically reduced and the obtained results can be considered unbiased. The following section reviews the main existing computer-aided CFS approaches.

2.3 Overview of the Existing Computer-Aided Approaches for the Three Craniofacial Superimposition Stages

2.3.1 Acquisition and Processing of the Materials

The original materials used in the CFS process include photographs of the person and the proper skull. Some of the most representative experts in craniofacial identification recently defined best practices to follow in CFS [DWK$^+$15]. Following these criteria, the experts filter the materials in a first phase and if they are not reliable, the CFS cannot be carried out. Regarding to the AM photographs, the face has to be clearly visible. Authors recommend to use more than one AM images in different poses and with a good quality, as well as to avoid images with obscuring objects, e.g. spectacles and beards. The photographs were acquired under some conditions that are fixed and usually unknown at the moment of the forensic analysis. With a digital image, the only possibility is to attempt to enhance its quality. If it is not in digital format, it can be scanned and transformed into a 2D digital image. Then, it can also be enhanced using digital image filters and/or processing algorithms. However, experts recommend to avoid as much as possible image manipulation, preserving the aspect ratio and keeping all the information contained within the original image, i.e., not cropping the images. For optimal examination of full frontal images the resolution of the face image should be at least 180 pixels corresponding to the width of the head, or roughly 90 pixels between the pupils of the eyes (ISO International Standard ISO/IEC JTC 1/SC 37 N506). On the opposite, the skull is an available physical object and its model needs to be obtained to accomplish the CFS procedure. Nevertheless, experts advise to use the real skull to confirm correct fit of the mandible with the cranium and to establish the centric occlusion. In addition, they suggest to locate and mark landmarks on the skull before scanning.

The computerized systems developed for the first stage of CFS are related to face enhancement and skull modeling procedures. Concerning the image of the face, enhancement techniques could be applied [GW08] in order to improve its quality. Such techniques depend on the available format (digital camera image or scanned photographic paper) and include frequency domain filters to fix artifacts due to aliasing and sampling problems present in scanned documents, as well as removal of non-uniform illumination effects and sharpening methods to deal with blurring and problems related to movements. Notice that the proper filter and its most suitable parameters are a choice that must be performed by the expert since they depend strongly on the acquisition conditions. Approaches that use human operated commercial software for the facial 2D image enhancement are considered non-automatic methods. Automatic methods perform such 2D image enhancement using computer programs with almost no human intervention [DCI$^+$11].

Skull 2D images, skull live images (video superimposition), or skull 3D models can be used in CFS. This subsection will mainly focus on contributions that include an automatic 3D modeling procedure since the other methods do not consider this stage and directly acquire a 2D projection of the skull (i.e., a skull photograph).

Obtaining an accurate skull 3D model has been considered a difficult task by forensic anthropologists in the past. Nowadays, it is an affordable and attainable activity using advanced scanning devices like laser range scanners. Current 3D scanners allow the forensic anthropologist to get skull 3D models with a precision lower than one millimeter in a reasonable time [PCK06]. The subject of the identification process, i.e. the skull, is a 3D object. Hence, the use of a skull 3D model instead of a skull 2D image should be preferred in CFS because it is definitively a more accurate representation. It has already been shown that 3D models are more informative in other forensic identification tasks [ASC$^+$09]. In the biomedical field, computed tomography (CT) scanning images are the starting data to reconstruct the skull [SLW$^+$09, FCPB08]. However, the possibilities of recording 3D forensic objects are not so many considering the available resources of a typical forensic anthropology laboratory. Indeed, many forensic laboratories are exploiting the

capabilities of laser range scanners nowadays since these devices present a greater availability and a lower cost.

Laser range scanners are based on the optical principle of triangulation and acquire a dense set of 3D point data in a very rapid, non-contact way [BR02]. Some laser range scanners are equipped with an additional positioning device named rotary table and appropriate software that permits the 3D reconstruction. Nevertheless, there are situations where that software does not provide suitable 3D models. Moreover, there are scenarios where it is not even possible to use a rotary table.

Before going on with the 3D reconstruction, every 3D view of the skull acquired by the laser range scanner must be preprocessed. This task involves the cleaning, smoothing, and filling of the view. Cleaning aims to remove those artifacts that were acquired by the scanner as part of the scene but which do not correspond to the skull. Meanwhile, smoothing is mainly concerned with the removal of some artificial vertexes that could have been wrongly included by the scanner on the borders of the surface because of a perspective distortion. Fortunately, this task is not needed so often. Finally, filling is used to avoid small holes to appear in those parts of the skull that are not properly scanned because they are too dark for the scanner capabilities or because they are located in shadow regions.

In order to accomplish the 3D model, some anthropologists are skilled enough to deal with the set of 3D views and they supervise the procedure with a commercial software like RapidForm^{TM}. Sometimes, this software does not provide the expected outcomes and the anthropologists even have to *stitch up* manually every couple of adjacent views. Hence, 3D image reconstruction software is a real need to construct the 3D model by aligning the views in a common coordinate frame. Such process is usually referred as range IR [IS01, ZF03]. It consists of finding the best 3D rigid transformation (composed of a rotation and a translation) to align the acquired views of the object.

Up to our knowledge, Nickerson et al. [NFKC91] were the first researchers to propose the use of a 3D model to tackle the CFS problem. In their work, a range scanner and a digital camera were used for 3D digitization of the skull surface mesh and the 2D AM facial photograph, respectively. Well known image processing algorithms were used for image enhancement (median filtering, histogram equalization, Wiener filtering) [GW08]. Rendering was done through computer graphics techniques. A feature-based algorithm to reduce the computational and memory complexities inherent in solid modeling was also described.

A completely different approach is presented in [BSA05] where the authors examined the applicability of holography in the 3D recording of forensic objects. Holography is an optical technique capable of recording the 3D data of an object. Two types of images, real and virtual, can be recorded in a holographically exposed film or hologram. Two superimposition systems using holographic images were examined in order to evaluate the potential use of this recording method. The authors claim that the performance of holography is comparable to that of the computer graphics system, which consists of an image scanner, software, and a display unit. Moreover, they argue it can even be superior to the computer technique with respect to the 3D reconstruction of images. The suitability of this technique needs further studies. In particular, the use of an automatic superimposition method and a comparison with a reconstructed 3D range image could have objectively proved the actual utility of holography in this field.

Galantucci et al. [GPA$^+$06] benchmarked two different acquisition techniques of images of a skull. In particular, CT and laser scanner performance was compared to ascertain which enabled more accurate reproductions of the original specimen. Comparison between the original object and every model yielded satisfactory results for both techniques. However, CT scanning demonstrated some advantages over the laser technique, as it provided a cleaner point cloud, enabling shorter pre-processing times, as well as data on the internal parts, which resulted in the reproduction of a more faithful model.

Santamaría et al. [SCD07a, SCD$^+$07b, SCD$^+$09] proposed a method, based on evolutionary algorithms (EAs), for the automatic alignment of skull range images. Different views of the skull to be modeled were acquired by using a laser range scanner. A two step pair-wise range IR technique was successfully applied to such images. The method includes a pre-alignment stage that uses a scatter-search-based EA [LM03] and a refinement stage based on the classical iterative closest point IR algorithm [BM92]. The method is very robust since it reconstructs the skull 3D model even if there is no turn table and the views are wrongly scanned.

Ballerini et al. [BCDS09] proposed the automatic reduction of the data provided by the laser range scanner used in the skull 3D model reconstruction task. The dense point cloud corresponding to every skull view is synthesized by considering heuristic features that are based on the curvature values of the skull surface. Those features guide the automatic skull 3D model reconstruction by means of an EA.

Finally, Campomanes et al. [CÁCD13] presented a new algorithm to deal with the 3D open model mesh simplification problem from an evolutionary multi-objective point of view. An open model refers to a surface with open ends. The problem is based on the location of a certain number of points in order to approximate a mesh as accurately as possible to the initial surface. The algorithm considers two conflicting objectives, the accuracy and the simplicity of a simplified 3D mesh.

2.3.2 Skull-Face Overlay

Several proposals have been presented in the literature to perform the SFO task. Non computer-aided methods follow a manual process to positioning the skull in the same pose of the face in the image. The forensic experts calculates the size of the head and the orientation in the image. Some of them perform the SFO by calculating an enlargement factor based on linear measurements of items within the AM image [GB37, Sek93] or by measuring the interpupillary distance and the size of dentition [ASM94]. In 1989, Maat proposed to use a set of landmarks and a reference line in order to calculate three components of head rotation to position the skull [Maa89]. Other works suggested the use of a vertical distance between the landmarks ectocanthion and tragion as a measure for calculating the extension of the head [Sek93]. In the case of video-superimposition, the orientation of the skull can be calculated in the same manner as in photographic superimposition, but the adjustment of the size of the skull is easier to achieve thanks to the zoom mechanism of the video camera [Yos12].

Concerning to computer-aided approaches, digital infraestructures are used for dealing with the SFO task. Lan and Cai [LC85] developed a dual projection system, named TLGA-1. Later, versions TLGA-2 was performed to improve the original. Then, a new version, called TLGA-213 was presented. It consists of a TV camera, a computer, an A/D and D/A converter, a mouse, and a software program. The angle of the photograph is calculated by measuring the distances between glabella and gnathion to nasion. The head size is estimated using the distance between the two ectocanthions and the deflection angle in the photograph. Those parameters are iteratively computed and considered for the manual SFO task [Tao86, LC88, LC93].

Ubelaker et al. [UBO92] developed a software which digitizes images of both skull and face and helps to assess the consistency between them. Then, the facial photograph is shown and experts can trace anatomical landmarks on a plastic slide taped on the monitor. The skull can also be visualized. It is manually manipulated to match the marked landmarks. This tool was used for solving a huge number of cases submitted to the Smithsonian Institute by the FBI.

Later, Ricci et al. [RMA06] proposed a method for comparing a skull 2D image with a facial 2D image. The corresponding anatomical landmarks are positioned in the skull and facial images, and then, the software tool calculates the distances between those landmarks. Finally, a forensic expert manually performs the superimposition of the images.

Other practitioners consider commercial software packages as Adobe Photoshop^{TM} to adjust the scale of the facial photograph and project it over the skull photograph using tools for moving, rotating, and resizing the images [SN02, BKA$^+$03, AAMG$^+$06, RMA06].

Regarding computer-aided automatic methods, CV techniques are used to make the SFO operation automatic. From a computational point of view, there is a clear relation between the superimposition task and the IR problem in CV [ZF03]. Besides, as explained before, artificial intelligence and machine learning algorithms such as NNs, EAs and FL are applied in order to ease the SFO task [Mit97, RM86, BFM97, Zad65].

Nickerson et al. [NFKC91] proposed a method to find the best overlay between a skull 3D model onto a digital facial image. The superimposition is achieved guided by four landmarks previously located in the skull and the face. The overlay is calculated with similarity transformations and a perspective projection basing on parameters of these operations. In order to find the best possible superimposition, these parameters have to be optimized. Three different AI methods, i.e., a heuristic method, a classic non-linear optimizer and a binary-coded genetic algorithm, are compared for this performance. The latter of these three achieved the best results.

Another automatic method for the SFO process was presented in [GS01]. The approach considers two different NNs. It takes into account the ambiguities due to soft tissue thickness during the selection of facial features. The system can implement an objective assessment of the symmetry between two nearly front 2D images, the skull and the facial image, that are the inputs. The output is the mapped skull image suitable for superimposition. The networks need to be trained separately because each of them can correctly map only a part of the cranial image.

Ibáñez et al. [ICDS09] presented an extension of the original proposal of [NFKC91]. The approach automatically superimposes a skull 3D model and an AM photograph in a fully automatic way. The forensic experts previously locate a number of different landmarks on the skull 3D model and on the facial image. The use of a large number of landmarks improves the robustness of the system. The optimization of the projection parameters is performed using a genetic algorithm. This proposal uses a real coding scheme and a better design of the genetic algorithm components with respect to the previous one. The authors accomplished a broader study for demonstrating that real-coded EAs are suitable approaches for SFO. In particular, they highlight the good performance and high robustness of the state-of-the-art covariance matrix adaptation evolution strategy (CMA-ES) [HO01] when applied to SFO.

Later, an improvement to the previous method was developed in [ICDS11]. This new proposal uses fuzzy sets to handle the uncertainty involved in the location of the facial landmarks in an invariable place [RPE$^+$95]. In this method, two different approaches to handle this uncertainty are introduced: weighted and fuzzy-set-based landmarks.

A novel SFO approach was presented by Ibáñez et al. in [ICD12]. This automatic system considers a cooperative coevolutionary algorithm to deal with the use of imprecise facial landmarks in the SFO task. Thus, the approach looks for both the best projection parameters and the best landmark locations at the same time.

Campomanes et al. [CAIN$^+$14] justified and analyzed the use of the computational methods to properly model the skull–face overlay problem. The authors also presented the automatic technical procedure using CV and SC and showed four overlays obtained in two craniofacial superimposition cases. This automatic procedure is proposed to be considered as a tool to aid forensic anthropologists to develop the skull–face overlay, automating and avoiding subjectivity of the most tedious task within craniofacial superimposition.

Finally, in [ICCÁ$^+$15], Ibáñez et al. presented the procedure and also created a ground truth dataset which allow to measure and compare the performance of both manual and automatic SFO methods following an objective and reliable way to assess the SFO result achieved. With this

ground truth dataset the performance of novel automatic methods can be objectively measured and compared against previous and future proposals.

2.3.3 Skull-Face Overlay Assessment and Decision Making

Once one or several skull-face overlays have been achieved for the same identification case, the main goal is to determine the degree of support for the assertion that the skull and the face of the photograph(s) belong to the same person or not. This is a subjective process that relies on the forensic expert's skills and the quantity and quality of the used materials. Although the systems reviewed in this section are labeled as automatic, the supervision and the final validation is always required as in any computer-aided medical diagnosis system [Ber07]. These algorithms are applied on the digitized images stored on the computer, after the determination of the orientation and the skull size by 'routine' SFO techniques.

There are just a few works tackling the automation of the analysis of craniofacial correspondences within the framework of CFS identification. This particular identification technique represents a challenging problem, very linked to the forensic anthropologist's expertise and non-automated yet. The existing literature was published more than 18 years ago and the works are very basic and limited. In addition, they do not consider the use of either skull 3D models or computer techniques to perform the SFO. The techniques used for shape analysis also implies manual interaction.

Tao [Tao86] developed the first procedure in which a computer was used for the decision making stage. That decision support system aimed to replace the previously used methods based on range estimation and subjective judgment. The system provided an identification conclusion by using distances between landmarks from the superimposed images. Later, Lan and Cai proposed 52 different superimposition identification indexes for that aim in the TLGA-213 system [Lan90, LC88, LC93]. These indexes were based on anthropometrical measures of Chinese adults, male and females, and were used together with proportion and distances between superimposed landmarks lines to automatically compute the final identification decision.

Pesce Delfino et al. [DCV$^+$86, PVP$^+$93] applied k-th-order polynomial functions and Fourier harmonic analysis to assess the fit between the outline of the skull and the face. Ten cases including positive and negative identifications were studied. The polynomial function was used to smooth the curve representing the investigated profile. The square root of the mean square error is taken to calculate the distance between polynomial function curves obtained for the skull and the face profile. The Fourier analysis considered the profile as an irregular periodic function whose sinusoidal contributors are found. Low-order harmonics (the first three or four) represented the basic profile shape and the high order harmonics corresponded to detail. The sum of the amplitude differences of the sinusoidal contributors between profiles of the skull and the face represented the second independent parameter for numerical comparison. A Janus procedure (so called by the authors because of the double-headed Latin god Janus, the bi-front) was used to evaluate the symmetry differences between the two profiles. This procedure takes into account the relationship between the total arc and the chord length and the area they delimit in the two faced profiles. All these parameters are calculated by a computer software package called Shape Analytic Morphometry. However, this method would be only applicable when lateral or oblique photographs are available. Furthermore, their contribution requires manual repositioning of the skull for the correct superimposition.

Bajnóczky and Királyfalvi [BK95] used the difference between the coordinate values of the pair of anatomical and/or anthropometrical points in both skull and face for judging the match between the skull and facial image obtained by the superimposition technique. Eight to 12 pairs of points were recorded and expressed as pixel units. Then, the final matrix, containing coordinates of measured points and calculated values, was established by computer-aided processing. Lacking

the appropriate information, their model assumed that all data in that matrix were independent and followed a normal distribution with the same variance. A part of that variance was σ^2, which was the square of the measurement error and was itself assumed to be the same for all the data. The model of the authors was based on assumptions that *the components of the error term are independent and distributed according to*:

$$N(0, 2\sigma^2). \tag{I.1}$$

The authors used a presupposed value of σ as part of the model assumptions. Under the assumption that the null hypothesis (Eq. I.1) is valid, it was statistically tested using two values for σ. Authors claimed that when a given case is evaluated it is crucial to know what value can be considered as measurement error. One skull and two photographs were used to test the method. Both frontal and lateral face photographs are considered. They noted that their method is suitable for filtering out false positive identifications. Although the results obtained from this method are objective and easily interpreted for lay people, the anatomical and anthropometrical consistency between the skull and the face should be assessed by forensic examiners who are well versed in the anatomy of the skull and face. The authors concluded that their method should be used only in combination with classic video superimposition and could be regarded as an independent check.

In [YMK$^+$97], Yoshino et al. presented a skull identification system where the distance between the landmarks and the thickness of the soft tissue of the anthropometrical points were semi-automatically measured on the computer screen for the assessment of the anatomical consistency between the digitized skull and face. The software also included polynomial functions and Fourier harmonic analysis for evaluating the match of outlines. To extract the outline, authors used edge segmentation thresholding operations. The final evaluation parameter is the root of the average quadratic error and sum of the difference in amplitude value of the fists Fourier harmonics. This final value was not normalized and authors did not establish a threshold value on which to base positive or negative identification through appropriate procedures of discriminant analysis.

A different approach was presented by Ricci et al. [RMA06], where an algorithm calculates the distance between crosses manually marked by the expert in both face and skull radiograph images. The mean value of the total distance between crosses represents the index of similarity between the given face and skull: the smaller the index value, the greater the similarity. Again, this index is not normalized, so this method is only valid for cross-comparisons but it does not provide an index of correspondence in a particular skull-face overlay. The algorithm matches a facial photograph to the correct skull in 100% of the cases following this approach: distances between anatomical points vs. the analysis of the consistency of morphological traits. However, the former is only valid under limited and unpractical conditions: the skull and face images have to be acquired in precise anterior position. Thus, the validity of the method has not been tested on real CFS identification scenarios and it serves more as a proof of craniofacial correspondence uniqueness. Besides, it is strongly sensitive to the SFO accuracy, a complex problem not addressed in that work since they superimpose images acquired under controlled conditions.

2.4 Computer Vision and Soft Computing for Craniofacial Superimposition

Several applications of CV and SC in forensic anthropology have been successfully developed, such as age estimation, facial soft thickness prediction, and facial identification [AFLGMFAL04, DMB$^+$11, ASC$^+$09].

As mentioned in the previous section, the latest studies that deal with CFS in an automatic way are based on CV and SC [SCD$^+$07b, SCD$^+$09, ICDS09, ICDS11, ICD12, CAIN$^+$14]. These computational methods can be extremely useful for the three stages of an advanced computer-aided CFS system.

Concerning the second stage, SFO can be formulated following an IR approach aiming to superimpose the skull 3D model over the facial image. However, this approach involves a complex optimization process. Robust and precise results are demanded, so IR methods based on EAs and FL are a promising solution for facing this challenging optimization problem as it has already been demonstrated in the clinical field [DCI$^+$11].

Evolutionary algorithms and FL show powerful tools to automate the SFO task:

- The manual procedure of SFO is a time-consuming task. The correct orientation of the skull with respect to the face in the photograph and the adjustment of its size can take several hours to complete [FHS08].

- The underlying task is an extremely complex IR optimization problem, so SFO can be naturally modeled as a 3D skull - 2D face photograph IR problem to be solved in an automatic way using EAs.

- Some landmarks present an imprecise definition [Boo97]. The poor quality of some photographs, the pose of the face and other elements that originate the occlusion of some landmarks cause an inaccurate location of the facial landmarks. Fuzzy sets can also be used to deal with these uncertainty sources in automatic procedures [ICDS11].

- Every forensic expert is prone to identify each specific landmark in a slightly different position, regardless of the means that are used to represent the involved objects (a skull and a face) [CGM$^+$13].

- The facial soft tissue, whose thickness differs for each corresponding pair of landmarks and varies among individuals, produces a mismatch among landmarks of the skull and the face, which can also be modeled using fuzzy sets [GS01, ICDS11].

Finally, regarding the decision making stage, automatic systems assist the forensic expert by applying decision support systems [Kee78] (see Sec. 3.3). These computer programs must use objective and numerical data for assessing the correspondence between the skull and the face. Based on this evaluation, the system has to be able to suggest an identification decision to the practitioner.

The third stage of CFS shows many characteristics which make CV and FL powerful tools to automate it:

- The delineation (or segmentation) of specific regions in both the skull and face can be valuable information for the study of their relationship.

- The morphological correspondence of face and skull structures is never precise due to the presence of soft tissues.

- The spatial relationship of face and skull structures is expressed in terms of prepositions of place.

- There is a need of considering multiple criteria together with their corresponding uncertainty for decision making.

- Uncertainty and degrees of confidence are inherent to the final identification result.

In the following three subsections we will review the basis of the automatic SFO system introduced in [ICDS09, ICDS11, ICD12] together with the very last proposals investigating landmark location dispersion [CÁIN$^+$15] and the procedure to generate ground truth cases [ICCÁ$^+$15]. This system represents the starting point of the work to be done within this dissertation, which will make use of the ground truth data set created in [ICCÁ$^+$15] to evaluate our upcoming proposals related with the second CFS stage.

2.4.1   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 AM photograph is the result of the 2D projection of a real (3D) scene that was acquired by a particular (unknown) camera [Fau93]. In such a scene, the living person 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 latter original scenario. To do so, a 3D model of the skull must be considered. Those models can be properly handled using the computer, making the automation of the SFO task easier (see Sec. 2.3.1). 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. In addition, the specific characteristics of the camera must also be replicated to reproduce the original as far as possible.

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 photograph [Yos12, FHS08]. Then a perspective projection of the skull 3D model is performed onto the facial photograph. Thus, the described SFO approach involves a 3D-2D IR task, which applies a geometric transformation f to the skull 3D model reproducing the face pose in the snapshot moment.

The 3D-2D IR task tries to reproduce the original scenario with several unknowns coming from two sources [ICDS09]:

- The camera configuration at the moment of the acquisition: the camera parameters have an influence in the SFO problem. Some of them are directly reflected on the photograph as the specific area of interest for the observer or the lightning conditions. However, other parameters cannot be easily derived directly from the photograph as the distance from the camera to the missing person or the focal length, which will determine what is finally projected into the picture.

- The skull 3D model: this skull model will have a specific orientation, resolution, and size given by the technical features of the considered scanner as well as by the skull modeling process. Notice that, the skull model size usually corresponds to the size of the real skull.

Hence, a 3D-2D IR process where all these unknown parameters have to be estimated seems to be the most appropriate formulation to automate SFO as it directly replicates the original scenario where the photograph was taken. Using the sets of cranial and facial landmarks, an automatic method will look for the geometric transformation (a translation, a rotation, a scaling, and a perspective projection) that minimizes the distances among the corresponding pairs of landmarks by properly overlaying the skull 3D model over the face in the 2D image.

2.4.2   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, $C = \{cl^1, ..., cl^n\}$ and $F = \{fl^1, ..., fl^n\}$, the process has to solve a system of equations with 12 unknowns [ICDS09]: 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 size of 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. Those parameters determine the geometric transformation f that projects every cranial landmark cl^i of the skull 3D model onto its corresponding facial landmark fl^i of the

photograph:

$$F = C \cdot R \cdot S \cdot T \cdot P \tag{I.2}$$

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. A complete description of the matrices of Eq.(I.2) is detailed in [HB97].

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 mean error (ME) fitness function [ICDS09]:

$$ME = \frac{\sum\limits_{i=1}^{N} d(f(cl^i), fl^i)}{N}, \tag{I.3}$$

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 3D cranial landmark projected on the 2D photograph by means of f, d is the 2D Euclidean distance, and N is the number of landmarks placed by the expert.

2.4.3   Modeling the Uncertainty Related to the Location of Facial Landmarks

The uncertainty related to the location of facial landmarks refers to the difficult task of precisely and invariably placing landmarks on a photograph [CGM$^+$13]. Among other reasons, the latter is caused by the fact that the definition of many anthropometric landmarks is imprecise in nature [Boo97]. The ambiguity of placing points in a photograph may also arise from additional reasons such as 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.

The fuzzy approach developed in [ICDS11] 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 increase when those landmarks are considered. This leads to a better specification 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.

Following the idea of fuzzy plane geometry [DK00] and of metric spaces [DP83], a fuzzy landmark is considered as a fuzzy convex set of points having a non-empty core and a bounded support. All its α-levels are non-empty bounded and convex sets. Dealing with 2D photographs with a $x \times y$ resolution, the facial fuzzy landmarks are defined 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 has a different size depending on the imprecision of its localizations but at least one pixel (i.e. crisp point related to a matrix cell) has a membership with degree equal to one.

These masks are built starting from two triangular fuzzy sets $\tilde{A}_x$ and $\tilde{A}_y$ modeling the approximate vertical and horizontal position of the ellipse determining the location of the landmark. Each triangular fuzzy set $\tilde{A}_p$ with $p \in \{x, y\}$ is defined by its center $c \in \{c_x, c_y\}$ and its offsets $l \in \{l_x, l_y\}$ and $r \in \{r_x, r_y\}$ as follows [KY95]:

$$\tilde{A}_p = \begin{cases} 1 - \dfrac{|p - c|}{c - l}, & \quad if \quad\quad x \leq 0 \\[3mm] 1 - \dfrac{|p - c|}{r - c}, & \quad if \quad\quad x \leq 0 \\[3mm] 0, & \quad\quad otherwise \end{cases} \tag{I.4}$$
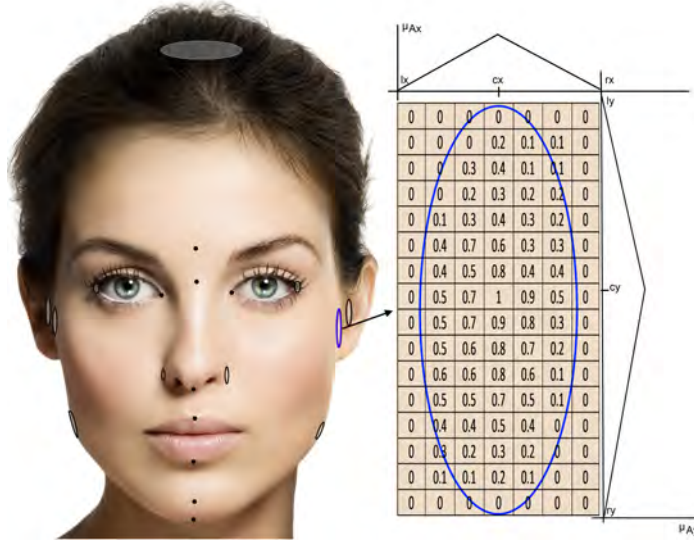
Figure 4: Imprecise location of facial landmarks (left) and representation of a fuzzy landmark with fuzzy sets (right).

In the case that $p = x$, the points would be $c = c_x$, $l = l_x$, and $r = r_x$; the fuzzy set corresponds to $\tilde{A}_x$. If $p = y$ then $c = c_y$, $l = l_y$, and $r = r_y$. Using this value, the defined fuzzy set is $\tilde{A}_y$. The 2D fuzzy set $\tilde{A}_{xy}$, which defines the ellipse, is the result of applying the product t-norm of the latter two 1D fuzzy sets $\tilde{A}_{xy} = \tilde{A}_x \cdot \tilde{A}_y$. Its membership function $\mu_{\tilde{A}_p}$ is determined as $\mu_{\tilde{A}_p} = \mu_{\tilde{A}_{xy}} = \mu_{\tilde{A}_x} \cdot \mu_{\tilde{A}_y}$.

As a consequence of handling those fuzzy facial landmarks, modeled as 2D fuzzy sets (see Fig. 4), the computation of the distances between the corresponding projected 3D cranial and 2D facial landmarks in Eq.(I.3) gets affected as the *distance from a point x to a fuzzy landmark* $\tilde{F}$:

$$d'(x, \tilde{F}) = \frac{\sum\limits_{k=1}^{m} d(x, \tilde{F}_k) \cdot \alpha_k}{\sum\limits_{k=1}^{m} \alpha_k}, \tag{I.5}$$

where x is a precise point, $\tilde{F}_k$ is the k_{th} element (pixel) of the fuzzy landmark $\tilde{F}$, d is the 2D Euclidean distance, α_k is the membership value of $\tilde{F}_k$, and m is the number of elements of the fuzzy landmark $\tilde{F}$.

The interested reader is referred to [ICDS11] for a deeper explanation and an example of the calculation of the distance between a point and a fuzzy landmark.

Considering the latter distance (Eq.(I.5)), the definition of the EA fitness function (Eq.(I.3)) is modified as follows [ICDS11]:

$$fuzzyME = \frac{\sum\limits_{i=1}^{N} d'(f(cl^i), \tilde{F}^i)}{N}, \tag{I.6}$$

where $f(cl^i)$ represents the 2D position of the 3D cranial landmark when projected on the photograph, $\tilde{F}^i$ represents the fuzzy landmark, and $d'(f(cl^i), \tilde{F}^i)$ is the distance between a point and a fuzzy landmark (Eq.(I.5)).

2.4.4 Ground Truth Data Generation for Skull-Face Overlay

The success of the final identification strongly relies on an accurate superimposition, since this is the previous step to analyze the anatomical correspondence between the face and the skull. Thus, reaching an accurate overlay is of paramount importance before proceeding with the final decision making stage.

The ground truth dataset created in [ICCÁ+15] allows us to measure and compare the performance of both manual and automatic SFO methods following an objective and reliable procedure to assess the SFO result achieved. With such a ground truth dataset a new horizon is opened for the development of novel automatic methods whose performance could be now objectively measured and compared against previous and future proposals.

In order to achieve a number of ground truth SFOs, frontal and lateral photographs were taken from living patients whose head had just been scanned with a CBCT. The DICOM images resulting from the CBCT machine were automatically processed to obtain the corresponding facial and skull 3D models. After positioning homologous points in both the facial 3D model and the photograph, the former was automatically projected onto the latter so they perfectly match. Then, the registration transformation originating that perfect match between the facial 3D model and the photograph was applied to the skull 3D model resulting in a perfect SFO. The latter should be the ground truth projection of the skull over the face photograph. Thus, for each case the 2D location (x and y pixels) of some landmarks are recorded to compare with. Figure 5 graphically shows an overview of the whole ground truth data creation process.
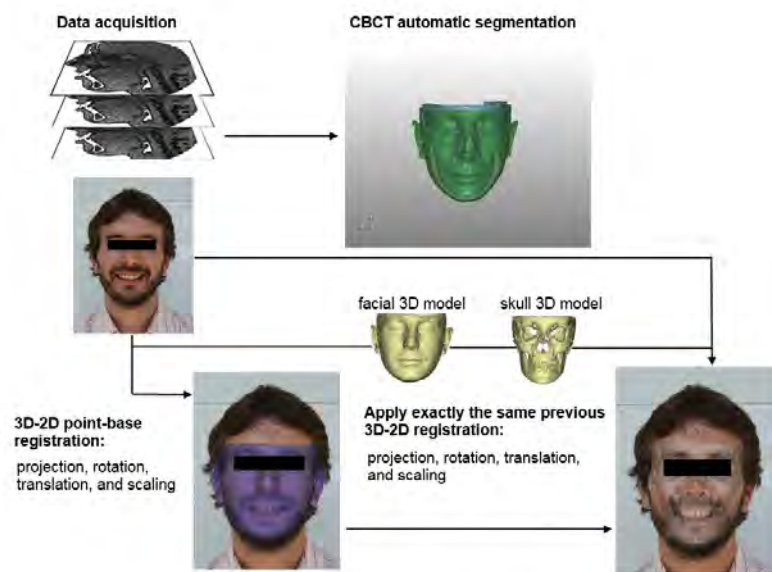
Figure 5: Overview of the ground truth data creation process proposed in [ICCÁ+15].

The facial 3D model enables to superimpose a face (3D model) over the same face (photograph), i.e. superimpose the same 'object' acquired by different sensors. Thus, the problem of superimposing the facial 3D model over the facial photograph is modeled following an IR approach based on the same 3D-2D IR approach as in the case of the SFO process explained in Sec. 2.4. Using two sets of homologous points located in both the facial 3D model and the photograph, the automatic method introduced in Secs. 2.4.2 and 2.4.3 will look for the geometric transformation that minimizes the distances among the corresponding landmark pairs by properly overlaying the skull 3D model over the face in the 2D image. The Euclidean distance in pixels is the measurement used in this algorithm.

Besides the error in pixels, the authors included an additional estimation of the total error in mm. By back-projecting the facial points located in the photograph, they calculate a back-projection line for a given geometric transformation f.

Once the facial 3D model is "perfectly" superimposed over the photograph of the face, the same geometric transformation is applied over the skull 3D model to perfectly superimpose it in the same manner. The geometric transformation, the whole picture of the skull projected in the 2D image or the location of a set of cranial landmarks can be used as ground truth data for comparison and method validation purposes.

3 Development

This section introduces the methods proposed in this PhD dissertation conforming together with some previous developments the first automatic CFS system. The first subsection presents a general overview of the obtained developments while the next two subsections describe each specific contribution within the respective CFS stage.

3.1 Overview

The current PhD dissertation aims to extend and improve the functionality of the automatic CFS method. The main objective is to provide a complete, reliable and robust computer-aided system for CFS. This work focuses on two differentiated goals. On the one hand, to incorporate the modeling of the imprecision related to the thickness between the cranial and facial landmarks to the existing automatic SFO method. On the other hand, to design and develop an automatic DSS studying the skull-face correspondence and taking into account the sources of information and uncertainty involved in the process. Thus, in this dissertation we focus on the second and the third stages of the CFS procedure. The obtained developments can also be classified attending to the stage of the CFS procedure that they belong to. Fig. 6 shows this classification.
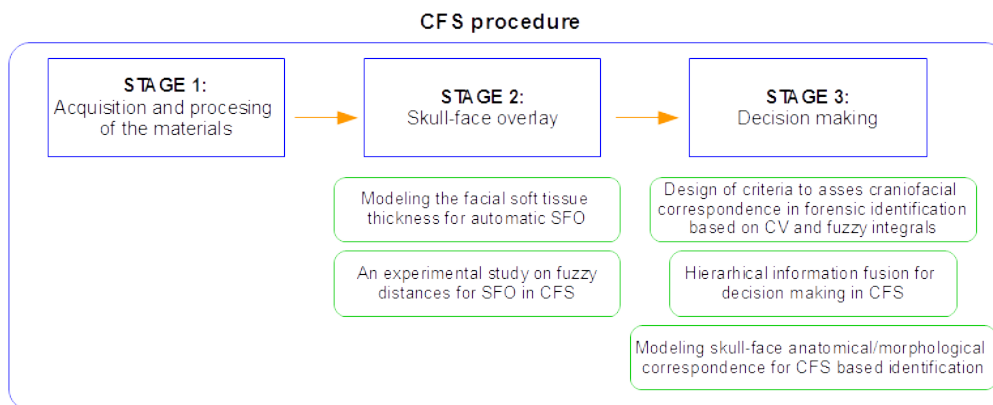
Figure 6: The extended system is composed of several methods embedded into the second and third stages of the CFS procedure.

Within the second stage, we have performed an analysis of the use of CV and SC to properly model the SFO problem. Then, we have introduced their use for automatically solving the SFO procedure dealing with the cranial-facial *landmark matching uncertainty*, which we define as the imprecision in the matching of two sets of landmarks that belong to two different objects, a face and a skull. Fuzzy sets will be considered in order to manage this uncertainty. Evolutionary algorithms will be used to automatically obtain the best fit between the found skull and the facial photograph.

Hence, we have designed a new automatic SFO procedure, based on EAs and fuzzy sets, which jointly tackles the landmark location and cranial-facial *landmark matching uncertainty*.

As said, this proposed method deal with the inherent uncertainty due to the presence of soft tissues in the face by considering fuzzy sets. Accordingly, our methodology requires computing two kinds of distance metrics: between a point and a fuzzy set, and between two fuzzy sets. The choice of a good distance metric can be crucial to the performance of our method. Thus, we aim to study the behavior and influence of the most significant and suitable fuzzy distances presented in the specialized literature, as well as some new ones proposed, on our SFO system.

Once the skull-face overlays are properly obtained, the next step is the decision making. In the literature, there is no previous work tackling the automation of this third stage. This is a subjective process that relies on the forensic expert's skills and the quantity and the quality of the used materials. Hence, we will design a complete framework for a DSS to make a decision based on the fusion of the available information sources in an objective way. It also leads to the application of CFS in identification scenarios with multiple comparisons, a possibility not yet explored due to the unaffordable time lapse needed to analyze all possible cross-comparisons. We will implement and test the DSS for CFS modeling a significant number of anatomical regions of the skull and their morphological correspondence with the face.

The approaches related to the second stage are introduced in Section 3.2. Then, the specific design method for the third stage is deeply explained and tested in Section 3.3.

3.2 Stage 2: Skull-Face Overlay

The following subsections describe the two proposals regarding to the stage 2 of the CFS. The next two include the corresponding experiments and their results.

3.2.1 Modeling the Landmark Matching Uncertainty

Previous works tackle SFO in an automatic way using EAs and fuzzy sets [ICDS09, ICDS11, ICD12]. Those 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 [CGM$^+$13]. That minimization process involves the search for the specific projection of the skull 3D model that reduces all the point-to-point distances between every pair of corresponding landmarks as much as possible. That is a good approximation to face the problem which provided reasonable results but it is not anatomically correct. In fact, the anatomical distance between a cranial landmark and its corresponding facial point (soft tissue depth) is not considered. In reality, the thickness of the facial soft tissue differs for each corresponding pair of landmarks, varies among individuals, and produces a mismatch among landmarks of the skull and the face [Ste09, GS12].

In this study, we present an automatic 3D-2D SFO method which considers the imprecision related to the matching of landmarks in the skull and face. We have modeled the soft tissue thickness between pairs of cranial and facial landmarks using fuzzy sets theory [Zad96]. This novel proposal uses the same optimization mechanism employed before [ICDS11, ICD12] but it changes the formulation of the problem to directly incorporate the spatial relationships between corresponding landmarks. Unlike traditional SFO approaches locating tissue depth markers on the physical or the 3D model skull [BKR$^+$10, GS12] (see Fig. 7), our proposal allows us to incorporate any soft tissue study easily. To our knowledge, the current proposal is the first automatic SFO procedure that considers forensic studies of inter-landmark distances.

Our first approach consists of building a fuzzy set whose center is the 3D cranial landmark with membership degree of zero. The rest of the points are calculated adding the positive and negative values of the *min*, *mean*, and *max* distances to the 3D cranial landmark position along the three main directions (X, Y, and Z). We thus define a 3D volume (sphere) in the space where each

Figure 7: Mean soft tissue depths (in mm) and spatial relationship between cranial and facial landmarks [SS08a, SS08b].

facial landmark is expected to be located according to a particular soft tissue depth study (Figs. 7 and 9a).
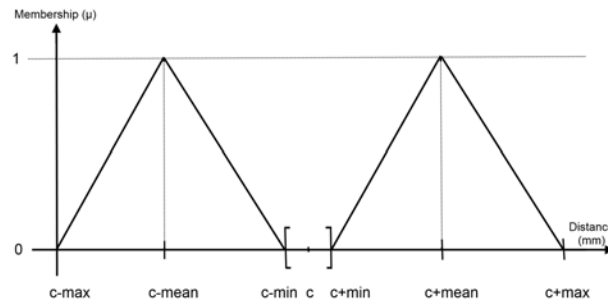
Figure 8: Graphical representation of one of the dimensions of the 3D fuzzy set $\tilde{B}_p$ for modeling the *landmark matching uncertainty* using a sphere. The other two dimensions are modeled in a homologous way.

Formally, we handle the *landmark matching uncertainty* using 3D masks represented as a matrix M with $m_x \times m_y \times m_z$ points. These masks are defined by three triangular fuzzy sets $\tilde{B}_x$, $\tilde{B}_y$, and $\tilde{B}_z$, which model the approximate vertical, horizontal, and depth position of the sphere that represents the place of each facial landmark in relation with its cranial counterpart. They thus become 3D fuzzy sets, where each triangular fuzzy set $\tilde{B}_p$, with $p \in \{x, y, z\}$, is defined by its center $c \in \{c_x, c_y, c_z\}$ (the 3D coordinates of the cranial landmark) and the *min*, *mean*, and *max* distances (see Fig. 8).

In the case that $p = x$, the point would be $c = c_x$. Hence, the calculated fuzzy set corresponds to $\tilde{B}_x$. If $p = y$ then $c = c_y$. Using this value, the defined fuzzy set is $\tilde{B}_y$. Finally, if $p = z$ implies that $c = c_z$. The resulting fuzzy set is $\tilde{B}_z$. The 3D fuzzy set $\tilde{B}_{xyz}$, which defines the whole sphere, is the result of applying the product t-norm of the latter three 1D fuzzy sets $\tilde{B}_{xyz} = \tilde{B}_x \cdot \tilde{B}_y \cdot \tilde{B}_z$.

Its membership function is determined as $\mu_{\tilde{B}_{xyz}} = \mu_{\tilde{B}_x} \cdot \mu_{\tilde{B}_y} \cdot \mu_{\tilde{B}_z}$.

$$
\tilde{B}_p = \begin{cases}
1 - \dfrac{|p - c + mean|}{max - mean}, & if & c - max \leq p < c - mean \\[2ex]
1 - \dfrac{|p - c + mean|}{mean - min}, & if & c - mean \leq p \leq c - min \\[2ex]
1 - \dfrac{|p - c - mean|}{mean - min}, & if & c + min \leq p \leq c + mean \\[2ex]
1 - \dfrac{|p - c - mean|}{max - mean}, & if & c + mean < p \leq c + max \\[2ex]
0, & otherwise
\end{cases} \tag{I.7}
$$
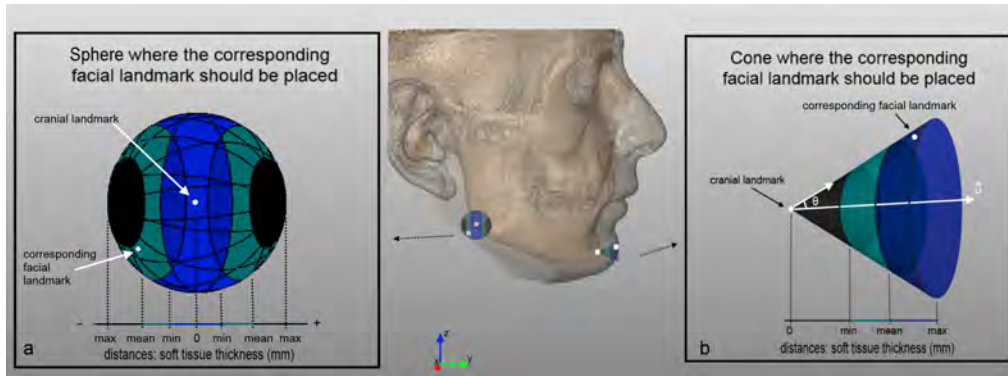
Figure 9: Facial landmark position from a cranial landmark using a sphere (a) and a cone (b). *Min*, *mean*, and *max* are the soft tissue depths. For the cone, $\vec{u}$ is the normal vector at the cranial landmark in the skull, θ is the rotation angle of $\vec{u}$.
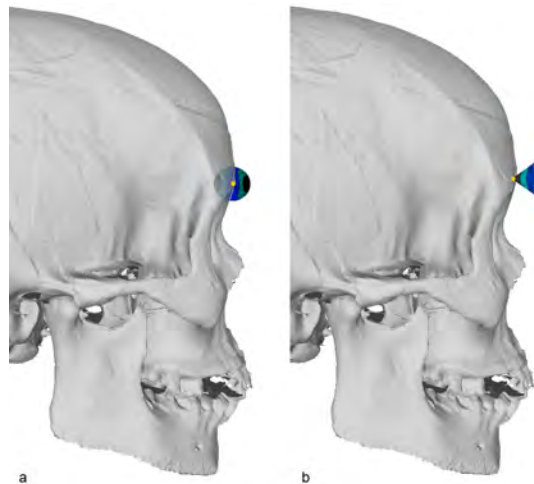

Figure 10: Facial landmark position from a cranial landmark using a sphere (a) and a cone (b).

This latter approach is a good approximation to tackle the SFO problem modeling the facial soft tissue thickness with the available tissue depth studies. However, it allows inconsistent anatomical solutions to be obtained because it does not model the positional relationships between

pairs of landmarks. For example, facial landmarks located inside the skull are considered to be correctly positioned (see the upper right part of the sphere in Fig. 9 and the whole left part of the sphere in Fig. 10).

A variation of the sphere-modeling method is proposed as follows. Instead of representing the soft tissue depth using a sphere with the cranial landmark in the center, we define a cone whose vertex is the cranial landmark. Each facial landmark is expected to be located inside the cone (Fig. 9b). Using a cone, we closely specify a narrower region where the facial landmark should be located. We refine the search of the facial landmarks placement to a specific region.

We assume a certain degree of perpendicularity between cranial and facial landmarks as most soft tissue studies do [SS08a, SS08b]. To do so, we consider the normal vector to the surface of the skull 3D model at each cranial landmark: $\vec{v} = (x_n, y_n, z_n)$. The unit vector of $\vec{v}$, which has its same direction but a magnitude of the unit, has been determined below:

$$\vec{u} = (\frac{x_n}{\|\vec{v}\|}, \frac{y_n}{\|\vec{v}\|}, \frac{z_n}{\|\vec{v}\|}) = (u_x, u_y, u_z), \tag{I.8}$$

where $\|\vec{v}\| = \sqrt{(x_n^2 + y_n^2 + z_n^2)}$ is the magnitude of $\vec{v}$.

In order to estimate the position of the facial landmarks, the $\vec{u}$ coordinates (u_x, u_y, u_z) are multiplied by the specific distance (*min*, *mean*, or *max*).

Since the correspondence between a pair of cranial-facial landmarks is not always perpendicular, different inclination angles are applied to the unit vector $\vec{u}$ in order to define the volume in which the facial landmark will be likely located. The amplitude of this area can be defined by a rotation equal to $\pm\theta$ along the three axes X, Y, and Z.

The 3D rotation of the unit vector $\vec{u}$ consists of three different rotations, i.e. a rotation of an angle $\pm\theta$ along the X, Y, and Z axes [HB97].

Likewise, we define a fuzzy set whose center is the 3D cranial landmark with membership degree of zero. The rest of the points are calculated multiplying the coordinates of the unit vector $\vec{u}$ by the values of the different distances *min*, *mean*, and *max*. Thus, a 3D cone is defined where each facial landmark is expected to be located according to a particular soft tissue depth study (Figs. 7 and 9b). The *landmark matching uncertainty* is defined using 3D masks in the same fashion as the spheres. Hence, a fuzzy set $\tilde{B}_p$ with $p \in \{x, y, z\}$ (see Fig. 11 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:

$$\tilde{B}_p = \begin{cases} 1 - \dfrac{|p - c + u \cdot mean|}{u(max - mean)}, & if & c - u \cdot max \leq p \leq c - u \cdot mean \\[2ex] 1 - \dfrac{|p - c + u \cdot mean|}{u(mean - min)}, & if & c - u \cdot mean \leq p \leq c - u \cdot min \\[2ex] 1 - \dfrac{|p - c - u \cdot mean|}{u(mean - min)}, & if & c + u \cdot min \leq p \leq c + u \cdot mean \\[2ex] 1 - \dfrac{|p - c - u \cdot mean|}{u(max - mean)}, & if & c + u \cdot mean < p \leq c + u \cdot max \\[2ex] 0, & otherwise \end{cases} \tag{I.9}$$

In the case that $p = x$, the points would be $c = c_x$ and $u = u_x$. Hence, the calculated fuzzy set corresponds to $\tilde{B}_x$. If $p = y$ then $c = c_y$ and $u = u_y$. Using these values, the defined fuzzy set is $\tilde{B}_y$. If $p = z$ it implies that $c = c_z$ and $u = u_z$, with the resulting fuzzy set being $\tilde{B}_z$.

The 3D fuzzy set $\tilde{B}_{xyz}$, which defines the whole 3D cone, is the product t-norm of the latter three 1D fuzzy sets $\tilde{B}_{xyz} = \tilde{B}_x \cdot \tilde{B}_y \cdot \tilde{B}_z$. Its membership function is determined as the product t-norm of the membership functions of each fuzzy set: $\mu_{\tilde{B}_{xyz}} = \mu_{\tilde{B}_x} \cdot \mu_{\tilde{B}_y} \cdot \mu_{\tilde{B}_z}$.
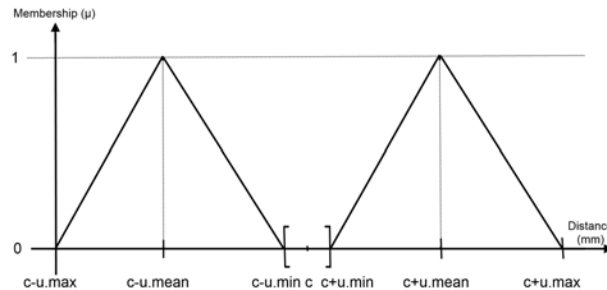

Figure 11: Graphical representation of one of the dimensions of the 3D fuzzy set $\tilde{B}_p$ for modeling the *landmark matching uncertainty* with a cone. The other two dimensions are modeled in a homologous way.

Our proposal allows experts to mark both precise and imprecise (fuzzy) facial landmarks. The latter landmarks can be used by forensic anthropologists as necessary.

3.2.2 Analysis of the Best Fuzzy Distance Metric for Landmark Matching

With the modeling of the *landmark matching uncertainty* proposed, the need to compute the distance between two fuzzy sets arises. One of these two fuzzy sets is the projection of the 3D cranial landmark on the facial photograph, which is composed of the precise 3D cranial landmark (called cl^i in Sec. 2.4.2) and the fuzzy set that models the *landmark matching uncertainty*. The other fuzzy set would be the fuzzy facial landmark of the photograph representing the imprecise position of a facial landmark (Fig. 4). 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 the 18 skull-face overlay instances from a ground truth dataset described in Section 2.4.4 [ICCÁ+15].

The approach introduced in Section 3.2.1 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 3.2.1. 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 kinds of distance measures arises: between a point and a fuzzy set and between two fuzzy sets.

To illustrate the latter concept, Figure 12 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. Once the facial-cranial distances (cones) are projected in the 2D image, we obtain a fuzzy set of
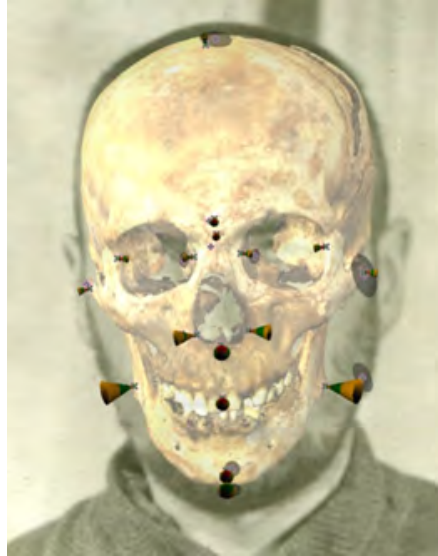
Figure 12: 3D cranial fuzzy landmarks (cones) projected onto a facial photograph with fuzzy landmarks (ellipses).

2D points. Now, we need to measure the spatial distance between the corresponding pairs of fuzzy sets.

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 have reviewed the existing fuzzy distance measures in the literature and their features. Using that information, we will 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 [Fre75, Dut91]. A review of several definitions of fuzzy distances is presented in [Blo99], 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 the 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. For this particular study we have chosen the kind of distances between fuzzy sets and between a point and a fuzzy set that always use the Euclidean distance 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.

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:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \tag{I.10}$$

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 $d''(\tilde{A}, \tilde{B})$.

The definitions of the proposed distances for our study between a point x and a fuzzy set $\tilde{B}$ are the following:

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 [Blo99]:

$$d_1'(x, \tilde{B}) = \frac{\sum_y d(x, y) \cdot \tilde{B}(y)}{\sum_y \tilde{B}(y)} \tag{I.11}$$

2. **Nearest point distance:** it calculates the distance to the nearest point taking into account the membership degrees [Blo96]:

$$d_2'(x, \tilde{B}) = \min_y \left\{ \frac{d(x, y)}{\tilde{B}(y)} \right\} \tag{I.12}$$

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:

$$d_4'(x, \tilde{B}) = \min_y \left\{ \frac{d(x, y) + 1}{\tilde{B}(y)} \right\} \tag{I.13}$$

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.

$$d_5'(x, \tilde{B}) = \begin{cases} d_4'(x, \tilde{B}) & \text{if x} \notin \tilde{B} \\ 1 - \tilde{B}(x) & \text{if x} \in \tilde{B} \end{cases} \tag{I.14}$$

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 [Blo96, GV86].

$$d_3'(x, \tilde{B}) = \sum_{i=1}^{m} d_H(x, \tilde{B}_\alpha) \cdot (\alpha_i - \alpha_{i-1}) \tag{I.15}$$

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

$$d_H(x, B) = \inf_{y \in B} d(x, y) \tag{I.16}$$

Note that denominators in Eqs. I.11, I.12, and I.13 cannot be zero. In our case this situation never happens due to the fact that we do not consider the points with membership degrees equal to zero as being part of the fuzzy set.

Distances between two fuzzy sets are the most widely addressed in the literature. These distances are 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 [Ros85]. 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.

We extend these previous definitions to model the corresponding distance between fuzzy sets as follows:

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 [Blo99, Ros85].

$$d''_1(\tilde{A}, \tilde{B}) = \frac{\sum_x \sum_y d(x,y) t[\tilde{A}(x), \tilde{B}(y)]}{\sum_x \sum_y t[\tilde{A}(x), \tilde{B}(y)]} \tag{I.17}$$

   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 (I.12).

$$d''_2(\tilde{A}, \tilde{B}) = \min_{(x,y)} \left\{ \frac{d(x,y)}{\tilde{A}(x) \cdot \tilde{B}(y)} \right\} \tag{I.18}$$

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 (I.13).

$$d''_4(\tilde{A}, \tilde{B}) = \min_{(x,y)} \left\{ \frac{d(x,y) + 1}{\tilde{A}(x) \cdot \tilde{B}(y)} \right\} \tag{I.19}$$

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 (I.14). It is based on the dissimilarity measure between two fuzzy sets taking into account only the intersection [ZCB87, Blo99].

$$d''_5(\tilde{A}, \tilde{B}) = \begin{cases} d''_4(\tilde{A}, \tilde{B}) & \text{if } \tilde{A} \cap \tilde{B} = \emptyset \\ 1 - \max\left\{ t[\tilde{A}(y), \tilde{B}(y)] \right\} & \text{if } \tilde{A} \cap \tilde{B} \neq \emptyset \end{cases} \tag{I.20}$$

   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 [Blo99, GV86].

$$d''_3(\tilde{A}, \tilde{B}) = \sum_{i=1}^m d(\tilde{A}_\alpha, \tilde{B}_\alpha).(\alpha_i - \alpha_{i-1}) \tag{I.21}$$

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 [Ros85] using the extension principle:

$$\tilde{D}(\tilde{A}, \tilde{B})(r) = \sup_{\substack{(x,y) \\ d(x,y)=r}} \inf\left[\tilde{A}(x), \tilde{B}(y)\right] \tag{I.22}$$

It was demonstrated in [Blo99] that the choice of the t-norm, among the most common ones (minimum, 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. I.17) and for the nearest point extension 2 distance (Eq. I.20).

As before, denominators in Eqs. I.17, I.18, and I.19 cannot be zero. In our case this situation never happens.

The fitness function *Fuzzy Mean Error* (FME) for the new automatic 3D-2D SFO task has been formulated taking into account the latter fuzzy distances (from 1 to 5):

$$FME = \frac{\sum_{i=1}^{Ncrisp}(d'(x_i, f(\tilde{C}^i)) + \sum_{j=1}^{Nfuzzy}(d''(\tilde{F}^j, f(\tilde{C}^j))}{N}, \tag{I.23}$$

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 landmarks considered ($N = Ncrisp + Nfuzzy$), 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. I.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) 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.

However, in the case of using fuzzy distances as fuzzy numbers (Rosenfeld distance, Eq. I.22), 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.

We use the α-cuts-based method for our purpose due to its lower computational complexity, an interesting requirement for an iterative process as ours. It defines a fuzzy set in the α-cut, $(\tilde{A} \star \tilde{B})_\alpha$, as

$$(\tilde{A} \star \tilde{B})_\alpha = \tilde{A}_\alpha \star \tilde{B}_\alpha, \tag{I.24}$$

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:

$$\mu \star \nu = \bigcup_{\alpha \in (0,1]} (\tilde{A} \star \tilde{B})_\alpha, \tag{I.25}$$

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.

Therefore, for these cases we call the average distance as *fuzzy mean distance* (FMD), and we denote it by $\tilde{D}_{FMD}$:

$$\tilde{D}_{FMD} = \frac{\bigcup_{\alpha \in (0,1]} \sum\limits_{i=1}^{N} \tilde{D}(\tilde{F}^i, f(\tilde{C}^i))_\alpha}{N} \tag{I.26}$$

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 [LK99]. 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:

$$FME = \frac{\sum\limits_{\alpha_{min}}^{\alpha_{max}} \alpha \cdot \tilde{D}_{FMD}(\alpha)}{\sum\limits_{\alpha_{min}}^{\alpha_{max}} \tilde{D}_{FMD}(\alpha)} \tag{I.27}$$

3.2.3 Experiments I: Modeling the Landmark Matching Uncertainty

The experimental design involves 18 SFO problem instances corresponding to nine cases of living people (from Spain and Italy). That will allow the study of the ground truth data and the objective evaluation of our automatic SFO method following an unbiased and reliable procedure. These 18 SFO instances correspond to the dataset previously built using the methodology presented in Section 2.4.4 [ICCÁ+15].

The skull 3D models and the facial photographs were stored using the Face2Skull^{TM} software. All the experiments have been performed on an Intel Corel^{TM} 2 Quad CPU Q8400 2.66 GHz, with 4GB RAM, running Windows 7Professional^{TM}.

The experimental setup is composed by two parts: the first one addresses to study the performance of the matching uncertainty proposal and select the best approach, and the second one is aimed to analyze the most appropriate function for this method.

We will compare our proposal against the state-of-the-art automatic approaches RCGA and CCGA-2 (see Section 2.4.2) [ICDS11, ICD12]. The parameter configuration used for testing the algorithm is that which obtained the best results in [ICDS09] and [ICD12] respectively. These proposals calculate the distance between a cranial and its corresponding facial landmark by means of the fitness function defined in Eq. (I.23) and the distance function that achieves the best results in the previous section, i.e., the weighted mean (Eqs. I.11 and I.17). Since these methods are based on stochastic processes, 30 independent runs were performed for each problem instance to compare the robustness of the methods and to avoid any possible bias. The EA with the best performance was chosen for incorporating the facial soft tissue modeling in it.

Table I.1 presents the mean error (in mm) 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 per algorithm. The distance error has been calculated for all the landmarks marked in each case study. 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 of a particular algorithm.

As a result of the comparison between RCGA and CCGA-2, RCGA achieved the best behavior in 13 of the 18 cases (Table I.1) with significant differences in most of them. Thus, we selected RCGA to incorporate the facial soft tissue modeling. We called the new proposals as RCGA-s and RCGA-c. RCGA-s corresponds to the approach that deals with the *landmark*

Table I.1: Mean error in mm 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

Case, pose	pl	il	CCGA	RCGA	RCGA-s	RCGA-c-0	RCGA-c-10	RCGA-c-30	RCGA-c-45
1,f	7	7	4.565	2.750	**2.727**	4.116	4.330	3.078	3.164
1,l	5	4	16.588	7.406	6.938	7.400	6.693	**5.788**	5.828
2,f	8	5	4.906	3.690	3.486	3.823	4.212	3.337	**3.331**
2,l	3	2	8.299	8.605	8.871	8.871	8.871	5.286	**3.688**
3,f	8	7	3.815	3.629	3.664	6.594	6.594	3.007	**2.941**
3,l	4	4	9.367	10.643	10.815	11.073	11.073	7.604	**7.001**
4,f	7	6	4.635	3.647	3.505	3.356	3.356	3.223	**3.045**
4,l	4	3	14.027	14.327	13.494	13.847	13.847	12.446	**12.171**
5,f	10	6	2.996	2.436	2.453	**2.204**	2.406	2.641	2.600
5,l	5	4	8.545	6.865	6.097	6.846	6.846	3.715	**2.825**
6,f	8	7	4.654	3.784	3.644	**2.618**	3.076	2.878	2.921
6,l	3	4	16.486	12.959	11.665	**2.204**	11.126	11.009	10.626
7,f	10	5	4.253	3.639	**3.617**	4.695	4.695	4.100	3.714
7,l	5	4	10.212	12.663	10.909	**9.649**	9.649	9.729	10.345
8,f	9	6	5.269	4.409	3.579	3.492	3.492	2.856	**2.882**
8,l	4	4	6.840	8.555	7.812	19.935	19.935	6.262	**5.412**
9,f	10	4	6.174	5.541	5.757	6.660	6.660	5.071	**4.887**
9,l	3	5	10.210	11.701	11.700	19.852	19.852	10.391	**9.509**
Average (mm)			7.778	6.976	6.548	7.759	8.416	5.585	**5.281**

matching uncertainty using spheres (see Sec. 3.2.1). In the case of RCGA-c, the *landmark matching uncertainty* is modeled with cones (see Sec. 3.2.1). Within RCGA-c, four different angles have been tested to define the cone amplitude: $\theta = \pm 0°$, $\theta = \pm 10°$, $\theta = \pm 30°$, and $\theta = \pm 45°$, so we have analyzed four variants of this proposal: RCCA-c-0, RCGA-c-10, RCGA-c-30, and RCGA-c-45. They both use the fitness function defined in Eq. (I.23). The experiment again consisted of 30 independent runs for each problem instance.

In view of the results reported in Table I.1, the best performance regarding the ground truth is obtained by RCGA-c-45 in most of the cases. In general, our new approaches incorporating the *landmark matching uncertainty* treatment clearly outperform the previous proposals not considering it. In frontal view, RCGA-c-0 obtains the more accurate results for cases 5 and 6, RCGA-s for case 1, and RCGA-c-45 for the rest of the cases. The mean error of best cases ranges from 2.204 to 4.887 mm. In lateral view, the more precise SFOs correspond to RCGA-c-45 and RCGA-c-0 for cases 4 and 5. RCGA-c-45 obtains the best results in all instances except for cases 1 (RCGA-c-30) and 7 (RCGA-c-0 and RCGA-c-10). Best mean errors are higher than in frontal view, they achieve values between 2.204 and 9.649 mm (Table I.1).

RCGA-c-45 and RCGA-c-30 achieve a total average error of 5.281 and 5.585 mm, respectively. RCGA-s and RCGA obtain more than 6 mm of total average error: 6.548 and 6.976 mm, and the worst algorithm is RCGA-c-10 with 8.416 mm (Table I.1).

Figure 13a, c shows the best superimposition obtained for the two photographs of the fifth case as a particular illustration of the methods' performance. Those outcomes have been achieved by RCGA-c-0 and RCGA-c-45 for the frontal and the lateral image, respectively. In particular, blue points refer to the cranial landmarks after overlaying the skull 3D model on the photograph.
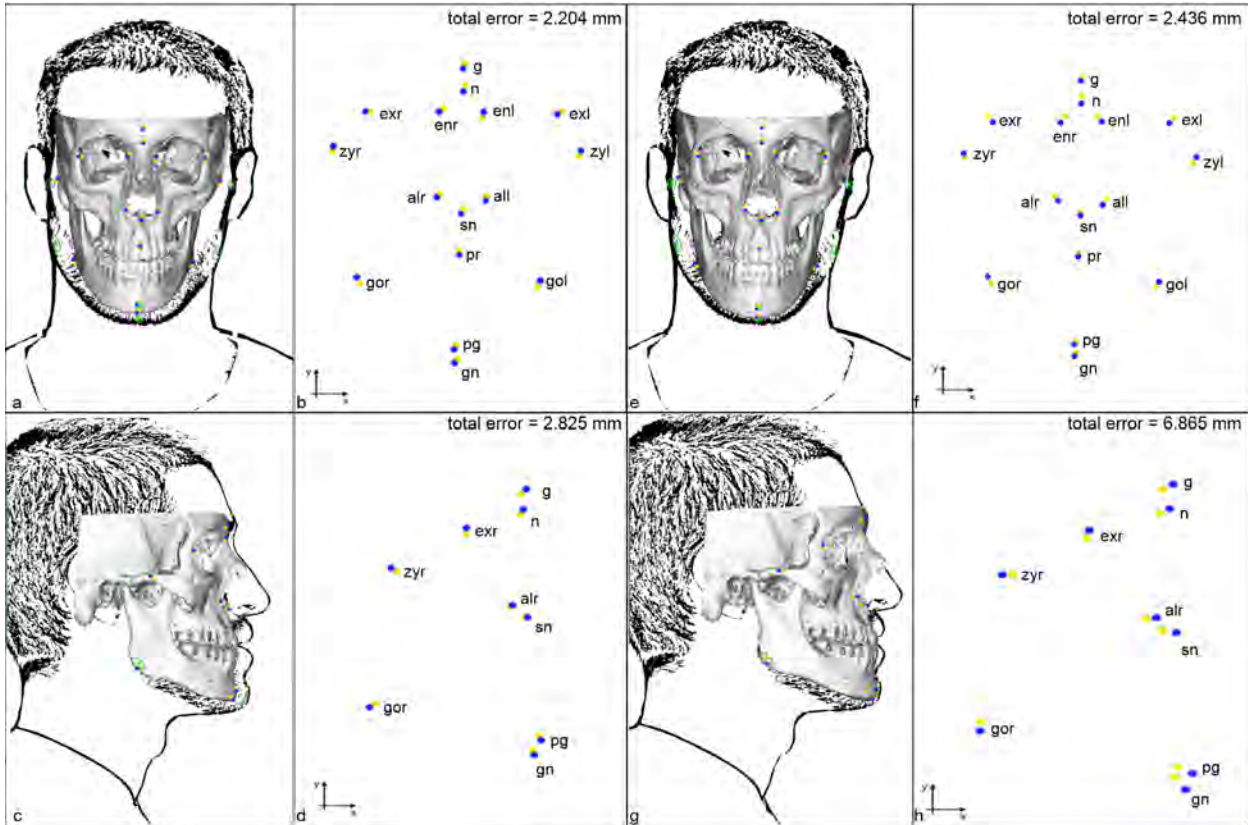
Figure 13: Case 5. Best (a, c) and RCGA (e, g) automatic SFOs for frontal and lateral poses. Best (b, d) and RCGA (f, h) resulting matching between pairs of projected cranial and actual ground truth landmarks for frontal and lateral views. Frontal images: glabella (g), nasion (n), exocanthion left, right (exl, exr), endocanthion left, right (enl, enr), alare left, right (all, alr), subnasale (sn), and prosthion (pr) were marked as precise facial landmarks. Zygion left, right (zyl, zyr), gonion left, right (gol, gor), pogonion (pg), and gnathion (gn) were placed as imprecise facial landmarks. Lateral images: Glabella (g), nasion (n), exocanthion right (exr), alare right (alr), and subnasale (sn) were placed in the photograph as a precise points. Zygion right (zyr), pogonion (pg), and gnathion (gn) were marked as imprecise facial landmarks. The homologous cranial landmarks were placed in the skull as precise points.

Yellow points are the actual landmarks achieved by the ground truth geometric transformation *g*. Green points 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. For better visualization, the resulting matching between pairs of cranial and ground truth landmarks for case 5 (frontal and lateral view) are depicted in Fig. 13b, d.

RCGA-c-0 obtains the most accurate solution regarding the ground truth in the frontal image, with a total average error of 2.204 mm (Table I.1). The two gonions present a larger distance with respect to their counterpart ground truth points (Fig. 13a, b). In the case of the lateral pose, the best overlay has been achieved by RCGA-c-45 with an error equal to 2.825 mm (Table I.1). Alare right and subnasale are the closest points to their corresponding ground truth landmarks (Fig. 13c, d).

Figure 13e and g presents the SFOs in frontal and lateral view for case 5 achieved by the best algorithm without incorporating the modeling of the *landmark matching uncertainty* (RCGA). The
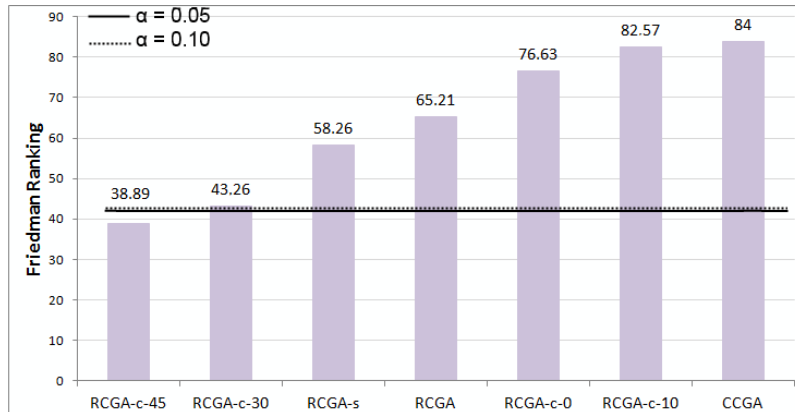
Figure 14: Friedman ranking and Bonferroni-Dunn lines for classification rate.

Table I.2: P values for the comparison between the control and the rest of approaches

Approach	Unadjusted p	p Bonf	p Holm
(RCGA-c-45 is the control)			
CCGA	**<0.0001**	**<0.0001**	**<0.0001**
RCGA	**0.00957**	0.20098	0.14356
RCGA-s	0.05028	1.0000	0.55313
RCGA-c-0	**0.00037**	**0.00779**	**0.00631**
RCGA-c-10	**<0.0001**	**0.00034**	**0.00034**
RCGA-c-30	0.62166	1.0000	1.0000

resulting matching between pairs of cranial and ground truth landmarks are depicted in Fig. 13f, h. RCGA obtains the larger error in the two ectocanthions, the two endocanthions, nasion, gonion right, alare, and zygion left for the frontal image (Fig. 13e, f). The mean total error of RCGA is 2.436 mm for the frontal photograph in case 5 (Table I.1). In the lateral image, RCGA presents a higher error in all landmarks comparing with RCGA-c-45 (Fig. 13d, h). We can also see that the skull is out of the face in the nasal region (Fig 13g). A manual refinement would be needed for this SFO result. The mean total error of RCGA in the lateral view is 6.865 mm (Table I.1).

In general, we obtain competitive matchings between pairs of corresponding cranial and facial landmarks in all the studied cases. As said, the best results are provided by the new proposals including the treatment of the *landmark matching uncertainty*.

We have performed a Friedman test [Fri40] to analyze whether significant differences exist among the performance of all the approaches. We have set the experiment level of significance in $\alpha = 0.05$. The statistic results of this test are a Friedman χ^2 equal to 43.9695, six degrees of freedom, and a p value of 7.496e-08. These data reveal significant differences among the behavior of the approaches with a p value < 0.0001, thus rejecting the null hypothesis. A Bonferroni-Dunn test [Dun61] is accomplished to detect significant differences among a control approach and the rest. RCGA-c-45 is the control algorithm because it outperforms the remaining methods, i.e. it obtains the lowest value in the Friedman ranking (Fig. 14). In the Bonferroni-Dunn test, we have obtained 3.093 and 2.856 as critical values using levels of significance $\alpha = 0.05$ and $\alpha = 0.10$, respectively.

Figure 14 summarizes the ranking obtained by the Friedman test. The bar height indicates the average ranking of each proposal. We have represented a line through all the graphic 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 cut line are the approaches whose performance is significantly worse than the control approach [GFLH09].

We have also applied a paired t test with a Bonferroni and a Holm correction, as well as an unadjusted p value to know the differences within approaches [Wri92].

Table I.2 details the pairwise comparisons considering RCGA-c-45 as the control approach. The p value is indicated in each comparison and we have distinguished in bold the approaches which are worse than the control considering a level of significance $\alpha = 0.05$.

Since the best performance in these experiments is obtained by RCGA-c-45, this approach is going to be used in our automatic SFO method for the following implementations.

3.2.4   Experiments II: Study on Fuzzy Distances

The experimental setup for the study of the different distances is the same than the previous experiments (Section 3.2.3). We have run the EA-based method explained in section 3.2.1 and its best performance (RCGA-c-45) using the distance definitions given in section 3.2.2.

Table I.3 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 the previous experiment, 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 I.4 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 I.3), five in frontal view and five in lateral view. With respect to the minimum fitness error (see Table I.4), it is the best in eight cases, five in frontal view and three in lateral view. The total average 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 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 can be seen in Table I.3, 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.

We also perform a Friedman test [Fri40] to analyze whether significant differences exist among the performance of the different fuzzy distances. 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

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 I.3: 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.

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 Bonferroni-Dunn test [Dun61] 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 15). 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 16).

Figures 15 and 16 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 [GFLH09].

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 [Wri92].

Table I.5 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 I.5). For the minimum fitness error, this distance is better than the remainder with the same confidence

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 I.4: 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.

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 I.5: P values for the comparison between the control and the rest of approaches for the mean error in 30 runs

level except for the mean α-cuts using symmetrical mean and Hausdorff distances (Table I.6). 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 15 and 16).

To better illustrate the different distance measures' performance, Figure 17 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.
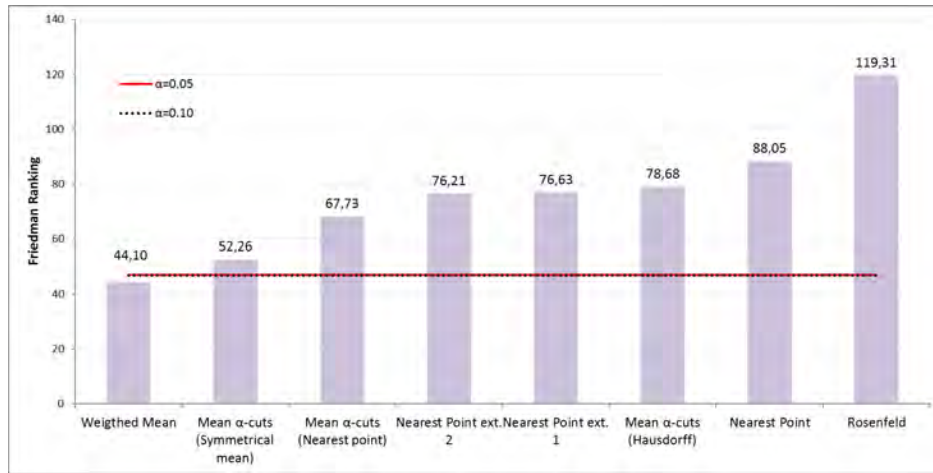
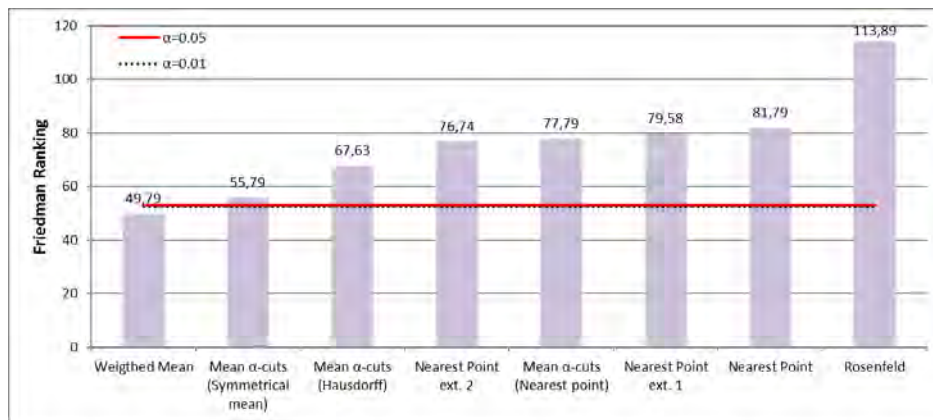Figure 15: Friedman ranking and Bonferoni-Dunn lines for classification rate for the mean error in 30 runs.

Figure 16: Friedman ranking and Bonferoni-Dunn lines for classification rate for the minimum fitness error in 30 runs.

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 I.6: P values for the comparison between the control and the rest of approaches for the minimum fitness error in 30 runs

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
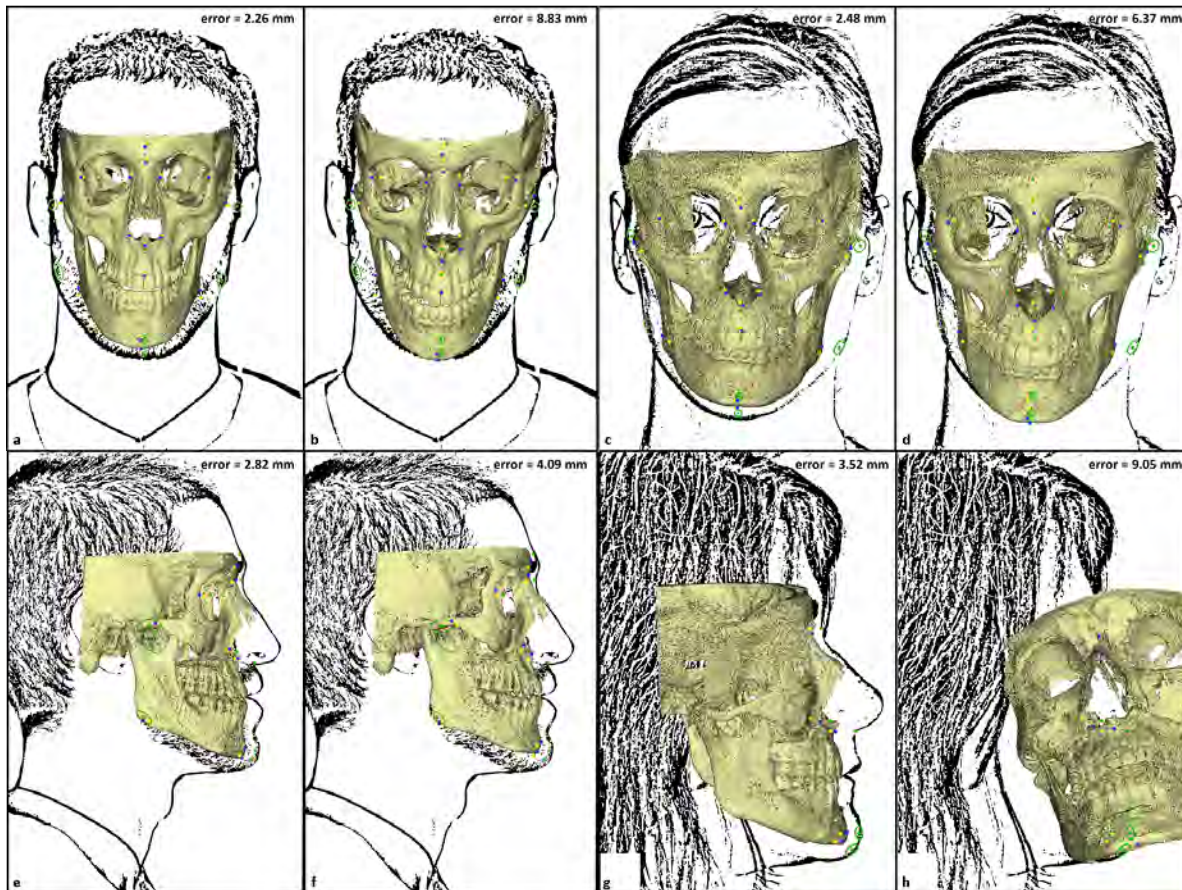
Figure 17: Visual results. Case 5, frontal view: (a) best result corresponds to the weighted mean distance, (b) worst result to the Rosenfeld distance. Case 8, frontal view: (c) best result with the weighted mean distance, (d) worst result with the Rosenfeld distance. Case 5, lateral view: (e) best result 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.

or imprecise (ellipses) landmarks. Each ellipse contains a grey point inside corresponding to its center.

Analyzing the fifth case in frontal view (Figures 17 (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 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 17 (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. 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 17 (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 17 (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 can be seen, 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. This distance operator together with the RCGA-c-45 approach to model the landmark matching uncertainty is going to be used in the following developments of this dissertation.

3.3 Stage 3: Decision Making

The performed work related to the stage 3 of the CFS process is divided in nine subsections. In the first, the design of the hierarchical DSS is described. The following three introduce the implementation of each level. In the fifth, the developments of the automatic CFS system as sort-listing and identification tool are explained. The final four subsections correspond to the experimental setup of the three performed tests and their results.

3.3.1 Design the Hierarchical Decision Support Framework for Craniofacial Superimposition

Once one or several skull-face overlays have been achieved for the same identification case the main goal is to determine the degree of support for the assertion that the skull and the face of the photograph(s) belong to the same person or not. This degree of support is based on the consistency of the matching between the face and the skull but it is also influenced by the quality and quantity of the materials used (photographs and skull). Note that the number of SFOs to be used for the identification depends on the available valid photos. The reliability of the technique relies on having more than one photo, at different poses, etc. [FHS08]. A scale for a craniofacial matching evaluation has been recently defined by some of the most representative experts in craniofacial identification within the MEPROCS project framework [DWK+15]. Accordingly, the final decision is provided in terms of strong, moderate or limited support.

This decision is guided by different criteria studying the anatomical relationship between the skull and the face. According to the literature [JSA01], we can distinguish the following families of criteria for assessing the craniofacial correspondence:

1. Analysis of the consistency of the bony and facial outlines/morphological curves. Forensic experts confirm if two particular curves (of skull and face) are anatomically consistent. That is, if two curves follow the same shape or, in other words, if one curve mirrors the other. An example of this criteria can see it in Fig. 18.a where the forehead curve of the face follows the forehead curve of the skull.

2. Assessment of the anatomical consistency by positional relationship. These criteria consist of a positional relationship analysis in order to assess anatomical consistency. Thus, the goal is to check if the relative position of a skull region against a facial region is similar in respect to anatomical reference. Fig. 18.b shows a case of this family: the consistency between the lateral angle of the eye and the cranial orbit.

3. Line location and comparison to analyze anatomical consistency. Experts analyze a set of marking lines (obtained by joining some reference landmarks) on the face and on the skull.

In terms of CV, these lines have to be parallel in an image. For example, the ectocanthion lines marked in the skull and the face (Fig. 18.c).

4. Evaluation of the consistency of the soft tissue thickness between corresponding cranial and facial landmarks. The last set of criteria involves analyzing the consistency of the facial soft tissue thickness considering distances between pairs of homologous landmarks (located on the skull and the face). Fig. 18.d shows an example of how the facial landmarks positions can be estimated from cranial landmarks using cones to model the soft tissue thickness. These distances can be checked using the skull-face overlay in existing studies relating to soft tissue thickness in different human populations [SS08a].

MEPROCS work group also discussed and quantitatively analyzed these criteria for the evaluation of the morphological skull-face correspondence, providing a set of the most discriminative and easy to assess criteria [IVC$^+$16].

Our long term goal is to automate the whole decision making process by modeling the most relevant criteria within the previous four families using CV and SC techniques. The resulting system would give as a result a global degree of support of a CFS identification to assist the forensic anthropologist to make her/his decision.

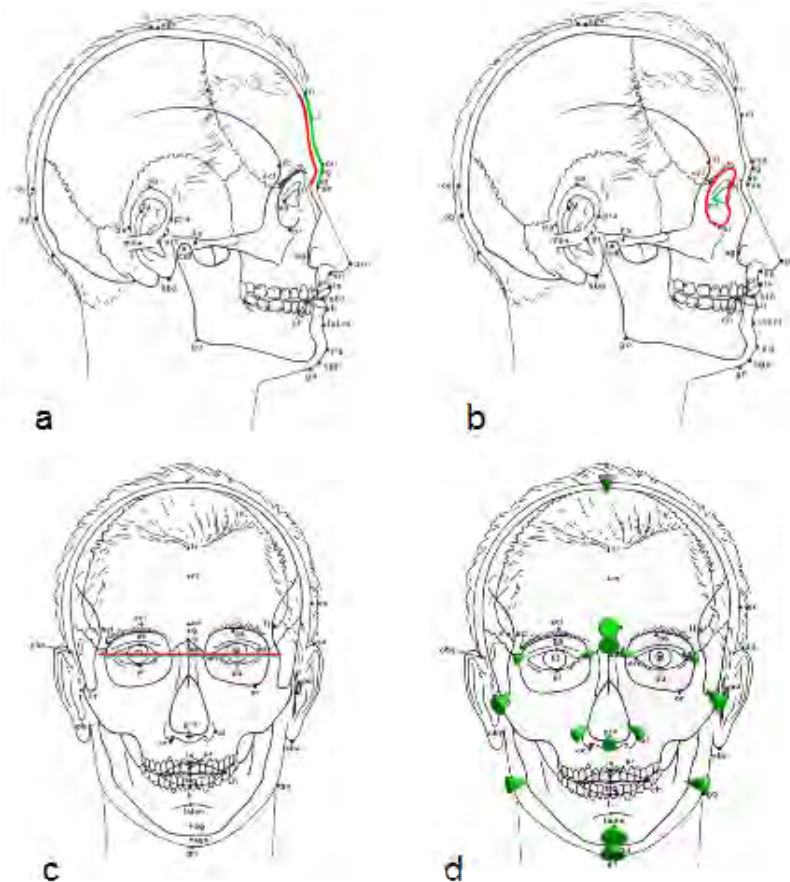Figure 18: Examples of the four families of criteria for assessing the craniofacial correspondence. a) Consistency between the outline of the frontal bone and the forehead outline in lateral view; b) Consistency between the lateral angle of the eye and the orbit; c) Ectocanthion lines marked in the skull and the face; d) Facial landmark position from a cranial landmark using cones to model the soft tissue thickness.

The whole CFS process is affected by several uncertainty sources and degrees of confidence that must be considered for decision making. In particular, we have distinguished the following sources of uncertainty from the forensic experts' experience.

Bone quality: the quality of the skeletal remains is an important issue during the CFS process. The condition of the bones depends on environmental factors, its preservation state has a direct influence on the confidence on the evaluation of face and skull anatomical correspondence.

Image quality: photographic quality is an additional criterion that has to be taken into consideration. The uncertainty inherent to the location of landmarks and regions in an image, already described in [DWK$^+$15], can be greatly affected by the quality of the image. In particular, the location and evaluation of each single region/landmark is affected by the following sources of imprecision: 1) the variation in the distribution of shadows which depend on the light; 2) an unsuitable focus, especially when the plane of focus is not deep enough and hence the critical objects are not sharp; 3) the image resolution: for optimal examination, experts recommend using photographs in which the facial image resolution is at least 180 pixels corresponding to the width of the head, or 90 pixels between the eyeballs (for full frontal images); 4) the pose of the face in the image, i.e. angle of view (frontal, lateral or oblique) and facial expression; 5) a complete or partial occlusion of a region due to the presence of elements such as glasses, clothes, or hair [DWK$^+$15].

Skull-face overlay accuracy: the confidence degree of the SFO accomplished in the previous stage is another important factor to be considered. This process focuses on achieving the best possible superimposition of the skull and a facial image and it can be influenced by different sources of uncertainty, as described in [ICDS11].

Morphological aspects: the degree of confidence of each particular criterion analyzing anatomical correspondence can be affected by several factors. Firstly, the expected craniofacial relationship of a specific region is affected (in a lower or higher degree) by the age, sex, ancestry (biological profile), and/or the body mass index (BMI) [ARP07, KS10]. Thus, these factors have to be considered for each particular criterion during the decision making process. For example, the chin shape of the skull follows the facial shape in young people, but after 45 years of age this relation is increasingly unreliable. This relation can also be distorted due to being overweight. Secondly, each isolated region can have a different discriminative identification power. This is considered as the rate of being able to make a positive identification taking into account only that region.

Automatic method modeling the spatial/morphological relationship accuracy: different computer methods can be considered to automatically evaluate the degree of matching for each specific criterion. The accuracy of a method is defined as how well the specific craniofacial relationship is modeled by that method for an specific criterion.

The DSS proposed in this work considers the evaluation of the skull-face anatomical correspondence at different levels. In each of them, the various sources of uncertainty introduced above are modeled, and different aggregation mechanisms account for information fusion and propagation. In particular, we have defined the following three levels in a decision hierarchy (see Fig. 19):

- Level 1: CFS evaluation.

- Level 2: SFO evaluation.

- Level 3: Criterion evaluation.

In this DSS scheme, the final degree of the CFS identification is obtained by aggregating all the SFO degrees. This corresponds with the highest level (**Level 1**) in the hierarchical system. In the next level (**Level 2**), for each SFO achieved in the previous stage, we aim to analyze the degree of fulfillment of several criteria studying their skull-face anatomical correspondence. For each criterion, facial and cranial regions are located (in the facial photograph and the 3D model,
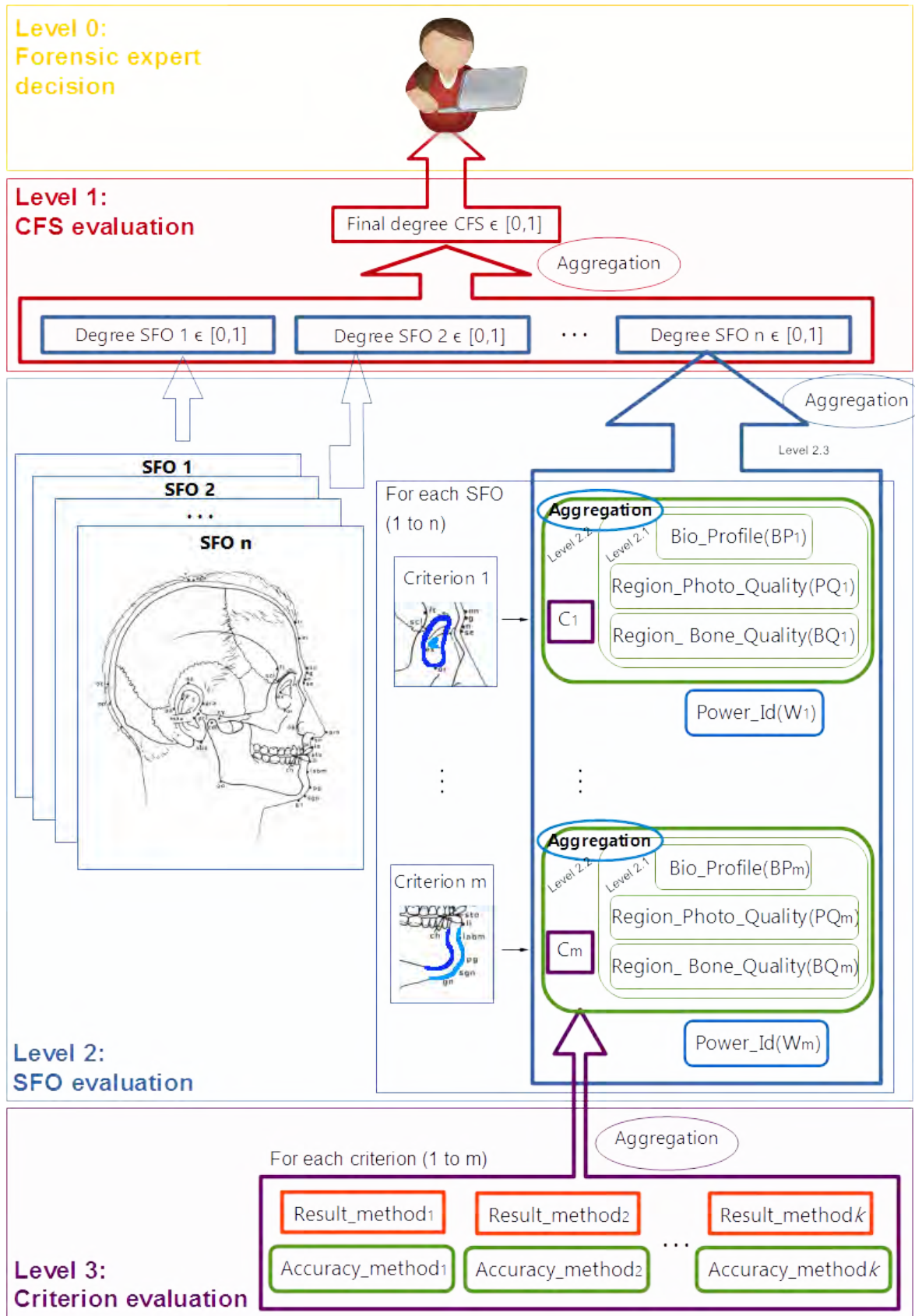
Figure 19: Hierarchical scheme of the DSS for CFS.

respectively) and a specific method to evaluate the skull-face correspondence is applied (the kind(s) of method(s) to be considered depend(s) on the nature of the specific criterion). The degree of craniofacial correspondence of a SFO is computed by aggregating the matching degree of each single criterion taking into consideration the confidence of that criterion. Thus, the skull-face consistency in a region is expressed by a value between 0 and 1, obtained in the previous level (C_m). This value is complemented by the region/criterion confidence based on the quality of the photo (PQ_m), the quality of the bone (BQ_m), the biological profile variability of the criterion (BP_m), and the discriminative power of the isolated region. At this level, we set three different aggregation sublevels. The first one consists of aggregating the first three sources of uncertainty to get a single uncertainty value associated to the criterion sample quality and biological variability. The second sublevel integrates this aggregation with the matching degree (C_m). Finally, at the third one, we obtain the degree of the SFO craniofacial correspondence by aggregating the different previous values for all the regions taking into account the discriminative power of the isolated region as weight. We denote these sublevels as level 2.1, level 2.2, and level 2.3, respectively. If there are more than one CV-based method to evaluate a specific criterion, there is a need to aggregate the results of all of them (**Level 3**). To do so, we take into account the accuracy of each method. Fig. 19 graphically summarizes the proposed hierarchical DSS for CFS.

In addition, but not included within our DSS, there is a Level 0 of evaluation, the one carried out by the human expert. At this higher level the input is the final degree of CFS matching provided by the DSS. Then, the forensic experts will make the final identification decision. In this sense, the MEPROCS international consortium agreed a set of possible decision degrees according to the quality and quantity of materials [DWK$^+$15]. The matching degree provided by our DSS could be directly incorporated within this scale, to any other, or considered in its own.

In the following subsections, the three levels of the hierarchy are described in detail.

3.3.2 Framework Implementation. Level 3: Criterion Evaluation

We have implemented different CV methods and aggregation techniques for each considered criterion. We classify them regarding to the family of criteria they belong to. The four families are described in Section 3.3.1.

- Modeling the anatomical consistency of the bony and facial outlines (first family of criteria)

  By analyzing the existing studies of this area, we have identified the three most common relations between the bony curves (B) and the face curves (F) used by forensic experts as follows: *F follows B*, *F is consistent with B*, and *F is the mirror image of B*. Regarding to their modeling, the former two have similar effects. Notice that these kinds of relations are established in different facial regions. For example, the outline of the frontal bone follows the forehead outline and the chin outline is consistent with the mental outline (see Fig. 18.a for a visual example).

  In order to compare both curves there is a need to extract the corresponding contour of the skull and facial regions (the bony outline) under study (see Fig. 18.a). Automating contour extraction using CV techniques is a difficult task when precision and robustness are demanded despite photographs and 3D models resolution, quality, noise or occlusions. In addition, it could require a different approach depending on the region at hand. Thus, manual delineation of both corresponding contours by the forensic expert have been preferred. The main idea is to select the corresponding curve of the skull that is located close to the facial curve of the photograph (See Fig 18.a). In this third stage, we just work with optimized overlays obtained in the previous step, so the skull will be (more or less) correctly superimposed onto the facial photograph. In the case when the bony outline crosses with the facial one, the studied criterion is evaluated as a zero matching since this fact is anatomically impossible.

As a proof of concepts of the proposed methodology we consider two different kinds of methods to model the *F follows B* relation: the spatial relation *along* and shape similarity measures.

The spatial relation *along* gives us the degree to which an object A is along an object B. The considered approach is based on computing a degree of elongatedness of the region lying *between* A and B, i.e. the between region [TJB05]. Thus, it directly models the specific case of the *F follows B* relation by measuring whether the projected bone curve follows its counterpart facial curve in the 2D facial image.

Two main steps have to be performed [BCC06]: i) calculate the *between* region (noted by β), and ii) measure how elongated it is (thus defining the degree to which A is along B). Once the *between* region is obtained, the *alongness* measure may be computed as:

$$S_{alongness1} = f^a(P(\beta)^2/A(\beta)), \tag{I.28}$$

$$S_{alongness2} = f^a(P(C_{IN})^2/A(\beta)), \tag{I.29}$$

where P and A represent the perimeter and the area of β, and C_{IN} is the inner contour that corresponds to those contours portions which belong to β (in Fig 24 $P(C_{IN}) = P(C_B) + P(C_F)$). To normalize this measure between 0 and 1, we use a sigmoid function as $f_a(x) = (1 - exp(-ax))/(1 + exp(-ax))$, taken from [TJB05]. These measures S_1 and S_2 tend towards 1 as β becomes more elongated. Absolute values can be changed by tuning parameter a to enhance the differences depending on the specific application.

Based on this idea, we propose another method to model the *F follows B* relationship. It consists of measuring the alongness of the *between* region using the classical definition of the circularity ratio of a shape [YKR+08]. The circularity ratio is the ratio of the area of a circle having the same perimeter:

$$C(\beta) = 4\pi \frac{A(\beta)}{P(\beta)^2} \tag{I.30}$$

where A is the area of the shape and P is the perimeter (in this case the area and the perimeter of the *between* region). This expression gives us a value in the interval [0,1] that tends towards 1 as the region β becomes more circular. So, the final formula for our model will be:

$$S_{circularity} = 1 - C(\beta) \tag{I.31}$$

where S_3 tends toward 1 as β becomes more elongated, and accordingly when a curve is along the other.

These three versions of the spatial relation *along* have the problem that they are sensitive to the distance between the contours under comparison together with their length. First and second version (Eqs. I.28 and I.29 respectively) can be adapted to each case tuning parameter a but this is not a desirable solution for an automatic system.

On the other hand, similarity measures between shapes have been widely studied in the computer vision community. They are usually employed to identify objects in images despite the different location, size, rotations, or deformations [Vel01]. They are obtained from a distance metric and thus are affected by the properties of such distance. Hence, there is a need to select the most appropriate one for the problem at hand to achieve a good performance.

A common way to study the similarity between two shapes involves comparing both shapes represented by their features (using shape parameters or shape description techniques). Shape parameters are simple geometric features that can be used to describe shapes in a general way (i.e.: circularity ratio, eccentricity, elliptic variance, convexity, etc.), but they can only discriminate shapes with large differences. For our study, we propose to compare the circularity ratio of both shapes (Eq. I.30). In order to determine the similarity between these two values (in the $[0, 1]$ interval), we use the following expression based on [PK93]:

$$S_{shape\ features} = 1 - \frac{|C(F) - C(B)|}{C(F) + C(B)} \tag{I.32}$$

where F and B represent the facial and the bony chin shape, respectively.

Alternatively, there are many different image analysis methods, contour-based and region-based, to measure the representation and description of a shape [ZL04]. We will only focus on contour-based methods due to the impossibility to precisely locate the bony region within a facial photograph. In addition, these measures are usually employed to compare objects of the same kind, so in general, the required features are scale, rotation, and translation invariance. As we have explained, that is not exactly the case of our problem.

Based on the overview presented in [YKR$^+$08], we have selected two methods considering one-dimensional functions for our application: shape signatures and chain code representation. A shape signature represents a shape by a one dimensional function derived from shape boundary points. For our application we will use the following shape signatures:

– Complex coordinates:
  A complex coordinates function is simply the complex number generated from the coordinates of boundary points, $P_n(x(n), y(n)), n \in [1, N]$:

$$z(n) = [x(n) - g_x] + i[y(n) - g_y] \tag{I.33}$$

  where (g_x, g_y) is the centroid of the shape, given by:

$$\begin{cases} g_x = \frac{1}{N} \sum_{i=0}^{N-1} x_i \\ g_y = \frac{1}{N} \sum_{i=0}^{N-1} y_i \end{cases} \tag{I.34}$$

– Centroid distance function:
  The centroid distance function is expressed by the distance of the boundary points from the centroid (g_x, g_y) (Eq. I.34) of a shape

$$r(n) = [(x(n) - g_x)^2 + (y(n) - g_y)^2]^{1/2} \tag{I.35}$$

– Area function:
  When the boundary points change along the shape boundary, the area of the triangle formed by two successive boundary points and the center of gravity also changes. This forms an area function which can be exploited as shape representation.

In order to measure the similarity between two curves we compute the Euclidean distance of their corresponding shape signatures. The following formula gives a similarity value between zero and one:

$$S_{shape\ signatures} = 1 - \sqrt{\frac{1}{N} \sum_{i=1}^{N} (s_F(i) - s_B(i))^2} \tag{I.36}$$

with s_F and s_B being the shape signature of each contour.

In addition to shape signatures, chain code is a common approach for representing contours. It describes an object by a sequence of unit-size line segments with a given orientation [ZL04]. Freeman [Fre61] introduced a chain code that describes the movement along a digital curve by using so-called 8-connectivity or 4-connectivity. The direction of each movement is encoded by the numbering scheme $\{i|i = 0, 1, 2, ..., 7\}$ or $\{i|i = 0, 1, 2, 3\}$ denoting a counter-clockwise angle of $45° \times i$ or $90° \times i$ regarding the positive x-axis.

In our implementation, we compute the similarity between boundaries by comparing their chain codes. Using 8-connectivity, the formula can be defined as follows:

$$S_{chain\ codes} = 1 - \sum_{i=1}^{N} d_i / N * 4 \tag{I.37}$$

where

$$d_i = \begin{cases} d_i^* & \text{if } d_i^* \leq 4 \\ 8 - d_i^* & \text{if } d_i^* > 4 \end{cases}, \tag{I.38}$$

$$d_i^* = |ch(F) - ch(B)| \tag{I.39}$$

with $ch(X)$ being the chain code representation of an object, d_i the distance between two chain code representations, and N the number of points of the objects. Notice that we normalize the final measure with respect to $N * 4$ since 4 is the largest distance between two chain codes using 8-connectivity (i.e., the opposite direction).

This method is very sensitive to noise or to small variations. Once we have the chain code representation of each contour, we calculate the mean of a number of consecutive elements (w) and we compare the final 'smoothed' chain code representations to reduce this sensibility.

These approaches (shape signatures and chain code comparison) imply both contours must have the same number of points. To solve this problem a sampling method is applied in the contour having the highest size.

- Measuring the relative position of two regions for studying anatomical consistency (second family of criteria)

The second group of criteria rely on a positional relationship analysis in order to assess anatomical consistency. Thus, the goal is to measure the relative position of a skull region against a facial region and determine the degree of anatomical consistency of the resulting relative position. For instance, the lateral angle of the eye must lie within the lateral wall of the orbit in both lateral and oblique views (see Fig. 18.b).

In a 2D image, any positional relation between two objects A and B can be obtained easily. The main directional spatial relationships are shown in Fig. 20. To model these relationships we take the relative position between two points as a base [BR03]. The angle between the segment joining two points a and b and the x-axis of the coordinate frame is denoted by $\vartheta(a, b)$ (see Fig. 21). This angle takes values in the range $[-180°, +180°]$.

Relations "left", "right", "above" and "below" can be modeled using $cos^2\vartheta$ and $sin^2\vartheta$ functions [BR03], as shown in Fig. 21. These functions can be directly used as fuzzy membership functions to establish a membership degree in $[0, 1]$ for each relation. To model
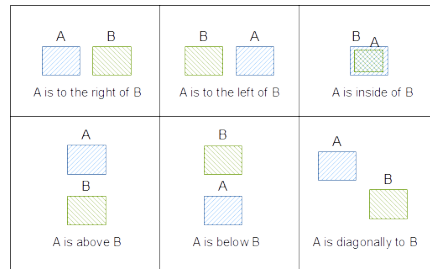
Figure 20: Main directional relative position between two objects.

"diagonally", we combine two of these relations. For the case of the Fig. 21, a is 0.5 to the left of b, 0.5 above b, 0.0 to the right of b, and 0.0 below b.

The positional relationship between two 2D objects can be computed in two different ways [KW95]:

1. Aggregation method. It involves computing the relationship between each pair of points in A and B and calculating the "A relation B" as the aggregated value of all of them. We choose the average mean for this computation.

2. Centroid method. It is based on calculating the positional relation between both centroids.

Thus, the positional relation between two objects A and B is expressed by a degree $\delta(A, B) \in [0, 1]$ for each position.
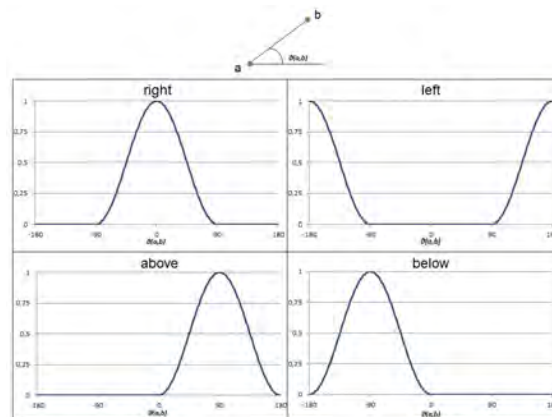
Figure 21: Functions to calculate the relative position between two points a and b.

Once the relative position between two regions is obtained, we need to assess if that position is anatomically consistent. One choice is comparing the current relative position against a reference. Specialized forensic anthropology literature [ASM94] normally refers to the relative position of anatomical regions according to lateral or frontal facial pose within the photograph. However, these criteria always depend on the photograph view and in most of the cases the facial pose is not exactly frontal or lateral. Alternatively, we propose to compare the actual relative position against the relative position given by a 3D reference model.

From a given set of 3D head models we could create a statistical mean facial 3D model and a skull 3D model. Then, we can apply the resulting geometric transformation obtained in the SFO stage to the reference 3D models (skull and face) onto a white image with the same size

as the photograph. Proceeding in this way, the positional relationship between skull and face regions in the reference case is obtained and can be compared with the positional relationship between the same regions in the given case (See Fig. 22).

We should note that this option is only applicable when the positional relationships are not significantly affected by population factors like ancestry, age, sex, etc.

In order to compare two different relative positions between two objects, we compute the similarity using the following expression based on [PK93]:

$$S_{rel\ position} = 1 - \frac{|\delta_a(A,B) - \delta_a(C,D)| + |\delta_b(A,B) - \delta_b(C,D)| + |\delta_r(A,B) - \delta_r(C,D)| + |\delta_l(A,B) - \delta_l(C,D)|}{\delta_a(A,B) + \delta_a(C,D) + \delta_b(A,B) + \delta_b(C,D) + \delta_r(A,B) + \delta_r(C,D) + \delta_l(A,B) + \delta_l(C,D)}$$
(I.40)

where δ_a, δ_b, δ_r, and δ_l are the degrees of the position relation "above", "below", "right", and "left", respectively.
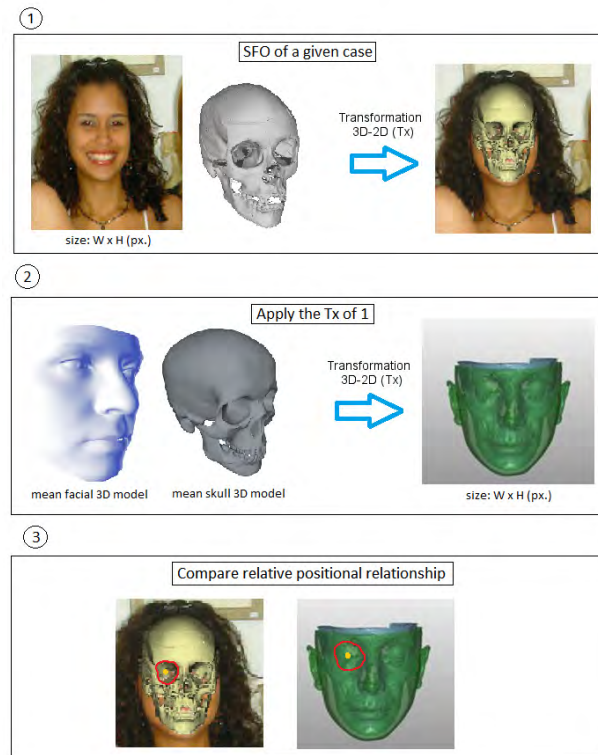
Figure 22: Scheme of the solution to solve the dependency on the photograph view to check relative position between skull and face.

- Modeling the anatomical consistency by the position of two bony regions (second family of criteria)

  Another of the sets of criteria that forensic experts take into account in the decision making stage is the consistency of hard tissue to hard tissue positions. From a CV point of view, the implementation consists of overlapping two regions. Manual marking of both corresponding regions has to be done by the anthropologists in an approximate way, keeping in mind that the only visible hard tissue in the face are dental pieces. The 3D regions are projected into the 2D image using the geometric transformation obtained in stage two of CFS for accomplishing SFO. Then, the aim is to check whether the two regions are located in the same place of

the image. Due to the quality of the materials, the skull regions could be greater than the photograph's ones or viceversa (dental pieces in photographs are in most of the cases partially occluded by the lips). Thus, we propose a 'special' kind of overlapping degree. That is, when one of the regions is entirely inside the other, the position of both objects is considered to be consistent (a value of 1 is assigned for the correspondence degree in that case). Note that the length of the regions is not taken into consideration. On the contrary, when the whole facial region is outside the corresponding cranial region, the position of the objects is not considered to be consistent at all (a value of 0 is assigned). For any intermediate situation of a partial matching between the two regions, the part of the object inside the other is considered to compute the consistency degree defined in $[0, 1]$ by means of Eq. I.41.

$$S_{overlap} = 2 \cdot \frac{A(Region_{skull}) \cap A(Region_{face})}{A(Region_{skull}) + A(Region_{face})}, \tag{I.41}$$

where $A(Region)$ is the area of a region.

- Modeling the anatomical consistency by line location and comparison (third family of criteria)

Following this set of criteria, experts analyze a set of marking lines, obtained by joining some reference landmarks on the face and skull. It is important to note that the cranial landmarks, which are used to obtain the corresponding cranial line, are 3D points. These points are projected into the 2D image using the geometric transformation of the SFO stage. Then, the comparison of both lines occurs in the 2D space. Depending on the lines at hand, two different aspects can be distinguished. Firstly, the study of the parallelism of both lines (cranial and facial). Secondly, the similarity of their lengths.

In terms of CV and regarding to our DSS framework, these criteria has to be given as a value in the interval $[0, 1]$. For the first case, the angle between both lines has to be calculated. To do so, we use the following formula:

$$\alpha = acos\left(\frac{\|\vec{v}_{skull} \cdot \vec{v}_{face}\|}{\|\vec{v}_{skull}\| \cdot \|\vec{v}_{face}\|}\right), \tag{I.42}$$

where $\vec{v}_{cranial}$ and $\vec{v}_{facial}$ are the cranial and facial lines, $\cdot$ represents the inner product, and $\|\vec{v}\|$ refers to the magnitude of the vector.

Thus, $\alpha \in [0^o, 360^o]$. Although the skull could not belong to the subject of the photograph, we can actually suppose that the SFO is correctly performed as it is guided by the landmark matching, so both lines will be more o less parallel. For this reason, we establish the worst case of this relation when the angle formed by the two lines is greater or equal than 45°. The final degree, which expresses the consistency between the cranial and the facial lines in a SFO regarding to the parallelism between them, is:

$$S_{parallelism} = \begin{cases} 1 - \frac{\alpha}{45}, & \text{if } \alpha \leq 45^o \\ 0, & \text{otherwise} \end{cases} \tag{I.43}$$

For the second case, we need to compare the lengths of the two lines. To do so, we establish the worst situation where the difference between both lengths is equal to the half of the greatest line. Thus, the similarity between the cranial and facial lines regarding to their length is:

$$S_{length} = 1 - \frac{\|\|\vec{v}_{skull}\| - \|\vec{v}_{face}\|\|}{\|\vec{v}\|_{greatest}/2}, \tag{I.44}$$

where $|\vec{v}|$ refers to the absolute value and $\|\vec{v}\|_{greatest}$ is the length of the greatest line. In the case that the latter formulae returns a value lower than 0, the similarity between the two lines will be 0.
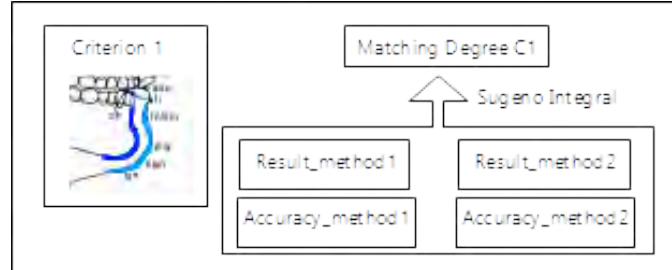

Figure 23: Scheme of the matching degree calculation for a criterion.

Every performed method returns a value between zero and one, which indicates how well the relation is achieved. Then, we aggregate the degrees of support of the individual methods (when there are more than one) to strengthen the final result. To do so, a measurement of importance of each one (the accuracy) is needed. Regardless the criterion type, this individual method accuracy is calculated as its capability to discriminate in the decision making process (ranking positive and identification negative cases). Hence, a database composed of real positive (including the 3D skull model and one or more subject photographs) and negative identification cases is required. These negatives cases are obtained combining real 3D skull models and non corresponding photos of similar subjects (same gender, age, ethnic group, ...). First, the corresponding value to apply each method for a specific criterion over the database of cases is achieved. Next, the matching values reported are used to rank the candidates based on their chance to be the actual subject. Then, a value between 0 and 1 is assigned to each positive case taking into account this ranking: if the method reported the highest value to a positive case (first position of the ranking), 1 is assigned. On the contrary, if the position of a positive case is the last of the ranking, 0 is assigned. The formula to assign the accuracy of the method x_i in the instance j is:

$$Acc(x_i)_j = 1 - \frac{r-1}{M_j - 1} \tag{I.45}$$

where r is the position of the positive case in the ranking and M_j is the worst (highest) value of the ranking for the instance j (all cases getting the same criterion-method value are supposed to have a draw, that is, they are assigned to the same ranking).

Then, the average of these accuracy values over all cases is calculated. As a result, an accuracy index in the interval $[0, 1]$ is achieved for each method. The final step is to aggregate the degrees of support of the best individual methods taking into account their accuracy. To do so, Sugeno integral [Sug74] is used. The scheme of this procedure is shown in Fig. 23.

The former criteria can be applied to different cases of study related to different facial regions. We have studied and modeled the following relations according to the expert knowledge [DIC17]:

- Chin contour

  Within the group of criteria that study the consistency of the bony and facial morphological curves, experts have stated that the chin outline has to be consistent with the mental outline [JSA01]. Our goal is to analyze if both shapes are similar, but not identical, since they belong to different objects (the skull and face have a different size). At the same time, the distance between them (according to soft tissue depth), their orientation, and their relative position

(anatomical correspondence) must be taken into account. A graphical example is depicted in Fig 26.i. Hence, unlike most common applications of similarity measures (mainly in the field of image retrieval), we need some degree of translation, scale, and rotation sensitivity while analyzing shape similarity. Rotation sensitivity is critical in order to model *F follows B* spatial relation. Translation and scale differences are expected due to the anatomical distance and size difference of the regions under comparison.

To do so, we use a method that involves the convex hull of the union of the two objects in order to segment the contours. The first step involves obtaining the region between both objects defined as [BCC06]:

$$\beta_{CH}(F, B) = CH(F \cup B) \cap F^C \cap B^C \tag{I.46}$$

where $CH(X)$ denotes the convex hull of X and X^C its complement.

The basic idea of this method is that the contour of the *between* region is composed of an alternate sequence of line segments of the convex hull and the boundary of the two objects. The intersection points are the pairs of subsequent vertexes belonging to different objects. These points divide each contour into two portions (inner and outer contours). The inner contours correspond to the skull and face contours that we want to analyze (refer to Fig. 24 for an illustrative example). In some cases this method has a wrong behavior under some circumstances as depicted in Fig. 24 b. That problem is solved setting the intersection points of the bony contour as the nearest points to the corresponding intersection points of the facial contour (Fig. 24 c). By using this segmentation method, the section of the contours under comparison changes according to their relative position and size. Thus, despite the method employed to model the *F follows B* relation, the resulting approach will be sensitive to translation and scale.
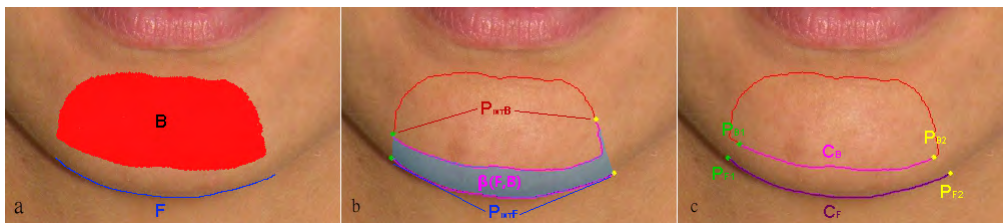
Figure 24: Chin contours segmentation. (a) Original regions, facial (F) and bony (B). (b) Convex hull $CH(F \cup B)$ with intersection points between $\beta(F, B)$ and CH. (c) Final contours with corresponding extreme points.

Once the contour segmentation has been accomplished, we have to study whether the facial curve follows the bony curve. With this segmentation method, when both curves cross, it is not possible to extract the contours, so the result value will be zero. This result concurs with the fact that the crossing of these two regions is morphologically impossible.

We have made a deep study to analyze which of the CV methods models this relation better. We have proven every method individually and the results have shown that the best performance is achieved with shape similarity measures. In particular, the area function comparison method and the complex coordinates' signature using Eq. I.36 achieved the highest accuracy (Eq. I.54). After applying the scheme of Fig. 23 with these two methods, we concluded that the accuracy index of the aggregation overcame the individual results.

The age and the BMI can affect to the chin criterion decreasing its reliability [ARP07, KS10]. For this study, the influence related with biological profile was defined by Prof. Wilkinson

according to her expert knowledge and represented using the fuzzy sets in Fig. 25. As can be seen, the chin criterion is less reliable after 60 years old, decreasing to 0.25 from 75. Similarly, the reliability decreases with BMI values above 35.
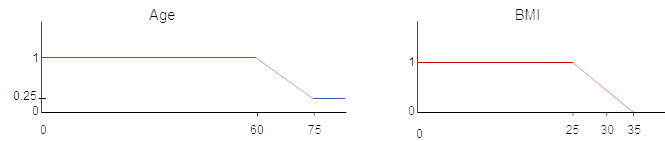

Figure 25: Defined fuzzy sets to model morphological aspects for the chin outline. At the left the age related confidence and at the right the BMI related confidence.

- Cranial contour

  The relationship between the shape of the head and the shape of the cranium is well established [DIC17]. Similar to the previous case, the head outline has to be consistent with the cranial outline (See Fig. 26.iii). However, there are translation and scale differences due to the anatomical distance between the two objects and the size difference of the regions. For this reason this region is modeled as the chin contour explained before. This relation is not affected by any biological aspect.

- Cranial orbit and eyeball center positional relationship

  One of the criteria that focuses on analyzing the relative position of two regions involves studying the anatomical consistency in the position of the orbits and the center of the eyeball (Fig. 26.ii). This relationship is independent of populations factors like race, age, sex, etc. [WM03]. For this reason, we use a reference model to compare with each specific case as shown in Fig. 22. To evaluate this criterion, we need to extract the appropriate contours both for the orbits (bony region) and the eyeball center (facial region). The desired orbital contour is the interior contour, selected as the smaller one, while the facial region is simply the contour of the marked zone on the photograph. Once the contours are segmented, the matching degree is calculated using Eq. I.40 with the two explained methods: aggregation and centroid. Finally, we prove that the aggregation of these two methods following the scheme of Fig. 23 overcomes the individual accuracy (Eq. I.54).

- Upper rim of the external auditory meatus and facial tragus (ear) positional relationship

  The ear morphology and its corresponding skeletal structure is one of the most understudied craniofacial relations. Some of the existing studies conclude that the tragus of the ear has an estable positional relationship with the upper rim of the external auditory meatus (see Fig. 26.v). Thus, we have modeled this criterion as those belonging to the second family of criteria (same than for the orbits and eyeball centers, see Fig. 22). On the image, the facial tragus is marked as a region. On the 3D model skull, the external auditory meatus is also marked as a region by the expert. Once the 3D region is projected into the image, the similarity value is obtained using the comparison of the positional relation of a reference model. This relation does not change with age, BMI, or ethnic group.

- Mouth occlusion length comparison

  There are many studies from orthodontic and anatomical fields that relate the mouth with the occlusion of the teeth. One of the most common relations used by experts is that the mouth corners are positioned on radiating lines (perpendicular to the palate arc) from the first premolar-canine junction. In order to model this we need information about the direction of the premolar-canine junction in 3D but it is not available. In [Ste03], the author tested
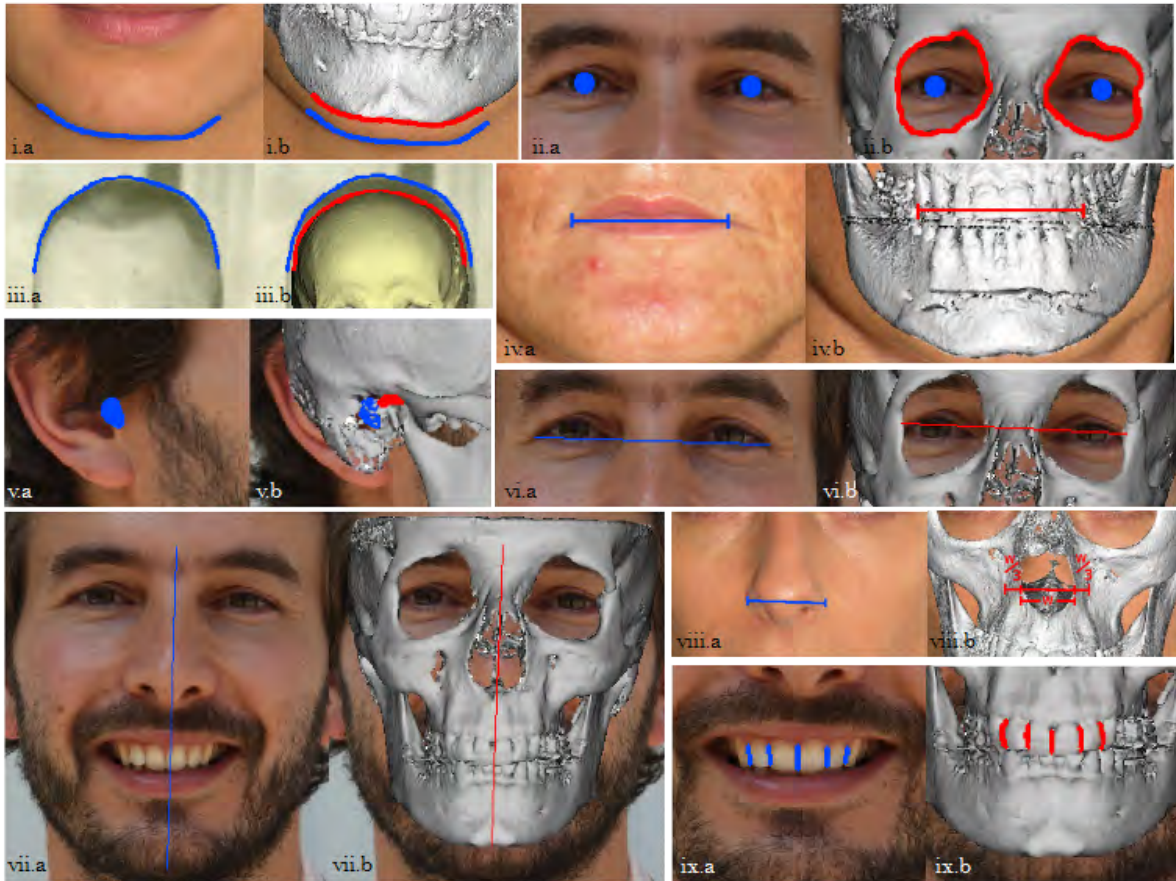
Figure 26: Different skull-face regions modeled using CV methods for the proposed DSS.

that the intercanine distance is 75% of the overall mouth width. Another study suggested that the mouth corners are located below the infraorbital foramina [SM08]. However, the results of this study cannot be generalized since it is based on a small sample of cases. For this reason, in our proposal we model the method based on the intercanine distance (a visual representation is shown in Fig. 26.iv).

To perform this comparison using CV, the facial occlusion line of the photograph (in frontal view and neutral pose) and the region between the two first canines in the skull 3D model have to be marked manually. Once the 3D region is projected into the 2D image, the two most remote points are joined. Hence, this length is comparable with the line of the facial image using Eq. I.44. The mouth corner position does not change with age or BMI, but the dental points may change due to loss of teeth.

- Nose width length and parallelism comparison

The most studied feature of the face is the nose. In [PU02], authors concluded that the bony nasal aperture at its widest point is three-fifths of the overall width of the soft nose (Fig. 26.viii). A post-study performed on living subjects of different ethnic groups confirmed this relationship [Ste03]. Accordingly, from a CV point of view, the maximum width of the soft nose is the width of the nasal aperture plus two-thirds of it. In order to model this relation, a line over the maximum width of the nose is marked on the facial photograph. In the 3D model, experts mark two 3D points corresponding to the maximum width of the nasal aperture and, in this line, a vector with the same direction and with a magnitude of a third

of this width is added to each extreme. Next, this whole line is projected into the 2D image and it is compared with the line marked on the photograph. For this criterion we use Eqs. I.43 and I.44, since the position of both lines must also be consistent. We aggregate these two values using the average, as both measurements have the same importance to compute the final consistency value. There is neither change in nasal width with increased age, nor variation for ethnic group or BMI.

- Ectocanthion lines parallelism comparison

  Commonly, experts focus on the lines on the face and the skull to assess the anatomical consistency in a SFO. In particular, this line is achieved joining the two excanthion landmarks in the skull and the two ectocanthion landmarks in the photograph, and both should be parallel in a positive case as it is depicted in Fig. 26.vi. For modelling this relationship, the 3D landmarks are projected into the 2D space and then the lines are compared using Eq. I.43. There should be no changes in position with age or BMI, but the visibility may change due to dropping of eyelids and shortening of fissure. Within our model, this fact is taken into account using the quality of the regions parameter.

- Frontal/central line parallelism comparison

  This is another way to analyze the anatomical correspondence in a SFO. The frontal line is achieved joining the glabella and the gnathion landmarks both in the skull and in the facial image (see Fig. 26.vii). As in the former criterion, the 3D landmarks are projected into the photograph and the degree of consistency is computed using Eq. I.43. Note that this criterion is only suitable for frontal view photographs. This relation does not change with age or BMI.

- Overlap of inter-dental lines

  When dental pieces appear in the AM photograph, experts give them a special attention since the same object can be compared in both skull and face. Unfortunately, it is almost impossible that the whole tooth is shown in the photograph or kept in good conditions in the skull. For this reason, we use the lines between the dental pieces to establish the comparison instead of the whole tooth. A graphical example of this criterion is shown in Fig. 26.ix. In this way, experts manually mark the corresponding lines on both surfaces. These marking lines are actually regions in the 3D skull model. Therefore, in the 2D image they also become regions (once they are projected using the geometric transformation of the SFO). In the photograph, lines are directly marked and then they must be in the same location than the projected regions. We use Eq. I.41 to measure that correspondence.

3.3.3 Framework Implementation. Level 2: SFO Evaluation

In level 2 we get a single value measuring the matching of the skull-face correspondence in an SFO by aggregating the degrees of the craniofacial matching the different regions considered weighted by the associated criterion uncertainty (see Fig. 27).

   We have the problem of how to choose an aggregation function for each sublevel within the vast variety of aggregation functions available in the literature. In [BPC07] the authors give some recommendations to select the most appropriate aggregation function for a specific application. First, the function must be consistent with the semantic of the aggregation process, i.e, if one is a disjunction, conjunctive or averaging aggregation functions are not suitable. Other important aspects to take into account are if the aggregation function should be symmetric, idempotent, or have a neutral or absorbing element. Whether the number of inputs is always the same and what is the interpretation of the input values are also important facts to make a good choice of a suitable family or class. The second criterion is to select the appropriate member of that family or class, in

order to produce adequate outputs for the given inputs. In our case, to address this second criterion we decided to perform an experimental study to analyze which aggregation function provides the best results in each case. To do so, we use a data set with positive and negative identification cases that helps us to make a good decision about the most suitable aggregation functions. In this sense, the SFO craniofacial correspondence of the positive cases must be ranked before the negatives in the DSS output.
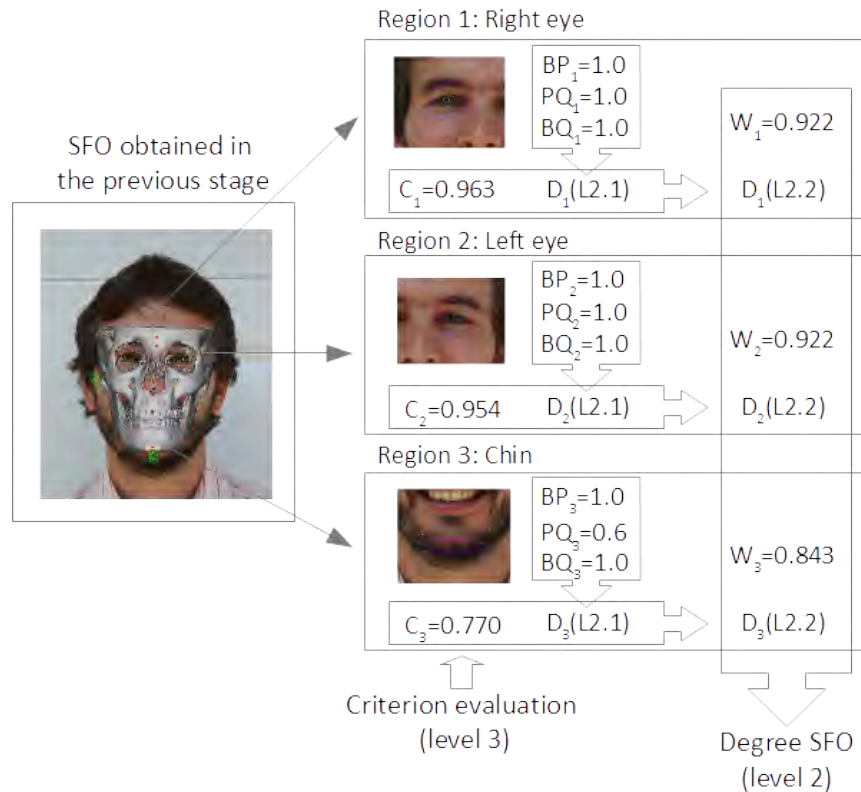
Figure 27: Graphical example of the level 2 of the fuzzy DSS framework (SFO evaluation).

Based on the latter guidelines, we justify the choice of the analyzed aggregation function for each sublevel as follows:

- Aggregation function to combine material quality assessments and biological profile (level 2.1)

  First, we have to aggregate the quality of the photo at the region (PQ_m), the quality of the bone at the region (BQ_m), and the biological profile variability of the criterion (BP_m) (see Fig. 28). These three aspects have a direct influence on the confidence of the matching degree of each particular region. Thus, we have decided to aggregate them in a single uncertainty value using an aggregation function denoted by $O_{Level2.1}(PQ_m, BQ_m, BP_m)$.

  Biological profile: the way biological profile affects the degree of confidence of each particular criterion can be easily modeled using fuzzy sets. According to the expert knowledge (task developed by Prof. Caroline Wilkinson, one of the most recognized experts in craniofacial identification), we have defined one or more fuzzy sets to model in which way the biological profile parameters (age, sex, and ancestry) and the BMI modify the degree of confidence of each particular criterion.

  Bone quality: the variations seen on bones can be described according to the bone's surface texture using the weathering stages [BU94]. They consider the taphonomic processes that may

have affected the bones of a subject. For our application, the accuracy of the 3D model must be taken into account at the same time. Values of quality are set, based on the weathering stages, and are specifically associated with each region of the skull. If a specific region is deteriorated, the method to analyze the skull-face correspondence for each criterion can use it but the confidence in that criterion is reduced, so it will consequently be associated with a lower support value. Similarly to [AAW10], where weathering stages were employed to modify the confidence of age estimation methods, we established quality indexes as ordinal numbers ranging in [0, 1]. They are assigned by a forensic expert according to the analysis of the state of the available skull. The assignation in this manner indicates the least weathering as being a perfect skull region (1) and the most weathering as being a faulty or not present region (0). We use a six-stage system in which stage 0 is determined to be a quality of 1.0, stage 1 of 0.8, stage 2 of 0.6, stage 3 of 0.4, stage 4 of 0.2, stage 5 of 0.1, and stage 6 is assigned a value of 0.0.
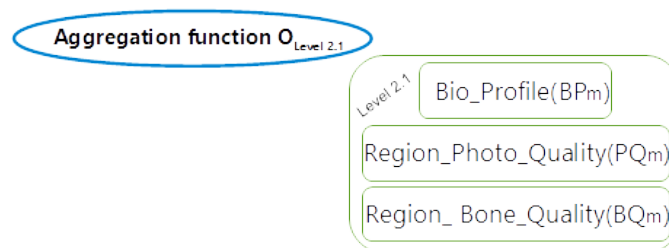

Figure 28: Aggregation scheme at level 2.1 of the DSS proposed framework.

Photo quality: we also use a six-stage system to establish the quality of each facial region. The highest stage means the facial region is clearly identified on the photograph and that is the ideal situation to apply the corresponding method to analyze the skull-face correspondence. On the contrary, a region of the lowest stage implies the impossibility to view that region in the photograph.

To aggregate these values, a conjunctive behavior seems to be the best choice since it does not allow for compensation. Thus, low scores for some criteria (in this quality or biological aspects) cannot be compensated by other scores. If the quality of the bone is very bad, no matter how well the other two sources are, it applies that the matching between the skull and the face will be less reliable. However, averaging mean could be a more conservative choice. On the other hand, we consider that these three aspects affect equally to the region criterion so the aggregation function has to be symmetric. For these reasons, we decide to analyze the behavior of our system using the aggregation functions Minimum, Product, and Arithmetic Mean. Thus, at this level we can state the aggregation function as $O_{Level2.1}(PQ_m, BQ_m, BP_m) = \{min, prod, mean\}$.

- Aggregation function to combine the matching degree and the uncertainty value of level 2.1 (level 2.2)

Secondly, we have to aggregate the previous uncertainty sources with the matching degree of the skull and the face at the corresponding region as can be seen in Fig. 29. For this application, an averaging procedure is required. The basic rule of this class of aggregation functions is that the total score cannot be above or below any of the inputs. The aggregated value is seen as some sort of representative value of all the inputs. In addition, we consider that the aggregated inputs do not have the same contribution to the total output, so a not symmetric weighted function is needed.
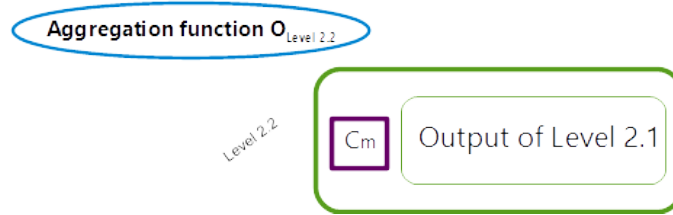
Figure 29: Aggregation scheme at level 2.2 of the DSS proposed framework.

Two of the most common weighted aggregation functions are the Weighted Arithmetic Mean and the Weighted Geometric Mean. Both functions are not symmetric. We establish different weights to each input based on the expert knowledge. This weight, a number $w_i \in [0,1]$, represents the importance of each one. These functions are abbreviated as *wam* and *wgm*, respectively.

Since the values of the weighting vector for the *wam* and for the *wgm* must sum up to 1, we apply a simple normalization of the accuracy index with respect to their sum:

$$w_i = \frac{W_i}{\sum_{i=1}^{n} W_i} \tag{I.47}$$

where W_i is the identification power of the i-region.

This aggregation function can be denoted as $O_{Level2.2}(C_m, Output_{level2.1}) = \{wam, wgm\}$

- Aggregation function to combine the identification power and the degree of level 2.2 (level 2.3)

The final step in level 2 is to obtain the SFO evaluation degree. As explained before, there is a need to aggregate multiple degrees of support with an associated weight. Each of these degrees corresponds to the skull-face matching degree in a specific region with the corresponding uncertainty integrated (Fig. 30). The weight in this case will be the identification power of each isolated part of the face.

The requirement for this aggregation function is that it must be non symmetric. The identification power of a region (the weight) reflects the relative contribution of each input to the final output: the degree of a SFO evaluation.

As in the previous sublevel, we consider the most common weighted aggregation functions: the *wam* and the *wgm*. In this case, the weight which represents the importance of each input is the power identification of each region. These weights are realistic and they are computed using a dataset of positive and negative identification cases in a similar way to the calculation of the accuracy of each method in level 3 (See Sec. 3.3.2 and Eq. I.54). Again, the values of the weighting vector for these functions must sum up to 1, so we apply a simple normalization of the power identification with respect to their sum (Eq. I.47).

Apart from the weighted aggregations, we also use fuzzy integrals as aggregation functions. These operators combine the data supplied by several information sources according to a fuzzy measure, that represents the background knowledge on the information sources. In this study, we use the Choquet and the Sugeno integral and we use a Sugeno λ-measure to determine the fuzzy measure. A fuzzy measure g is a real valued function defined on the power set of X (the universe of discourse), 2^X, with range $[0,1]$, satisfying the following properties (with A and B being two subsets from X):

1. $g(\phi) = 0$, $g(X) = 1$ (boundary condition)
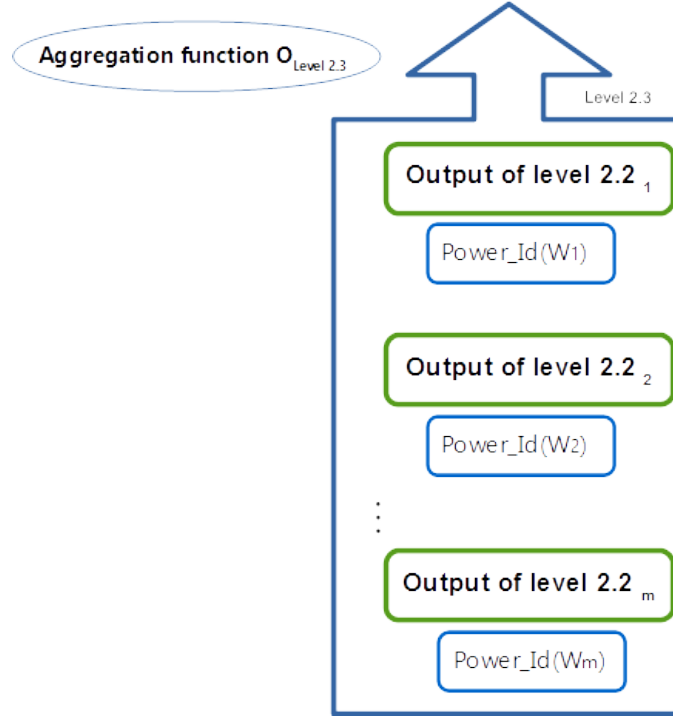
Figure 30: Aggregation scheme at level 2.3 of the DSS proposed framework.

2. $A \subseteq B$ implies $g(A) \leq g(B)$ for all $A, B \in F$ (monotonicity)

A fuzzy measure specifies the opinion of the 'worth' or 'goodness' of each subset of information sources in evaluating a particular hypothesis. Each information source gives a belief or confidence in the hypothesis and the measure lets you know how to weight that belief or confidence, in this case the identification power or each facial region. To determine a fuzzy measure on X, we must identify $2^p - 2$ coefficients satisfying $p.2^{p-1}$ conditions. To solve this drawback, some approaches have been proposed to reduce the number of parameters to be determined [Sug74]. In this paper, we use a Sugeno λ-measure defined as in [AAW10]:

Definition 1 *Let g be a fuzzy measure, then g_λ is a Sugeno λ-measure if there exists $\lambda > -1$ such that*

$$g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A)\mu(B) \tag{I.48}$$

holds for all $A, B \in F$.

It is to be noted that, for a Sugeno λ-measure,

$$\prod_{j=1}^{p}(1 + \lambda g_\lambda(\{x_j\})) = 1 + \lambda \tag{I.49}$$

holds because of the boundary condition [IT03]. Using the above definitions $g_\lambda(X)$ can be constructed from the fuzzy densities of the elements of X. Given the set of densities, the value λ can be easily found as the unique root greater than -1 of a simple polynomial [TK90].

Once λ is found, the fuzzy integral can be calculated.

– *The discrete Choquet Integral*

The discrete Choquet integral with respect to a λ-fuzzy measure is given by

$$C_g(\mathbf{x}) = \sum_{i=1}^{n} [x_{(i)} - x_{(i-1)}] g(H_{(i)}) \tag{I.50}$$

where $\mathbf{x}_\searrow = (x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ is a non-decreasing permutation of the input $\mathbf{x}$, $x_{(0)} = 0$ by convention, and $H_i = (i), \ldots, (n)$ is the subset of indices of $n - i + 1$ largest components of $\mathbf{x}$.

– *The Sugeno Integral*

The Sugeno integral with respect to a λ-fuzzy measure is given by

$$S_g(\mathbf{x}) = \max_{i=1,\ldots,n} \min(x_{(i)}, g(H_{(i)})), \tag{I.51}$$

where $\mathbf{x}_\searrow = (x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ is a non-decreasing permutation of the input $\mathbf{x}$, and $H_i = (i), \ldots, (n)$.

In the following we refer to these aggregation functions as *choq* and *sug*, respectively. Hence, at this level we can state the aggregation function as $O_{Level2.3}(Output_{Level2.2_i}, Power\_ID(W_i)) = \{wam, wgm, choq, sug\}$.

3.3.4 Framework Implementation. Level 1: CFS Evaluation

From a theoretical point of view, the hierarchical framework for decision making proposed in this work covers all sources of information as well as their aggregation and propagation. Nevertheless, there is still a need of providing an implementation for the level 1 of the hierarchy, i.e., the CFS level. In this level, the matching degree of individual SFOs (belonging to different AM images of the same person) have to be fused to produce a unique and final CFS matching degree.

The influence of considering multiple SFOs of the same person within the same CFS identification process was already studied more than 20 years ago in [ASM94], leading to a significant improvement of false matches (from 9% to 0.6%). Similarly, in another study developed in [YIMS95], the unknown skull was positively identified as the missing person, in 35 out of the 37 cases with more than one photograph available, based on more than 13 acceptable matches. In contrast, when the skulls in question were examined with only the frontal face photograph of the missing person (15 cases), the examiners could only found less than 12 matching criteria, leading to a probable identification rather than a positive.

Within our DSS we could either produce a unique CFS matching degree in order to be able to order a set of given candidates according to this value, or we could instead provide a CFS matching degree together with a confidence degree of such a matching value. One of the main variables that should be used to produce a confidence value for a given CFS case is the quality/accuracy of the SFOs considered. However, we have not found the way to measure such parameter, so in this work we focused on providing a single CFS value at Level 1, assuming that optimal SFOs have been achieved in the stage 2 of the whole CFS system.

Therefore, we propose to analyze the aggregation of the different SFO degrees of the same CFS case using the following four operators: the mean, the maximum, the minimum, and a weighting function based on the number of regions analized in each SFO. The latter function is defined in Eq. I.52:

$$Agg\_Nreg = \frac{\sum_{n=1}^{N} (D\_SFO_n \cdot Nreg_n)}{\sum_{n=1}^{N} Nreg_n} \tag{I.52}$$

where N is the number of SFOs in the CFS case and $Nreg$ is the number of regions taken into account in each specific SFO.

3.3.5 Framework Implementation. Automatic CFS as sort-listing and identification tool

The different trials developed to analyze the reliability of the CFS technique along the history have not been successful and they are fraught with limitations [YIMS95, CLT⁺92, ASM94, JSA01, RMA06, GS12, GODA⁺16]. On the one hand, the absence of an objective measure of the skull-face overlay match, technical limitations of the equipment, disregard for accurate landmarks location while performing landmark-based methods, absence of soft tissue data for the tested population, insufficient quality of the 3D cranial models, postmortem photographs, reduced samples, absence of appropriate statistical analysis, and absence of inter and intra-observer studies are but a few. Statistically significant reliability studies tackling these challenges are required to obtain a more solid picture on the reliability of CFS.

On the other hand, the CFS technique has been performed by the experts without a common standard during its existence. The first and major effort to develop a common methodology has been done under the umbrella of the MEPROCS European project. Different quantitative and qualitative inter-lab studies served to develop the first best practice document in the field [IVN⁺15], identified the set of morphological criteria with a greater discriminative power [DWK⁺15], and validated the recently created methodology by measuring the performance of different forensic anthropologists in similar studies accomplished following their own methodology and confronting it with the one developed within the MEPROCS framework [IVN⁺16].

Our proposed system follows the majority of MEPROCS recommendations (still not all, since the articulation of the mandible has not been addressed, and our system, due to its automatic nature, does not consider the physical skull but a 3D model representation). In this dissertation, and as a continuation of the works developed within the MEPROCS framework, we also present a complete methodology to evaluate the identification capabilities of a given method / process. It is composed of two different types of experiments.

The first ones are addressed to evaluate the performance of our system in the identification task, that is to say, a binary response of positive or negative identification must be provided for each particular CFS case. The same kind of test has been already carried out by MEPROCS partners in [IVN⁺15]. The experimental set up involves the comparison of unknown skulls and multiple candidates. For each single case, the experts were asked to report the final identification decision (either positive or negative) along with the rationale supporting the decision. We will thus compare the results obtained by our automatic system with those manually achieved by the forensic experts in this 0study.

The second experiment is designed to study the capability of our system for identifying the correct individual among a list of several negative cases and a single positive case. That is to say, the goal is to measure the sort-listing capabilities of a given system, for which Cumulative Match Characteristic (CMC) curves [JL05] are used. A CMC curve captures the percentage (or probability) that the correct match of a case appears in a list of r best matches, where r denotes the rank. In this rank, we also take into account the percentage with respect the total sample size.

3.3.6 Framework Implementation. Experimental Setup

The experimental design of this study consists of 100 AM photographs and 24 3D models of real skulls. Nine of them are Cone Bean Computed Tomography (CBCT) models of living individuals, and the remaining 15 are skull 3D models of deceased people acquired using a 3D structure light scanner (seven of them acquired with the Artec MHT scanner and eight of them using the Fastscan

Polhemus Scorpion scanner). In order to create the experimental dataset, a cross-comparison was performed, in which each skull 3D model was superimposed with a variable number of photographs, obtaining 591 different SFOs. Each skull has one or more positive SFOs where the skull belongs to the subject of the AM photograph. A previous filter of meaningful candidates based on sex and age was made, so there is not the same number of overlays for each skull. At SFO level (level 2), the dataset is composed of 43 positive and 548 negative overlays. Since there is more than one AM photograph for the same person in some cases, we have to aggregate them by CFS cases. Thus, at CFS level (level 1), the experimental dataset involves 324 cases, 24 positive and 300 negative cases. Table I.7 details the experimental dataset.

Table I.7: Experimental dataset summary

Skull Model	Positive SFOs	Negative SFOs	Positive CFSs	Negative CFSs
CBCT 1	2	24	1	11
CBCT 2	2	24	1	11
CBCT 3	2	24	1	11
CBCT 4	2	24	1	11
CBCT 5	2	33	1	18
CBCT 6	2	33	1	18
CBCT 7	2	33	1	18
CBCT 8	2	33	1	18
CBCT 9	2	33	1	18
3D Model 10	3	24	1	14
3D Model 11	1	26	1	14
3D Model 12	2	25	1	14
3D Model 13	4	24	1	14
3D Model 14	1	27	1	14
3D Model 15	1	21	1	10
3D Model 16	3	19	1	10
3D Model 17	1	18	1	11
3D Model 18	1	17	1	10
3D Model 19	1	16	1	10
3D Model 20	2	16	1	11
3D Model 21	1	16	1	10
3D Model 22	1	14	1	9
3D Model 23	1	14	1	9
3D Model 24	2	10	1	6
Total	43	548	24	300

The SFOs employed have been obtained by our automatic method using the best parameter values obtained in Sections 3.2.3 and 3.2.4. Exceptionally, the CBCTs positive cases have been achieved by means of a ground truth dataset, whose overlays are assumed to be optimal, according to [ICCÁ+15].

3.3.7   Experiments I: Performance Analysis of the DSS

The aim of this experiment is to determine the optimal design for our DSS in the identification task. To do so, we have to analyze different aggregation functions proposed for each level or sublevel. Then, the best of them is chosen based on the accuracy index by ranking the positive and negative cases. As a final step, a threshold has to be set for labeling each case as positive or negative.

The experimental dataset is divided into training and test sets to validate the system in a correct way. The 74.1% of the instances compose the training dataset. Thus, it is composed of 240 CFS cases, 16 positive (CBCTs 1 to 9 and 3D Models 10 to 16) and 224 negative ones. In terms of SFO level, this corresponds to 33 positive and 427 negative SFOs. Meanwhile, the remaining 25.9% of the cases form the test set, with 84 CFS cases, 8 positive (3D Models 1 to 8) and 76 negative ones. Correspondingly, this set has 129 cases at SFO level, 10 positive and 121 negative SFOs (3D Models 17 to 24). Fig. 31 summarizes the structure of the dataset and the experimentation procedure. Notice that, due to the intrinsic nature of CFS, we are dealing with a imbalanced classification problem [HG09].
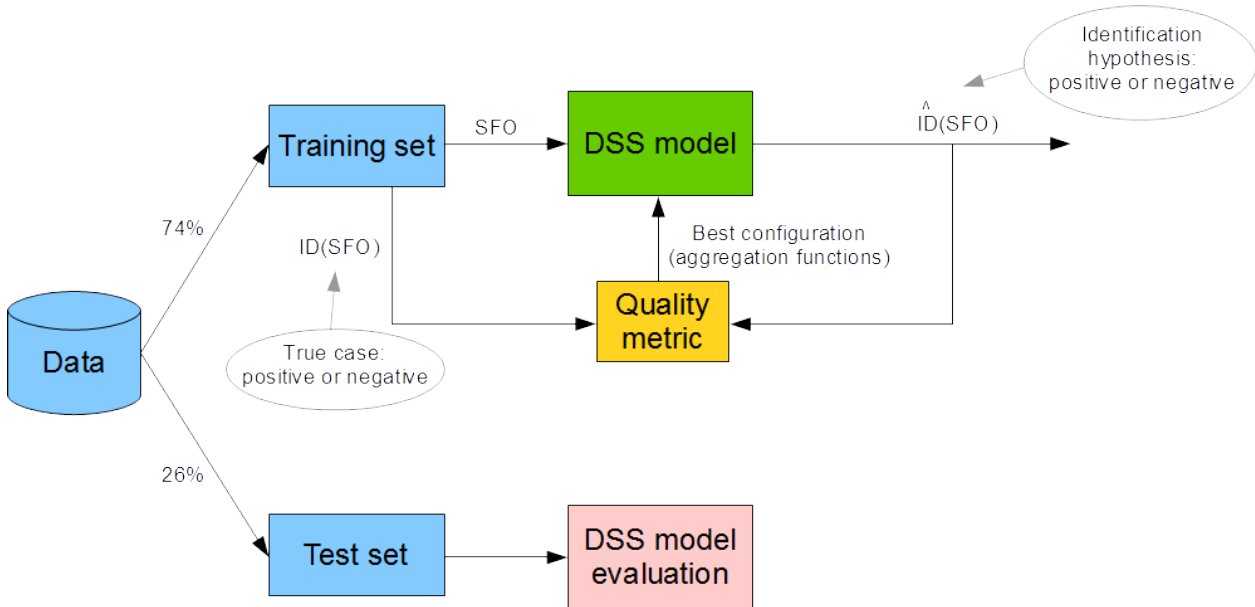

Figure 31: Experimentation setup diagram.

Following our DSS framework, the age and the BMI have to be set for each available photograph. Besides, the available regions of each photograph have to be marked and the related quality of each one established in a scale between 0 and 1. In this study, forensic experts must delineate in each image from one to eight regions in each image (depending on the visibility of them): chin contour, cranial contour, eyeball center right, eyeball center left, nose width, mouth occlusion, tragus right and tragus left; and from zero to 18 inter-dental lines. The corresponding 26 zones are also marked on the skull 3D models. Again, experts set the quality of the bone region based on the weathering stages. Then, the anthropometric landmarks used to achieve the SFOs are considered to form the other two criteria: ectocanthion lines and frontal central lines (both in the 3D model skulls and images). Figure 33 depicts an example of some of these regions and landmarks marked on a skull 3D model and on a photograph.

The experts established that the uncertainty sources have a third of influence and the matching degree have two thirds of influence. Thus, the weighted vector used in level 2.2 is $\mathbf{w} = (\frac{1}{3}, \frac{2}{3})$.

Finally, the discriminative power of each region has to be determined using the training dataset. This value is used as a weight for the aggregation at level 2.3. Then, with these parameters, we study the different aggregation functions mentioned in Section 3.3.3 for each sublevel of level 2, and in Section 3.3.4 for the level 1. Then, we need to study the behavior of our system using a threshold to give the final decision as a positive or a negative case. Finally, we validate the obtained results over the test dataset using the obtained DSS design and threshold.
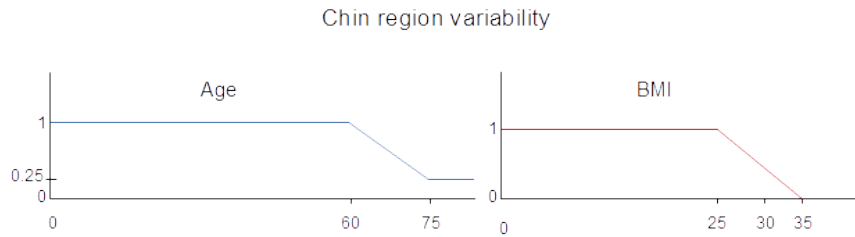
Figure 32: Defined fuzzy sets to model morphological aspects for the criteria. The left one refers to the age related confidence and the right one to the BMI related confidence of the chin region.

The influence related with the biological profile for the implemented criteria was defined by Prof. Wilkinson according to her expert knowledge and represented using the fuzzy sets in Fig. 32. As can be seen, the chin criterion is less reliable after becoming 60 years old, decreasing to 0.25 from 75. The same criterion is unreliable with BMI values above 35 (changes in fat will alter the shape of the chin). The remaining criteria are not affected by any morphological aspect.
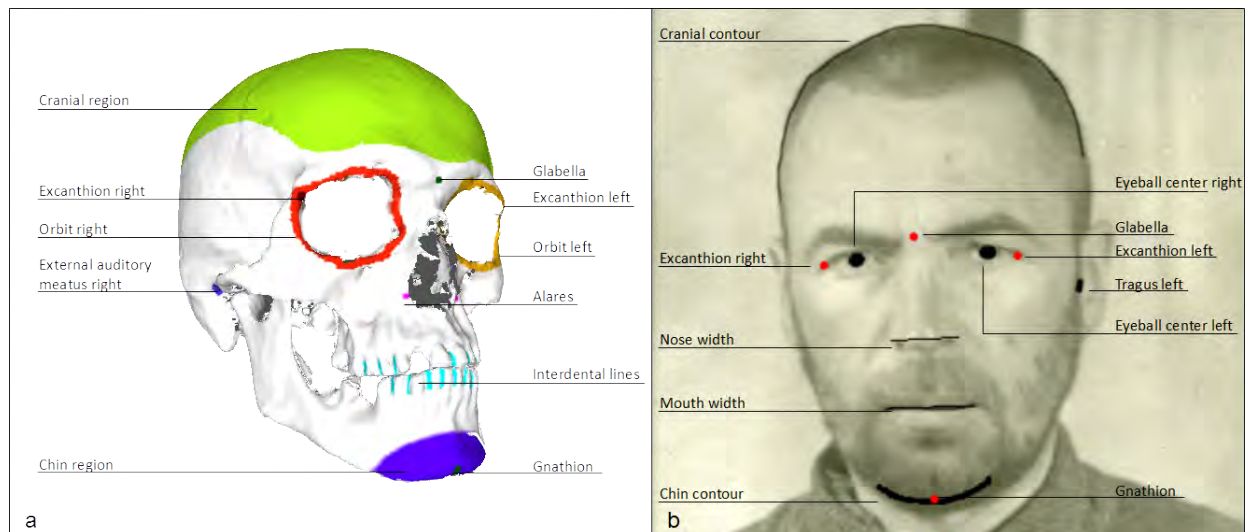
Figure 33: Example of marking regions on a skull 3D model (a) and on a photograph (b).

The training dataset is used to set the best configuration of our system. To do so, the following steps are performed:

1. To calculate the power of identification over the training dataset: this is understood as the capability of each isolated criterion to discriminate in the decision making process. Firstly, the matching degree of the craniofacial correspondence by only using a specific criterion is calculated over the database of cases. These values are used to rank the candidates from 1 to n based on their chance to be the actual subject. Then, a value between 0 and 1 is assigned to each positive case taking into account the ranking. The formula to assign the power of identification of criterion W_i in instance j is:

$$Power_{Id}(W_i)_j = 1 - \frac{r - 1}{M_j - 1} \tag{I.53}$$

where r is the position of the positive case in the ranking and M_j is the lowest value of the ranking for instance j (all cases getting the same criterion-method value are supposed to have a draw, that is, they are assigned to the same ranking). Finally, the average of these values

over all cases is calculated, getting the value $Power_{Id}(W_i)$ for each criterion over the whole database.

2. To analyze and compare the performance of different aggregation functions within our framework at the SFO level: in order to calculate the most suitable functions for our system, we compute the accuracy degree for identifying positive cases at the SFO evaluation level (level 2). The way to obtain the accuracy degree at SFO level follows the same methodology as before. First, all the aspects that are involved in the SFO evaluation are aggregated (quality of the materials, biological profile, craniofacial matching degree of each region, and power of identification), obtaining the degree of a SFO. We analyze the aggregation functions explained in Section 3.3.3. That is, for sublevel 2.1, the minimum (*min*), the product (*prod*), and the arithmetic mean (*mean*). For sublevel 2.2, the weighted arithmetic mean (*wam*) and the weighted geometric mean (*wgm*). Finally, for sublevel 2.3, the *wam*, the *wgm*, the Sugeno (*sug*) integral, and the Choquet (*choq*) integral [Sug77] as considered. Accordingly, 24 different combinations are analyzed overall. The obtained SFO degree is used to rank the candidates in decreasing order. Again, a value between 0 and 1 is assigned to each positive case taking into account this ranking. The formula to obtain the accuracy for each configuration of the system is:

$$ACC(Combination_a)_j = 1 - \frac{r-1}{M_j - 1} \tag{I.54}$$

where r is the position of the positive case in the ranking and M_j is the lowest value of the ranking for instance j. The final step calculates the average of these values over all the cases, $ACC(Combination_a)$. The parameters that obtain the highest accuracy will be selected for the configuration of the system.

3. To analyze and compare the performance of different aggregation functions within our framework at level 1: the final CFS degree is obtained by aggregating the degrees of all the SFOs that belong to the same case. As explained in section 3.3.4, we study four different operators at this level: mean, minimum, maximum, and weighted average by the number of regions. These SFO degrees are obtained using the best combination of operators from of the previous step. The accuracy degree at this level serves to identify the most appropriate operator following the same way as in the previous stage.

4. To establish a threshold for labeling each CFS case as positive or negative. The CFS degree is given in the interval $[0, 1]$. The threshold sets the limit to consider the case as negative (below this value) or positive (above this value). Four different scenarios can be featured taking into account the real nature of each case (positive or negative) as well as the label given by our system:

   (a) If the DSS labeled a positive case as positive, it is a true positive case (TP).

   (b) If the DSS labeled a positive case as negative, it is a false negative case (FN).

   (c) If the DSS labeled a negative case as negative, it is a true negative case (TN).

   (d) Finally, if the DSS labeled a negative case as positive, it is a false positive (FP).

   In order to measure the effectiveness of our system for each threshold value, an appropriate evaluation metric must be used. As said, the nature of the data is considered imbalanced since it exhibits an unequal distribution between the two classes, positive and negative identification. In fact, negative cases are significantly much more common than positive

cases. For this reason, we analyze the best threshold for the performance of our DSS using the G-Mean metric [HG09]:

$$G - mean = \sqrt{\frac{TP}{TP + FN} \times \frac{TN}{TN + FP}} \tag{I.55}$$

This metric evaluates the degree of inductive bias in terms of a ratio of positive and negative accuracy. The threshold which achieves the highest value of this metric will be considered the most appropriate for the performance of our system.

Finally, the test dataset is used to validate the system performance. The designed DSS is applied to this set of cases and the value of the $G - mean$ metric is computed as a final performance measure, together with some other imbalanced classification measures.

Thus, starting from point 1 of the previous list, the discriminative power for each criterion obtained using Eq. I.53 over the training dataset is shown in Table I.8. With these values, the next experiment consists of analyzing the behavior of the 24 combinations of the aggregation functions at level 2. To do so, we calculate the accuracy using the values of the 24 different configurations over all the training cases in Eq. I.54 as explained before.

Table I.8: The power of identification of each isolated region (Criterion level, level 3).

Region	Power of identification
Chin contour	0.75
Cranial contour	0.75
Eyeball (left and right) center position	0.72
Ear (left and right) position	0.50
Mouth occlusion	0.67
Nose width	0.83
Ectocanthion lines	0.60
Frontal central lines	0.54
Inter-dental lines overlap	0.33

Table I.9 details the mean accuracy for the DSS at level 2 using each combination of the aggregation functions in decreasing order of performance. The combination *mean-wgm-wgm* achieves the highest value, with 0.828 of accuracy. In general, we can see that the best aggregation functions for sublevel 2.3 are *wgm* and *wam*.

Therefore, we use the best configuration (*mean-wgm-wgm*) for the next experiments performed at level 1. The obtained results for the analysis of the four different ways to aggregate the SFO degrees are reported in Table I.10.

As can be seen, the best performance of the DSS is achieved aggregating the SFO degrees, using a weighted average by the number of regions considered. The accuracy obtained with this operator is 0.917 while the remaining functions do not reach a 0.9 of accuracy.

The latter configuration of the system is used to choose the most appropriate threshold for labeling each case as positive or negative. For the final identification task, ten different values are tested for the threshold: from 0.80 to 0.90, with steps of 0.01. Table I.11 summarizes the obtained results after applying the methodology explained before and calculating the $G - mean$ metric (Eq. I.55) for each threshold value.

As can be seen in Table I.11, a threshold of 0.86 achieves the highest $G - mean$ value. Concerning this result, our DSS correctly classified a relatively high number of positive (13 of 16, i.e. 81.25%) and negative cases (186 of 224, i.e. 83.03%). Therefore, the best performance of our identification system is reached using the configuration reported in Table I.12.

Table I.9: Mean accuracy of each combination method at SFO level (level 2).

Combination method	Mean accuracy
mean-wgm-wgm	**0.828**
mean-wam-wgm	0.825
mean-wam-wam	0.809
min-wgm-wgm	0.808
mean-wgm-wam	0.807
min-wam-wgm	0.807
prod-wam-wgm	0.805
prod-wgm-wgm	0.804
min-wam-wam	0.787
prod-wam-wam	0.787
min-wgm-wam	0.786
prod-wgm-wam	0.785
min-wgm-sug	0.718
min-wgm-choq	0.716
mean-wgm-choq	0.648
mean-wgm-sug	0.648
mean-wam-sug	0.644
prod-wgm-choq	0.627
mean-wam-choq	0.625
prod-wgm-sug	0.616
prod-wam-sug	0.616
min-wam-sug	0.613
prod-wam-choq	0.609
min-wam-choq	0.609

The test dataset is used to validate the former configuration of our DSS. Table I.13 details the obtained results after applying the system over the set of test cases. Apart from reporting the G-Mean value, the accuracy of the system is calculated using Eq. I.56 [HG09]:

$$Accuracy = \frac{TP + TN}{P_c + N_c}, \tag{I.56}$$

where P_c and N_c are the total number of positive and negative cases, respectively; and the rate of true positives and true negatives are defined as:

$$TP_{rate} = \frac{TP}{P_c}; \qquad TN_{rate} = \frac{TN}{N_c}. \tag{I.57}$$

3.3.8 Experiments II: Comparison between the Automatic DSS and a Manual Approach Performed by Experts

A second experiment has been developed in order to compare the performance of real forensic practitioners with that of our CFS DSS. To do so, our system is applied to the same experimental dataset of [IVN+15]. In that study, 26 participants from 17 different institutions were asked to deal with 14 identification scenarios, some of them involving the comparison of multiple candidates and unknown skulls. A total number of 60 SFO problems were tackled. Table I.14 shows the mean value of the results of the 26 experts, the results of the three best experts, and the outcomes of

Table I.10: Mean accuracy of the DSS using each aggregation operator at CFS level (level 1).

Aggregation operator	Mean accuracy
Weighted average by the number of regions	**0.917**
Mean	0.893
Minimum	0.860
Maximum	0.839

Table I.11: Results of the DSS for different values of the threshold.

Threshold	TP	FN	TN	FP	*G-Mean*
0.80	13	3	153	71	0.745
0.81	13	3	156	68	0.752
0.82	13	3	158	66	0.757
0.83	13	3	161	63	0.764
0.84	13	3	166	58	0.776
0.85	13	3	180	44	0.808
0.86	13	3	186	38	**0.821**
0.87	12	4	190	34	0.798
0.88	11	5	200	24	0.783
0.89	10	6	203	21	0.753
0.90	7	9	214	10	0.646

Table I.12: Configuration for the best performance of our automatic DSS.

Framework level	Task	Operator
Level 2.1	Aggregate the sources of uncertainty of bone and image and biological profile	Mean
Level 2.2	Aggregate the skull-face matching degree and the uncertainty of level 2.1	Weighted geometric mean
Level 2.3	Aggregate all level 2.2 degrees weighting by the discriminative power	Weighted geometric mean
Level 1	Aggregate all the SFO degrees of the same case	Weighted average by number of regions
Threshold	Classify each case as negative or positive	0.86

our automatic DSS. Note that expert 2 and 3 did not complete the whole study, leaving one and two positive cases without decision, respectively. Detailed performance indicators are shown such as the percentage of correct decisions, the number of positive and negative decisions given in each case, and the corresponding rates of TP, FP, TN, and FN. 'Ground Truth' refers to the actual decision of each CFS case, i.e., positive when the skull and the facial photographs belong to the same person, negative in the contrary case. DSS-0.86 refers to our DSS method using the best configuration set and validated in the previous experiment.

In view of these results, and according to the conclusions from [IVN+15], experts generally achieve higher rates of TN than TP. Table I.14 shows that the mean TP rate was 52.63% while the mean TN rate was 84.20%. We can observe that the three best participants achieved TN rates equal or higher than 90%. Meanwhile, the same three experts achieved 80.00%, 66.67%, and 62.50% of TP rates, respectively. The overall accuracy (total correct decisions) is 78.99% in average, and

Table I.13: Performance of the DSS over the test dataset.

TP	FN	TN	FP	TP_{rate}	TN_{rate}	Accuracy	*G-Mean*
5	3	74	2	0.625	0.974	0.940	0.780

Table I.14: Performance of the different CFS approaches for test 2

Method	Correct Decisions	Ground Truth	Decision		Decision(%)	
			Positive	Negative	Positive	Negative
Experts Mean	78.99%	Positive	100	90	52.63%	47.37%
		Negative	152	810	15.80%	84.20%
Best Expert 1	93.33%	Positive	8	2	80.00%	20.00%
		Negative	2	48	4.00%	96.00%
Best Expert 2	88.14%	Positive	6	3	66.67%	33.33%
		Negative	4	46	8.00%	92.00%
Best Expert 3	86.21%	Positive	5	3	62.50%	37.50%
		Negative	5	45	10.00%	90.00%
DSS-0.86	90.00%	Positive	6	4	60.00%	40.00%
		Negative	2	48	4.00%	96.00%

93.33%, 88.14%, and 86.21% for the three best participants, respectively. The overall performance (correct decision rate) of our system is better than the mean of the experts with respect to the correct decisions rate (90.00%) and better than 25 out of the 26 forensic experts who participated in the studio. The TN rate is also very good (96.00%), exactly the same that the best performing expert. Meanwhile, the TP rate is not so high (60.00%) but still above the experts mean and really similar to that of the expert performing the third. One of the advantages of our automatic DSS tool is that classifies all cases on the contrary of some human experts.

3.3.9   Experiments III: Performance Analysis of the DSS as a Sort-Listing Tool

The final experiment of this work aims to study the capability of our system to identify the correct individual between a list with one positive and several negative cases. Note that for this analysis, the threshold value is not needed since we use the CFS degree to rank the candidates. Fig. 34 shows the CMC curve of the system (red line). As we can see, our tool is able to rank more than 60% of all the correct individuals in the first position of the list (the 4% of the total samples) and more than 70% in the two first positions. This percentage increase to more than 80% for rank 3, which corresponds with the 11% of the total cases of the data set.

In addition, we analyze the performance of the system ignoring the cases with only one image. Nowadays, the most common scenario is that the relatives of a missing person have a lot of AM photographs of him/her. We want to study the influence of taking into account only the cases of our data set with more than one SFO. Fig. 34 shows also the results of applying the DSS in this case, regardless of the number of available photos per case. As can be seen, our system is still not able to identify correctly every case (rank 1 values are still in the 60% range). However, the behavior improves for the sort listing task. It ranks all the positive cases between the six first positions (51% of the data set). Meanwhile, using all the cases the system has to reach the 60% of the samples to get the same value.
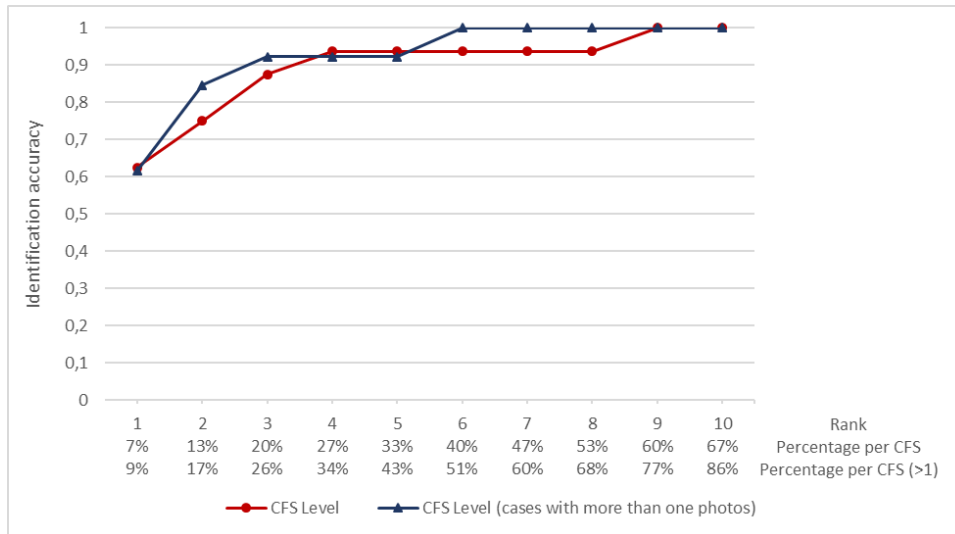
Figure 34: CMC curves corresponding to the experimentation at CFS level with all the cases and only with the cases with more than one photographs.

4 Summary and Discussion of the Obtained Results

The current section summarizes the main results obtained in this PhD dissertation with a specific subsection for each of them.

4.1 Skull-face Overlay Stage: Modeling the Landmark Matching Uncertainty and Fuzzy Distances Study

Our automatic procedure, based on CV and SC, performs SFO in the most natural way by replicating the original scenario in which the photograph was taken. Our technical procedure for SFO automates this error-prone and repetitive approach. The orientation of the skull and its size adjustment with the face in the photograph can be performed in an automatic fashion simplifying the work of the experts. That strongly reduces the time required to get a valid superimposition and automates the application of the method.

The performance of our automatic SFO method (see Section 3.2.1) and the two previous ones [ICDS11, ICD12] that do not include the *landmark matching uncertainty* treatment have been objectively evaluated considering the ground truth dataset described in Section 2.4.4 in 18 case studies. The different variants of our method achieve a competitive matching between pairs of corresponding cranial and facial landmarks because of the natural modeling of the landmark correspondences using SC techniques. In particular, RCGA-c-45 significantly outperforms the remaining automatic SFO methods tested. It obtains the best performance regarding the ground truth in most of the cases. The RCGA-c-30 variant also achieves a good ranking in the computed statistical analysis (in fact, it is the same method with different parameters). Hence, the new proposals that model the *landmark matching uncertainty* get better results than the previous approaches that do not consider the facial soft tissue thickness.

Nevertheless, although really good results are obtained, it is important to note that a distance or residual error is still identified when evaluating a SFO result against its corresponding ground truth overlay.

On the other hand, 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 independent runs.

This distance function between pairs of landmarks only considers 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 must satisfy a specific position criteria depending on the pose. We plan to carry out this extension as a future work.

4.2 Decision Making Stage: Hierarchical Decision Support System Development

The complete framework proposed for a DSS in CFS develops information fusion concerning skull-face anatomical correspondence at three different levels: criterion evaluation, SFO evaluation, and CFS evaluation. We classify the uncertainty sources and degrees of confidence involved in this process as related to bone, image, skull-face overlays, morphological aspects, and used methods.

Concerning the criterion evaluation stage, experts distinguish four different families for assessing craniofacial correspondence: 1) consistency between cranial and facial outlines; 2) anatomical consistency by positional relationship; 3) line location and comparison; and 4) consistency of the soft tissue thickness between corresponding cranial and facial landmarks. In this work, we have studied and modeled several skull-face morphological criteria which belong to the firsts three groups: chin and cranial contour (first family); cranial orbit and eyeball center position, upper rim and ear position, mouth occlusion length, nose width, overlap of inter-dental lines (second family); and ectocanthion lines and frontal and central lines parallelism (third family). In this way, the craniofacial correspondence evaluation comprises the whole face. Thus, we have considered from one to eight regions and from zero to 18 inter-dental lines. Using CV techniques, we have studied different methods to measure the consistency between each bony and facial region. In some of them, the best performance is obtained by combining two or three techniques thanks to the use of the Sugeno integral.

Within the SFO evaluation stage, we distinguish three sublevels with different conditions of information fusion. For this reason, we perform an experimental study to analyze which aggregation function provides the best results. The first sublevel aggregates the sources of uncertainty of the bone and the image, and the biological profile variability. For this sublevel, we propose to use the minimum (*min*), the product (*prod*), and the arithmetic mean (*mean*) as aggregation functions. The second sublevel consists of integrating this aggregation with the matching degree of the skull and the face. In this case, we study the use of the weighted arithmetic mean (*wam*) and the weighted geometric mean (*wgm*). Finally, at the third sublevel, we obtain the degree of the SFO craniofacial correspondence by aggregating the different previous values for all the regions taking into account the discriminative power of the isolated region as weight. To perform this aggregation, we test the *wam*, the *wgm*, the Sugeno (*sug*) integral, and the Choquet (*choq*) integral. Thus, at the CFS evaluation level, we analyze the aggregation of the different SFO degrees using the following four operators: the mean, the maximum, the minimum, and a weighting function based on the number of regions analyzed in each SFO. In view of the results obtained in the experiments, we conclude that the best current performance at level 2 is obtained using the combination of aggregation functions *mean-wgm-wgm*. Then, the best behavior of the system is achieved when the different SFO degrees of the same case are aggregated by weighting by the number of regions considered in each SFO. Finally, we have studied several thresholds for establishing a binary classification (positive or negative), obtaining 0.86 as the best value. Concerning this result, our DSS correctly classified a relatively high number of positive (13 of 16, i.e. 81.25%) and negative cases (186 of 224, i.e. 83.03%). These parameters have been achieved over a training data set. The performance of the system has been evaluated using a different unseen set of cases (test data set), obtaining a

worse rate in positive cases (62.5% of TP) but a very good behavior in negatives ones (97.4% of TN).

From the analysis of the experimentation developed we can conclude that our current system is suitable for:

- Filtering (sort-listing) cases: in 62.5% of the occasions the positive case is ranked in the first position. Besides, given a data set of initial candidates, the positive case was included in the 60% best ranked candidates with a 100% of probability and within the 27% best ranked candidates with a 92% of probability.

- Establishing exclusion: the ability to determine a negative identity was performed with a 97.4% of accuracy (97.4% of TN rate over the test data set and 83.03% over the training data set).

Thus, the designed DSS can be considered the first automatic tool for classifying couples of unknown faces and skulls as positive or negative cases with an accuracy similar to that of the best performing forensic experts.

5 Final Conclusions and Future Works

In this PhD dissertation, we have improved an existing computer-aided CFS system in order to extend its functionality for developing a more reliable and robust tool. We have used SC and CV techniques as well as anatomical forensic studies.

In particular, we have developed an automatic superimposition method, based on EAs and fuzzy set theory, which models the imprecision related to the matching of cranial and facial landmarks, i.e. the soft tissue depth distances between a pair of cranial and facial landmarks.

Besides, we have performed a deep study for applying the most appropriate fuzzy distance operators in this method in order to obtain the best possible superimposition.

On the other hand, we have also present a hierarchical framework for information fusion to support the anthropologist expert in the decision making stage. The automation of this process is based on the use of several skull-face anatomical criteria combined at different levels by means of fuzzy aggregation functions. For the first time, we present a complete system which provides a final degree of craniofacial correspondence.

In addition, we have validated our system as an automatic identification tool analyzing its capabilities in closed and open lists and comparing its performance with the manual results achieved by forensic experts, obtaining a remarkably good performance. The proposed system has demonstrated to be valid for sort-listing a given data set of initial candidates (in 62,5% of the decisions the positive case is ranked in the first position) and to serve as an exclusion method (97,4% and 83.03% of TN rate in training and test, respectively).

In general, the results obtained show that we have reached the global aim. We also achieved the different sub-objectives defined at the beginning of this PhD dissertation such as the performance improvement in the automatic SFO method. The good outcomes obtained are reflected in four scientific journal publications, another one recently submitted [CAMFI+17], and several conference papers. In addition, other two journal articles have been published due to the participation in the MEPROCS project [IVC+16, IVN+16].

Even so, we cannot still consider CFS as a solid identification technique. Although CFS has been applied and developed for more than a century, its reliability is still unclear. On the one hand, it is essential to work with large datasets to reinforce the conclusions of our approach. Enlarging the size of the dataset presents many opportunities to generalize the process as well as increasing the likelihood of producing consistent, accurate, and reproducible results. On the other hand,

an objective and more precise automatic system, which considers more information from different sources, is necessary. In this work, there are some types of uncertainty sources and degrees of confidence which are still ignored, such as the quality of the SFOs achieved in the second stage (highly improbable); the 3D and 2D regions delimitations; and the number of regions evaluated. Although we have taken into account up to nine regions, some studies suggest the importance of evaluating even more. For instance, authors in [YIMS95] use more than 13 criteria to make identification decisions and a similar approach is followed in [CLT$^+$92].

We are aware that the quality of the SFOs has a strong influence in the performance of the DSS. For this reason, future works will be addressed in order to enhance this quality: a new parametrization of the camera and a innovative design of the optimization algorithms for the SFO stage. Related to this, the modeling of the mandible articulation is crucial for improving the overlays obtained by the superimposition algorithm. Furthermore, developing new accurate soft tissue studies, by including spatial directions and a robust statistic metric, could increase the performance of the method. Besides, a feature selection regarding the criteria can be done in order to improve the classification of our system. This feature selection could be apply over different age, genre, ancestry groups or even frontal, lateral and oblique poses.

6   Conclusiones Finales y Trabajos Futuros

En esta tesis doctoral se ha ampliado la funcionalidad del procedimiento de superposición cráneo-facial creando un método más automático, fiable y robusto. Para llevar a cabo este objetivo se ha hecho uso de técnicas de SC, visión por computador y estudios anatómicos forenses.

En concreto, se ha diseñado un nuevo método automático de solapamiento cráneo-cara que incorpora los estudios disponibles de las distancias de tejido blando entre los *landmarks* faciales y craneales mediante algoritmos evolutivos y lógica borrosa.

También se ha llevado a cabo un estudio en profundidad de las métricas más adecuadas para este método con el objetivo de obtener las mejores superposiciones posibles.

Por otro lado, hemos propuesto un marco de trabajo jerárquico de ayuda al forense mediante el cual se integra toda la información disponible en la etapa de la toma de decisiones. La automatización de este proceso se basa en el uso de distintos criterios anatómicos del cráneo y la cara combinados a diferentes niveles mediante funciones de agregación. Por primera vez, se presenta un sistema completo que proporciona como resultado un grado de correspondencia cráneofacial.

Además, se ha validado el sistema como herramienta autómatica de identificación analizando sus capacidades en casos abiertos y cerrados y comparando su funcionamiento con el de expertos forenses, obteniendo muy buenos resultados. El sistema propuesto ha demostrado ser válido para filtrar listas de posibles candidatos (el 62,5% de los casos positivos se asignaron a la primera posición) así como para ser usado como método de exclusión (97,4% y el 83.03% de verdaderos negativos en los datos en entrenamiento y test, respectivamente).

En general, los resultados obtenidos demuestran que hemos alcanzado el objetivo pricipal de la tesis. También se han alcanzado los distintos sub-objetivos definidos al inicio de la misma. El éxito del trabajo realizado se refleja en cuatro publicaciones en revistas científicas, otro artículo recientemente enviado [CAMFI$^+$17] y varias contribuciones en congresos. Además, han sido publicados otros dos artículos gracias a la participación en el proyecto MEPROCS [IVC$^+$16, IVN$^+$16].

Aún así, la superposición craniofacial todavía no puede considerarse todavía una técnica de identificación sólida. Aunque lleva aplicándose y desarrollándose durante más de un siglo, su fiabilidad no está todavía clara. Por un lado, resultaría esencial trabajar con conjuntos de datos con un mayor número de casos para poder reforzar las conclusiones de nuestro sistema. Aumentar el tamaño de los conjuntos de datos utilizados presentaría muchas posibilidades para

generalizar el proceso y aumentaría la probabilidad de producir resultados consistentes, precisos y reproducibles. Por otro lado, sería necesario un sistema automático más objetivo y fiable que considere más información de diferentes fuentes. En este trabajo, aún se ignoran varios tipos de fuentes de incertidumbre y de grados de confianza, como la calidad de las superposiciones obtenidas en la segunda fase, las delimitaciones de las regiones 3D y 2D y el número de regiones evaluadas. A pesar de que se han tenido en cuenta hasta nueve regiones, algunos estudios sugieren la importancia de evaluar aún más. Por ejemplo, en [YIMS95] se usan más de 13 criterios para tomar la decisión de identificación final, al igual que en [CLT$^+$92] donde se sigue un planteamiento similar.

Somos conscientes de que la calidad de las superposiciones influye significativamente en el funcionamiento de nuestro sistema de ayuda a la decisión. Por esta razón, los trabajos futuros van encaminados a mejorar esta calidad mediante una nueva parametrización de la cámara y un diseño innovador de los algoritmos de optimización en la fase del solapamiento cráneo-cara. En este aspecto, también es crucial el modelado de la articulación de la mandíbula con el objetivo de mejorar las superposiciones obtenidas con nuestro algoritmo. Además, el desarrollo de nuevos estudios de los tejidos blandos entre cráneo y la cara que incluyan la dirección espacial y una métrica estadística robusta, podrían incrementar la funcionalidad del método. Por último, se podría realizar una selección de características para mejorar la clasificación de nuestro sistema. Esta selección podría aplicarse diferenciando grupos de edad, género, etnia o incluso poses frontales, laterales y oblicuas de los casos.

Part II. Publications

This chapter presents all the scientific papers published. Together they show the work carried out to achieve the stated objectives in this PhD dissertation. Four sections present each of the contribution developed.

1 Modeling the Soft Tissue Thickness for Automatic Skull-Face Overlay

- B. R. Campomanes-Álvarez, O. Ibáñez, C. Campomanes-Álvarez, S. Damas, and O. Cordón. Modeling the Soft Tissue Thickness for Automatic Skull-Face Overlay, IEEE Transactions on Information Forensics and Security, vol. 10, no 10, pp. 2057 - 2070, 2015. DOI: 10.1109/TIFS.2015.2441000.

  - State: Published.
  - Impact Factor (JCR 2015): 2.441.
  - Category: COMPUTER NETWORKS AND COMMUNICATIONS. Order: 11/203. Q1.

Modeling Facial Soft Tissue Thickness for Automatic Skull-Face Overlay

B. Rosario Campomanes-Álvarez, Oscar Ibáñez, Carmen Campomanes-Álvarez, Sergio Damas and Oscar Cordón

Abstract—**Craniofacial superimposition involves the process of overlaying a skull with a number of ante-mortem images of an individual and the analysis of their morphological correspondence. Within the craniofacial superimposition process, the skull-face overlay stage focuses on achieving the best possible overlay of the skull and a single ante-mortem image of a missing person. This technique has been commonly applied following a non automatic trial-and-error approach. Automatic skull-face overlay methods have been developed obtaining promising results. In this work, we present two new variants that are an extension of existing 3D-2D methods to automatically superimpose a skull 3D model on a facial photograph. We have modeled the imprecision related to the facial soft tissue depth between corresponding pairs of cranial and facial landmarks which typically guide the automatic approaches. As an illustration of the model's performance, the soft tissue distances associated to studies for Mediterranean population have been considered for dealing with this landmark matching uncertainty. Hence, we directly incorporate the corresponding landmark spatial relationships within the automatic skull-face overlay procedure. We have tested the performance of our proposal on 18 skull-face overlay instances from a ground truth dataset obtaining valuable results. The current proposal is thus the first automatic skull-face overlay method evaluated in a reliable and unbiased way.**

Index Terms—**forensic anthropology, craniofacial identification, craniofacial superimposition, 3D-2D skull-face overlay, soft tissues, fuzzy sets and distances.**

I. Introduction

THE techniques used to identify a missing person from skeletal information have been under continuous investigation within forensic anthropology [1], [2]. Craniofacial superimposition (CFS) [3], one of the approaches in craniofacial identification [4], [5], involves superimposing a skull onto a number of ante-mortem images of an individual and the analysis of their morphological correspondence.

Three consecutive stages for the whole CFS process have been distinguished in [6]: 1) Acquisition and processing of the skull and the ante-mortem facial photographs, together with

B. R. Campomanes-Álvarez, O. Ibáñez, S. Damas, and O. Cordón are with the Department of Fuzzy Evolutionary Applications, European Centre for Soft Computing, Mieres Asturias 33600 Spain (e-mail: rosario.campomanes@softcomputing.es; oscar.ibanez@softcomputing.es; oscar.cordon@softcomputing.es; sergio.damas@softcomputing.es).

O. Ibáñez, C. Campomanes-Álvarez, and O. Cordón are with Department of Computer Science and Artificial Intelligence, University of Granada, Granada 18014 Spain (e-mail: oscar.ibanez@decsai.ugr.es; carmen.campomanes@decsai.ugr.es, ocordon@decsai.ugr.es).

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

the location of anatomical landmarks on the skull and the face; 2) Skull-face overlay (SFO), which focuses on achieving the best possible superimposition of a skull (or a skull 3D model) and a single ante-mortem image of a missing person. This process is iteratively repeated for each available photograph, obtaining different overlays. Skull-face overlay thus refers to what traditionally has been known as the adjustment of the skull size and its orientation with respect to the facial photograph [3]. It is the most time consuming stage of the whole CFS procedure. 3) Decision making where the degree of support shows that the skull and the available photograph belong to the same person or not (exclusion). This task requires a thorough analysis of the face/skull correspondence provided by SFO (second stage) to determine if the skull and the face actually belong to the same person. Currently, decision making in CFS must be manually taken by the human expert, a forensic anthropologist. This decision is guided by different criteria involving 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.

An important limitation of the CFS technique is the absence of a commonly accepted methodology[1]. Experts try to solve the CFS problem by applying a specific strategy which uses their knowledge and the available technologies. During the SFO stage, the focus of this contribution, most forensic anthropologists follow a trial-and-error approach until they attain a good enough superimposition. The appropriate projection of the skull onto the facial photograph is a very challenging and time-consuming part of the CFS technique. This task involves the estimation of the best correspondence and it can take hours to arrive at the best possible fit [7], [8]. In addition, an inherent uncertainty exists because of overlaying two different "objects" (a skull and a face) [9]. Hence, a systematic and automatic SFO method is a real need in forensic anthropology [8].

Computational methods in the fields of computer vision (CV) and soft computing (SC) can be extremely useful of this proposal. Computer vision includes techniques for processing, analyzing, segmenting and registering image data in an automatic way [10]. 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) [11]. Meanwhile, SC

---

[1]There is a European project, entitled "New Methodologies and Protocols of Forensic Identification by Craniofacial Superimposition" (MEPROCS), which aims to develop a common methodology for the application of CFS. The interested reader is referred to http://www.meprocs.eu/

aims to design intelligent systems that process uncertain, imprecise and incomplete information [12]. Two of the main SC techniques are fuzzy logic (FL) [13] and evolutionary algorithms (EAs) [14]. The former extends classical logic to provide a conceptual framework of knowledge representation under imprecision and the consequent uncertainty. The latter combines powerful bio-inspired search and optimization tools to automate problem solving in areas such as modeling, simulation or global optimization [14]. Specifically, fuzzy sets have largely demonstrated their capability to deal with vagueness and imprecise information [13].

Our previous work tackles SFO automatically using EAs and fuzzy sets [9], [15]–[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 [18], [19]. This minimization process involves the search for the specific projection of the skull 3D model that reduces all the distances between every pair of corresponding landmarks as much as possible. This is a good approximation to deal with the problem, which provides reasonable results but it is not anatomically correct. In fact, the anatomical distance between a cranial landmark and its corresponding facial point (soft tissue depth) is not considered. In reality, the thickness of the facial soft tissue differs for each corresponding pair of landmarks, varies among individuals and produces a mismatch among cranial and facial landmarks [5], [20]. Furthermore, another drawback of our proposals is that they have only been validated in a subjective way, based on the visual evaluation of the superimposition results by forensic experts. This is due to the lack of an objective assessment methodology in the area.

In this contribution, our proposal deals with the automation of the SFO task that is focused on the projection of a 3D skull model over a 2D photograph. We present an automatic 3D-2D SFO method which considers the imprecision related to the matching of landmarks in the skull and face. Thanks to fuzzy sets [21], we have modeled the soft tissue thickness between pairs of cranial and facial landmarks. This novel proposal uses the same optimization mechanism employed before [9], [16] but it changes the formulation of the problem to directly incorporate the spatial relationships between corresponding landmarks. Unlike traditional SFO approaches locating tissue depth markers on the physical or the 3D model skull [20], [22], our proposal allows us to incorporate any soft tissue study easily.

In addition, we have performed an objective and quantitative evaluation of the SFO results. We included the soft tissue depth measurements of specific studies [23]–[25] within the SFO task. We tested the performance of our approach on 18 SFO problem instances that belong to the first and unique ground truth dataset [26]. This allows us to evaluate our automatic SFO method following an unbiased and reliable procedure. To our knowledge, the current proposal is the first automatic SFO procedure that incorporates forensic studies of inter-landmark distances and considers a ground truth dataset to validate the results.

The paper is organized as follows. Section II reviews the related work and the problem statement. Section III describes our proposals. Section IV presents the experiments and results. The discussion and conclusions are detailed in Section V.

II. BACKGROUND

A. Craniofacial Superimposition

Craniofacial superimposition [3] is one of the approaches in craniofacial identification [4], [5]. It involves the superimposition of a skull with a number of ante-mortem images of an individual and the analysis of their morphological correspondence. Since the first documented use of CFS for identification purposes [27], the technique has been under continuous improvement. Although the foundations of the CFS method were laid at the end of the ninetieth century [28], [29], the associated procedures evolved as new technology became available. Therefore, three different main approaches have been developed: 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) [3], [4], [30].

The role of computerized systems in CFS is very important nowadays. Thus, special attention has been given to them in the more recent surveys in the field [6], [31]. In particular, they highlight the importance of stating the difference between non automatic and automatic computer-aided methods. Among the latter group, a reduced number of proposals deal with the same problem tackled in this manuscript: the SFO task. They obtain unbiased results and drastically reduce the time taken for SFO. These proposals are based either on photograph to photograph comparison [32] or on skull 3D model to photograph comparison [9], [15], [16], [30], [33].

B. Skull-Face Overlay as a Computer Vision Problem

Skull-face overlay requires positioning the skull in the same pose as the face in the photograph. From a purely 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 [34]. In such a scene, the living person was somewhere inside the camera's field of view in a given pose.

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 used. Current 3D scanners provide skull 3D models with a precision of less than one millimeter in a few minutes [35]. These models can be properly handled using the computer, making the automation of the SFO task easier. 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. In addition, the specific characteristics of the camera must also be replicated to reproduce the original situation as much as possible.

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 at the same angle as the face in the photograph [3], [7]. Then a perspective projection of the skull 3D model is placed 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, that process directly replicates the original scenario in which the photograph was taken [15], [36].

C. 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 to the skull 3D model and the facial photograph.

Hence, given two sets of cranial and facial landmarks, $C = \{cl^1, ..., cl^n\}$ and $F = \{fl^1, ..., fl^n\}$, the process has to solve a system of equations with 12 unknowns [15]: 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 using 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 angle of view ϕ. These parameters determine the geometric transformation f that projects every cranial landmark cl^i of the skull 3D model onto its corresponding facial landmark fl^i of the photograph:

$$F = C \cdot R \cdot S \cdot T \cdot P \qquad (1)$$

The rotation matrix R turns the skull to the same pose as the head in the photograph. S, T, and P are scaling, translation and perspective projection matrices respectively [15]. A complete description of the matrices of Eq. (1) is detailed in [37].

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 mean error (ME) fitness function [15]:

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

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 onto the photograph, d is the 2D Euclidean distance, and N is the number of landmarks placed by the expert.

D. Modeling the Uncertainty Related to the Location of Facial Landmarks

The uncertainty related to the location of facial landmarks refers to the difficult task of placing landmarks on a photograph [18], [19]. Among other reasons, the definition of many anthropometric landmarks is imprecise in nature.

Using precise landmarks, forensic anthropologists can only place the facial landmarks that they clearly identify in the facial photograph. The fuzzy approach developed in [9] 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 these landmarks are employed. This leads to a better description of the skull-face correspondence thanks to the new
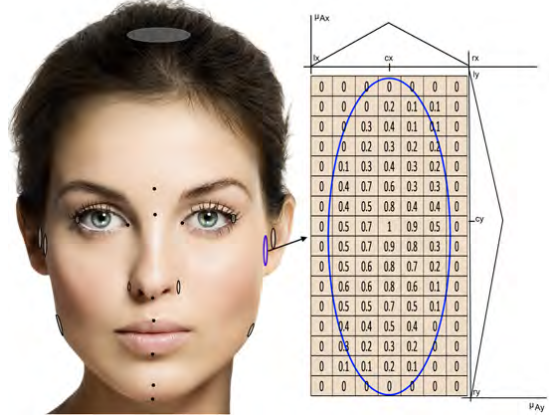

Fig. 1: Imprecise location of facial landmarks (left) and representation of a fuzzy landmark with fuzzy sets (right).

pairs of cranial points and fuzzy landmarks in the face. The performance of the automatic SFO method is thus improved.

As a consequence of handling these fuzzy landmarks, modeled as bi-dimensional fuzzy sets (see Fig. 1), the computation of the distances between the corresponding cranial and facial landmarks in Eq. 2 is affected as follows:

Distance from a point x to a fuzzy landmark $\tilde{F}$:

$$d'(x, \tilde{F}) = \frac{\sum_{k=1}^{m} d(x, \tilde{F}_k) \cdot \alpha_k}{\sum_{k=1}^{m} \alpha_k}, \qquad (3)$$

where x is a precise point, $\tilde{F}_k$ is the k_{th} element (pixel) of the fuzzy landmark $\tilde{F}$, d is the 2D Euclidean distance, α_k is the membership value of $\tilde{F}_k$, and m is the number of elements of the fuzzy landmark $\tilde{F}$.

The interested reader is referred to [9] for a more in-depth explanation and an example of the calculation of the distance between a point and a fuzzy landmark.

Considering this latter distance (Eq. 3), the definition of the EA fitness function (Eq. 2) is modified as follows [9]:

$$fuzzyME = \frac{\sum_{i=1}^{N} d'(f(cl^i), \tilde{F}^i)}{N}, \qquad (4)$$

where $f(cl^i)$ represents the 2D position of the 3D cranial landmark when projected onto the photograph, $\tilde{F}^i$ represents the fuzzy landmark, and $d'(f(cl^i), \tilde{F}^i)$ is the distance between a point and a fuzzy landmark (Eq. 3), and the remaining parameters are those in Eq. 2.

III. INCORPORATING FACIAL SOFT TISSUE MODELING TO THE SKULL-FACE OVERLAY METHOD

Facial soft tissue depth varies for each landmark correspondence and for different groups of people. Some facial and cranial landmarks show a very close relationship as glabella, dacryon and frontotemporale. Meanwhile, others do not exactly overlap because of varying thicknesses in the soft tissue between them, e.g. gnathion, zygion and alare [20]. This

Fig. 2: Mean soft tissue depths (in mm) and spatial relationship between cranial and facial landmarks [23], [24].
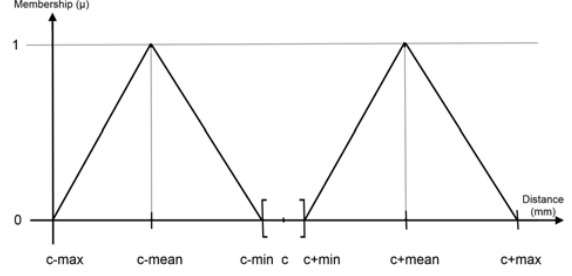
.

Fig. 3: Graphical representation of one of the dimensions of the 3D fuzzy set $\tilde{B}_p$ for modeling the landmark matching uncertainty using a sphere. The other two dimensions are modeled in a homologous way.

variability has been widely studied in many populations and considering different age and gender subgroups [23], [24].

Our proposal goes beyond the use of tissue depth markers [22] as it directly incorporates the corresponding landmark spatial relationships and distances within the automatic SFO procedure (Fig. 2). To do this, we model the minimum (*min*), mean (*mean*) and maximum (*max*) distances between a pair of cranial and facial landmarks using fuzzy sets (what we defined as landmark matching uncertainty). These distances are obtained from any study looking at the specific population group considered.

In this section, we present two alternative approaches that deal with the landmark matching imprecision in SFO.

A. Modeling the Landmark Matching Uncertainty Using Spheres

Our first approach consists of building a fuzzy set whose center is the 3D cranial landmark with a membership degree of zero. The rest of the points are calculated adding the positive and negative values of the *min*, *mean* and *max* distances to the 3D cranial landmark position along the three axes (X, Y and Z). We thus define a 3D volume (sphere) in the space where each facial landmark is expected to be located according to a particular soft tissue depth study (Figs. 2 and 4a).

We have chosen a triangular fuzzy set to define the masks for handling the landmark matching uncertainty. This choice is motivated by the large amount of literature available on the topic. Although different membership function shapes can be considered, both piece-wise (e.g. trapezoidal-shaped) and continuous (e.g. Gaussian), piece-wise triangular fuzzy membership functions are simpler and easier to handle. In addition, some works have proven that a linear piece-wise fuzzy membership function can approximate a continuous function to the desired degree, achieving similar results [38].

Formally, we handle the landmark matching uncertainty using 3D masks represented as a matrix M with $m_x \times m_y \times m_z$ points. These masks are defined by three triangular fuzzy sets $\tilde{B}_x$, $\tilde{B}_y$, and $\tilde{B}_z$, which model the approximate vertical, horizontal, and depth position of the sphere that represents

the place of each facial landmark in relation with its cranial counterpart. They thus become 3D fuzzy sets, where each triangular fuzzy set $\tilde{B}_p$, with $p \in \{x, y, z\}$, is defined by its center $c \in \{c_x, c_y, c_z\}$ (the 3D coordinates of the cranial landmark) and the *min*, *mean*, and *max* distances as follows (see Fig. 3):

$$
\tilde{B}_p = \begin{cases} 1 - \dfrac{|p - c + mean|}{max - mean}, & \text{if } c - max \leq p < c - mean \\[2mm] 1 - \dfrac{|p - c + mean|}{mean - min}, & \text{if } c - mean \leq p \leq c - min \\[2mm] 1 - \dfrac{|p - c - mean|}{mean - min}, & \text{if } c + min \leq p \leq c + mean \\[2mm] 1 - \dfrac{|p - c - mean|}{max - mean}, & \text{if } c + mean < p \leq c + max \\[2mm] 0, & otherwise \end{cases}
\tag{5}
$$

In the case that $p = x$, the point would be $c = c_x$. Hence, the calculated fuzzy set corresponds to $\tilde{B}_x$. If $p = y$ then $c = c_y$. Using this value, the defined fuzzy set is $\tilde{B}_y$. Finally, if $p = z$ implies that $c = c_z$. The resulting fuzzy set is $\tilde{B}_z$. The 3D fuzzy set $\tilde{B}_{xyz}$, which defines the whole sphere, is the result of applying the product t-norm of the latter three 1D fuzzy sets $\tilde{B}_{xyz} = \tilde{B}_x \cdot \tilde{B}_y \cdot \tilde{B}_z$. Its membership function is determined as $\mu_{\tilde{B}_{xyz}} = \mu_{\tilde{B}_x} \cdot \mu_{\tilde{B}_y} \cdot \mu_{\tilde{B}_z}$.

B. Modeling the Landmark Matching Uncertainty Using Cones

The previous approach is a good solution to tackle this SFO problem of modeling the facial soft tissue thickness with the available studies. However, it produces inconsistent anatomical solutions because it does not model the positional relationships between pairs of landmarks. For example, facial landmarks located inside the skull are considered to be correctly positioned (see the upper right part of the sphere in Fig. 4 and the whole left part of the sphere in Fig. 5a).

A variation of the previous method is proposed as follows: instead of modeling the soft tissue depth using a sphere with the cranial landmark in the center, we define a cone whose vertex is the cranial landmark. Each facial landmark is supposed to be located inside the cone (Fig. 4b). Using a cone, we closely specify a narrower region where the facial landmark should be located. We refine the search of the facial landmarks placement to a specific region. We assume a certain degree of
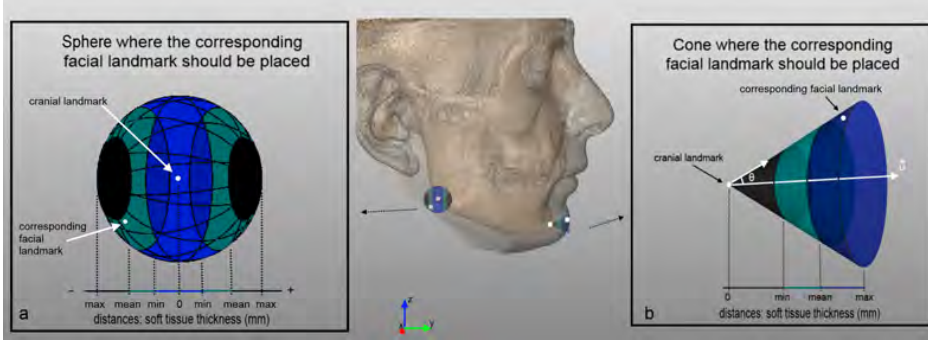
Fig. 4: Facial landmark position from a cranial landmark using a sphere (a) and a cone (b). *min*, *mean*, and *max* are the soft tissue depths. For the cone, $\vec{u}$ is the normal vector at the cranial landmark in the skull, θ is the rotation angle of $\vec{u}$.
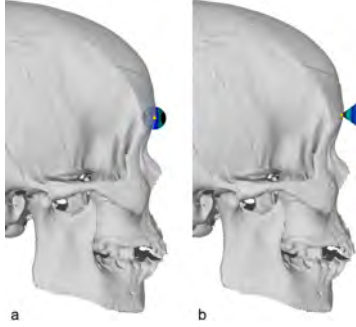

Fig. 5: Facial landmark position from a cranial landmark using a sphere (a) and a cone (b).

perpendicularity between cranial and facial landmarks as most soft tissue studies do [23], [24].

To do this, we consider the normal vector [2] on the surface of the skull 3D model at each cranial landmark: $\vec{v} = (x_n, y_n, z_n)$. The unit vector of $\vec{v}$, which has the same direction, but a magnitude of the unit, has been determined below:

$$\vec{u} = (\frac{x_n}{\|\vec{v}\|}, \frac{y_n}{\|\vec{v}\|}, \frac{z_n}{\|\vec{v}\|}) = (u_x, u_y, u_z), \tag{6}$$

where $\|\vec{v}\| = \sqrt{(x_n^2 + y_n^2 + z_n^2)}$ is the magnitude of $\vec{v}$.

In order to estimate the position of the facial landmarks, $\vec{u}$ coordinates (u_x, u_y, u_z) are multiplied by the specific distance (*min*, *mean*, or *max*).

Since the correspondence between a pair of cranial-facial landmarks is not always perpendicular, different inclination angles are applied to the unit vector $\vec{u}$ in order to define the volume in which the facial landmark is likely to be located. The amplitude of this area can be defined by a rotation equal to $\pm\theta$ along the three axes X, Y and Z.

The 3D rotation of the unit vector $\vec{u}$ consists of three different rotations, i.e., a rotation of an angle $\pm\theta$ along the X, Y and Z axes [37]. The expressions for each axis rotation

are the following:

$$X - axis\ rotation = \begin{cases} x' = u_x \\ y' = u_y \cdot cos\theta - u_z \cdot sin\theta \\ z' = -u_y \cdot sin\theta + u_z \cdot cos\theta \end{cases} \tag{7}$$

$$Y - axis\ rotation = \begin{cases} x' = u_x \cdot cos\theta + u_z \cdot sin\theta \\ y' = u_y \\ z' = -u_x \cdot sin\theta + u_z \cdot cos\theta \end{cases} \tag{8}$$

$$Z - axis\ rotation = \begin{cases} x' = u_x \cdot cos\theta - u_y \cdot sin\theta \\ y' = u_x \cdot sin\theta + u_y \cdot cos\theta \\ z' = u_z \end{cases} \tag{9}$$
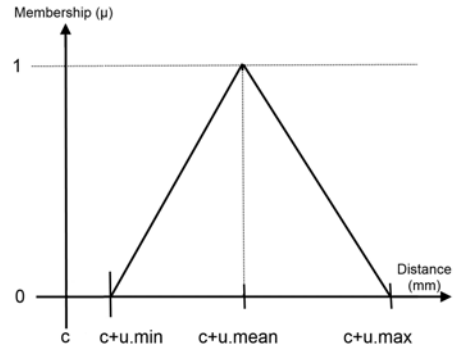
Fig. 6: Graphical representation of one of the dimensions of the 3D fuzzy set $\tilde{B}_p$ for modeling the landmark matching uncertainty with a cone. The other two dimensions are modeled in a homologous way.

Likewise, we define a fuzzy set whose center is the 3D cranial landmark with a membership degree of zero. The rest of the points are calculated by multiplying the coordinates of the unit vector $\vec{u}$ by the values of the different distances *min*, *mean*, and *max*. Thus, a 3D cone is defined where each facial landmark is expected to be located according to a particular soft tissue depth study (Figs. 2 and 4b). The landmark matching uncertainty is defined using 3D masks in the same fashion as the spheres (see Sec. III.A). Hence, a fuzzy set $\tilde{B}_p$ with $p \in \{x, y, z\}$ (see Fig. 6 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:

$$\tilde{B}_p = \begin{cases} 1 - \dfrac{|p - c - u \cdot mean|}{u(mean - min)}, if\ c + u \cdot min \leq p \leq c + u \cdot mean \\ 1 - \dfrac{|p - c - u \cdot mean|}{u(max - mean)}, if\ c + u \cdot mean < p \leq c + u \cdot max \\ 0,\ otherwise \end{cases}$$

$$(10)$$

In the case that $p = x$, the points would be $c = c_x$ and $u = u_x$. Hence, the calculated fuzzy set corresponds to $\tilde{B}_x$. If $p = y$ then $c = c_y$ and $u = u_y$. Using these values, the defined fuzzy set is $\tilde{B}_y$. If $p = z$ it implies that $c = c_z$ and $u = u_z$, with the resulting fuzzy set being $\tilde{B}_z$.

The 3D fuzzy set $\tilde{B}_{xyz}$, which defines the whole 3D cone, is the product t-norm of the previous three, one-dimensional fuzzy sets $\tilde{B}_{xyz} = \tilde{B}_x \cdot \tilde{B}_y \cdot \tilde{B}_z$. Its membership function is determined by the product t-norm of the membership functions of each fuzzy set: $\mu_{\tilde{B}_{xyz}} = \mu_{\tilde{B}_x} \cdot \mu_{\tilde{B}_y} \cdot \mu_{\tilde{B}_z}$.

C. Distance Between Two Fuzzy Sets and Fitness Function Definition

Our proposal allows experts to mark both precise and imprecise (fuzzy) facial landmarks. These latter landmarks can be used by forensic anthropologists as necessary.

With the modeling of the landmark matching uncertainty proposed in the previous subsection, the need to compute the distance between two fuzzy sets arises. One of these two fuzzy sets is the projection of the 3D cranial landmark on the facial photograph, which is composed of the precise 3D cranial landmark (called cl^i in the two previous sections) and the fuzzy set that models the landmark matching uncertainty. The other fuzzy set would be the fuzzy facial landmark of the photograph representing the imprecise position of a facial landmark (Fig. 1).

The distance between two fuzzy sets $\tilde{F}$ and $\tilde{G}$ can be stated by [39]:

$$d''(\tilde{F}, \tilde{G}) = \frac{\sum\limits_{k=1}^{m} \sum\limits_{l=1}^{n} (d(\tilde{F}_k, \tilde{G}_l) \cdot min[\alpha_k, \beta_l])}{\sum\limits_{k=1}^{m} \sum\limits_{l=1}^{n} min[\alpha_k, \beta_l]}, \quad (11)$$

where $\tilde{F}_k$ is the k_{th} element of the fuzzy set $\tilde{F}$, $\tilde{G}_l$ is the l_{th} element of the fuzzy set $\tilde{G}$, d is the 2D Euclidean distance, α_k is the membership value of $\tilde{F}_k$, β_l is the membership value of $\tilde{G}_l$, m is the number of elements of the fuzzy set $\tilde{F}$, and n is the number of elements of the fuzzy set $\tilde{G}$.

Therefore, the fitness function *Fuzzy Mean Error* (FME) for the current automatic 3D-2D SFO task has been formulated by taking into account the latter fuzzy distances:

$$FME = \frac{\sum\limits_{i=1}^{Ncrisp} (d'(x_i, f(\tilde{C}^i))) + \sum\limits_{j=1}^{Nfuzzy} (d''(\tilde{F}^j, f(\tilde{C}^j)))}{N}, \quad (12)$$

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 bi-dimensional fuzzy sets, *N* is the total number of landmarks considered (*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 (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.

Solving the SFO problem results in an extremely complex optimization task with a highly *multimodal* landscape. This scenario led us to face the problem of considering robust EAs to search for the optimal values of the 12 registration transformation parameters, as discussed in Sec. II.C.

IV. EXPERIMENTS

Some detailed experiments have been developed to analyze the performance of our approaches, including the treatment of the landmark matching uncertainty in comparison with the state-of-the-art SFO automatic methods proposed in [9], [16], which do not consider this capability.

A. Materials

The experimental design involves 18 SFO problem instances corresponding to nine cases of live people (from Spain and Italy) that will allow the study of ground truth data and the subsequent objective evaluation. These instances were created as follows [26]: 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.

During the same clinical session, two digital photographs were taken in frontal and lateral poses. The patients were in an orthostatic position and asked to maintain a neutral expression. Unfortunately, CBCTs only scan from the lowest part of the mandible to the eyebrows without including the top of the head (see Fig. 4). The photographs were taken at a distance of between 1 m and 1.5 m, using a digital camera with a minimum resolution of 4 Mpx.

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 projection of the skull onto the facial photograph that can later be compared with the outcome of every SFO method. As a result, we obtained the ground truth data which is 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^{TM} software [40], which has been developed by our team. This software allows forensic 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^{TM} also integrates and runs the proposed automatic SFO algorithms. All the experiments have been performed on an Intel Corel^{TM} 2 Quad CPU Q8400 2.66 GHz, with 4GB RAM, running Windows 7 Professional^{TM}.

B. Experimental Design

First, we compared the performance of the state-of-the-art automatic approaches RCGA and CCGA-2 [9], [16], neither of which model the landmark matching uncertainty. These proposals calculate the distance between a skull and its corresponding facial landmark by means of the fitness function defined in Eq. (4). We developed this experiment because the previous proposals had not tested using a ground truth dataset. Furthermore, they were compared only with imprecise metrics in six SFO problem instances corresponding to three cases [16]. The parameter configuration used for testing these approaches is that which obtained the best results in [15] and [16] respectively. The EA with the best performance was chosen to incorporate the facial soft tissue.

In the next step, we included triangular membership functions in the best EA of the previous analysis, to model the facial soft tissue following the two new proposals based on spheres and cones.

Furthermore, the influence of the fuzzy membership function type was analyzed using the method that showed the best performance in the experiment. The two other most common fuzzy membership functions (piece-wise trapezoidal and Gaussian) were tested.

Finally, to better understand the behavior of our proposal, we have included a study based on negative cases. We tested the best algorithm, i.e., RCGA-c-45 using a triangular fuzzy set function when a 3D skull and a photograph belong to different persons.

Since all these methods are based on stochastic processes, 30 independent runs were performed for each problem instance to compare the robustness of the methods and to avoid any possible bias.

C. Performance Analysis of the Existing CCGA-2 and RCGA Approaches

Table I presents the mean error achieved for each case and pose (f = frontal view and l = lateral view) by RCGA and CCGA-2 in the 30 runs performed as well as the total average error per algorithm.

TABLE I: Mean error in mm regarding the ground truth obtained in 30 runs for each case, comparing CCGA-2 with RCGA. 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

Case, pose	pl	il	CCGA-2	RCGA
1,f	7	7	4.565	**2.750**
1,l	5	4	16.588	**7.406**
2,f	8	5	4.906	**3.690**
2,l	3	2	**8.299**	8.605
3,f	8	7	3.815	**3.629**
3,l	4	4	**9.367**	10.643
4,f	7	6	4.635	**3.647**
4,l	4	3	**14.027**	14.327
5,f	10	6	2.996	**2.436**
5,l	5	4	8.545	**6.865**
6,f	8	7	4.654	**3.784**
6,l	3	4	16.486	**12.959**
7,f	10	5	4.253	**3.639**
7,l	5	4	**10.212**	12.663
8,f	9	6	5.269	**4.409**
8,l	4	4	**6.840**	8.555
9,f	10	4	6.174	**5.541**
9,l	3	5	**10.210**	11.701
Average (mm)			7.778	**6.976**

The distance error has been calculated for all the landmarks marked in each case study. 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 of a particular algorithm.

As a result of the comparison between RCGA and CCGA-2, RCGA achieved the best behavior in 13 of the 18 cases with significant differences in most of them. Thus, we selected RCGA to incorporate the facial soft tissue modeling.

D. Performance Analysis of the New Proposals Incorporating the Facial Soft Tissue Modeling

We call the new proposals RCGA-s and RCGA-c. RCGA-s corresponds to the approach that deals with the landmark matching uncertainty using spheres. In the case of RCGA-c, the landmark matching uncertainty is modeled with cones. Within RCGA-c, four different angles were tested to define the cone amplitude: $\theta = \pm 0°$, $\theta = \pm 10°$, $\theta = \pm 30°$, and $\theta = \pm 45°$, so we analyzed four variants of this proposal: RCCA-c-0, RCGA-c-10, RCGA-c-30 and RCGA-c-45. They both use the fitness function defined in Eq. (12). The experiment again consisted of 30 independent runs for each problem instance.

The best performance regarding the ground truth is obtained by RCGA-c-45 in most of the cases (see Table II). In frontal views, RCGA-c-0 obtains the most accurate results for cases 5 and 6, RCGA-s for case 1, and RCGA-c-45 for the rest of the cases. The mean error of the best cases ranges from 2.204 to 4.887 mm. In lateral views, the most precise SFOs correspond

TABLE II: Mean error in mm regarding the ground truth obtained in 30 runs for each case using different parametrization for RCGA, f = frontal and l = lateral poses of the face in the photograph

Case, pose	RCGA s	RCGA c-0	RCGA c-10	RCGA c-30	RCGA c-45
1,f	**2.727**	4.116	4.330	3.078	3.164
1,l	6.938	7.400	6.693	**5.788**	5.828
2,f	3.486	3.823	4.212	3.337	**3.331**
2,l	8.871	8.871	8.871	5.286	**3.688**
3,f	3.664	6.594	6.594	3.007	**2.941**
3,l	10.815	11.073	11.073	7.604	**7.001**
4,f	3.505	3.356	3.356	3.223	**3.045**
4,l	13.494	13.847	13.847	12.446	**12.171**
5,f	2.453	**2.204**	2.406	2.641	2.600
5,l	6.097	6.846	6.846	3.715	**2.825**
6,f	3.644	**2.618**	3.076	2.878	2.921
6,l	11.665	**2.204**	11.126	11.009	10.626
7,f	**3.617**	4.695	4.695	4.100	3.714
7,l	10.909	**9.649**	**9.649**	9.729	10.345
8,f	3.579	3.492	3.492	2.856	**2.882**
8,l	7.812	19.935	19.935	6.262	**5.412**
9,f	5.757	6.660	6.660	5.071	**4.887**
9,l	11.700	19.852	19.852	10.391	**9.509**
Average (mm)	6.548	7.759	8.416	5.585	**5.281**

TABLE III: P values for the comparison between the control and the rest of the approaches

Approach	Unadjusted p	p Bonf	p Holm
(RCGA-c-45 is the control)			
RCGA-s	0.0502	1.0000	0.5531
RCGA-c-0	**0.0003**	**0.0077**	**0.0063**
RCGA-c-10	**<0.0001**	**0.0003**	**0.0003**
RCGA-c-30	0.6216	1.0000	1.0000

ground truth in the frontal image, with a total average error of 2.204 mm (Table II). The two gonions present a larger distance with respect to their counterpart ground truth points (Fig. 7a, b). In the case of the lateral pose, the best overlay has been achieved by RCGA-c-45 with an error equal to 2.825 mm. Alare right and subnasale are the closest points to their corresponding ground truth landmarks (Fig. 7c, d).

Figure 7e and g presents the SFOs in frontal and lateral views for case 5, achieved by the best algorithm without incorporating the modeling of the landmark matching uncertainty (RCGA). The resulting matching between pairs of cranial and ground truth landmarks are depicted in Fig. 7f, h. RCGA obtains the larger error in the two ectocanthions, the two endocanthions, nasion, gonion right, alare and zygion left for the frontal image (Fig. 7e, f). The mean total error of RCGA is 2.436 mm for the frontal photograph in case 5 (Table I). In the lateral image, RCGA-c-10 presents a higher error in all landmarks compared with RCGA-c-45 (Fig. 7d, h). We can also see that the skull protrudes from the face in the nasal region (Fig. 7g). A manual refinement would be needed for this SFO result. The mean total error of RCGA in the lateral view is 6.865 mm (Table I).

In general, we obtained competitive matchings between pairs of corresponding cranial and facial landmarks in all the studied cases. As previously noted, the best results are obtained by the new proposals, which include the treatment of the landmark matching uncertainty.

We performed a Friedman test [41] to analyze whether significant differences exist among the performance of all the approaches. The aim was to test a null hypothesis stating that the mean total errors of all the approaches were the same. We set the experiment level of significance at $\alpha = 0.05$. The statistic results of this test are a Friedman χ^2 equal to 43.9695 and a p value of 7.496e-08. This data reveals significant differences among the behavior of the approaches 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 [42] was carried out to detect significant differences among a control approach and the rest. RCGA-c-45 was the control algorithm because it outperformed the remaining methods, i.e., it obtained the lowest value in the Friedman ranking (Fig. 8). In the Bonferroni-Dunn test, we obtain 3.093 and 2.856 as critical values using levels of significance $\alpha = 0.05$ and $\alpha = 0.10$ respectively.

Figure 8 summarizes the ranking obtained by the Friedman test. The bar height indicates the average ranking of each proposal. We drawn a line through the chart whose value

to RCGA-c-45 and RCGA-c-0 for cases 4 and 5, while RCGA-c-45 obtains the best results in all instances except cases 1 (RCGA-c-30) and 7 (RCGA-c-0 and RCGA-c-10). The best mean errors are higher than in the frontal view, it achieves values between 2.204 and 9.649 mm.

RCGA-c-45 and RCGA-c-30 achieve a total average error of 5.281 and 5.585 mm respectively. RCGA-s gives more than 6 mm of total average error: 6.548 mm, and the worse algorithm is RCGA-c-10 with 8.416 mm (Table II).

Figure 7a, c shows the best superimpositions obtained for the two photographs of the fifth case as a particular illustration of the methods' performance. These outcomes have been achieved by RCGA-c-0 and RCGA-c-45 for the frontal and the lateral image respectively. In particular, blue points (dark gray in the black and white version of this manuscript) refer to the cranial landmarks after overlaying the skull 3D model onto the photograph. Yellow points (light gray in the black and white version) are the actual landmarks achieved by the ground truth geometric transformation *g*. Green points (gray in the black and white version) are the facial landmarks marked by the expert in the photograph. We should remind readers that these facial points have been placed as either precise or imprecise (ellipses) landmarks. Each ellipse contains a gray point corresponding to its center. For better visualization, the resulting matching between pairs of cranial and ground truth landmarks for case 5 (frontal and lateral view) are depicted in Fig. 7b, d.

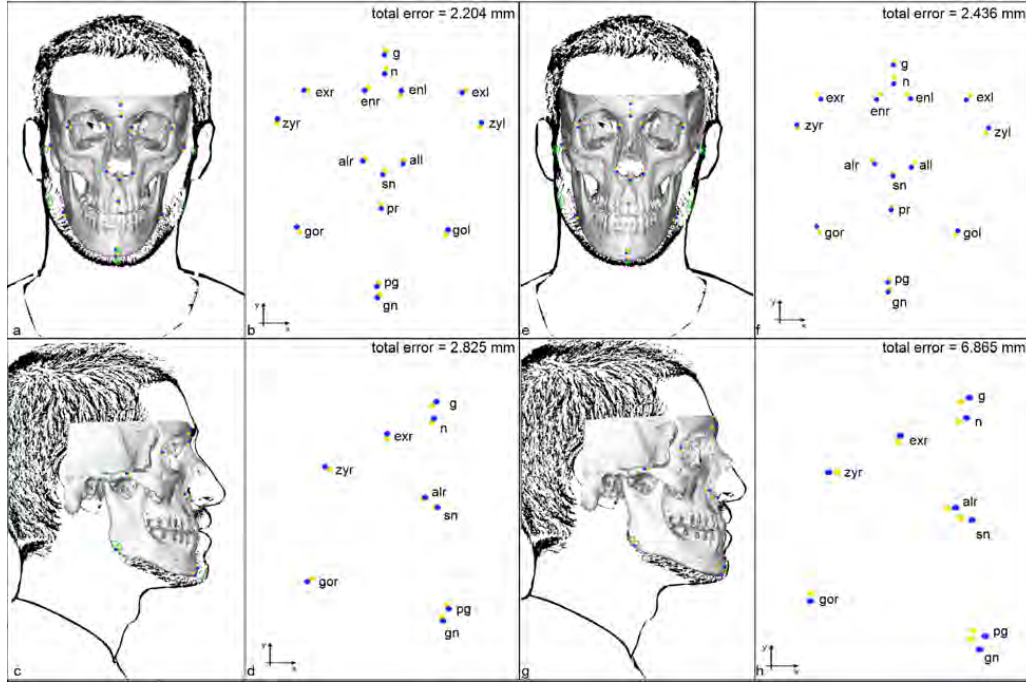RCGA-c-0 obtains the most accurate solution regarding the

Fig. 7: Case 5. Best (a, c) and RCGA (e, g) automatic SFOs for frontal and lateral poses. Best (b, d) and RCGA (f, h) matching between pairs of projected cranial and actual ground truth landmarks for frontal and lateral views. Frontal images: glabella (g), nasion (n), exocanthion left, right (exl, exr), endocanthion left, right (enl, enr), alare left, right (all, alr), subnasale (sn) and prosthion (pr) were marked as precise facial landmarks. Zygion left, right (zyl, zyr), gonion left, right (gol, gor), pogonion (pg) and gnathion (gn) were placed as imprecise facial landmarks. Lateral images: Glabella (g), nasion (n), exocanthion right (exr), alare right (alr) and subnasale (sn) were placed in the photograph as a precise points. Zygion right (zyr), pogonion (pg) and gnathion (gn) were marked as imprecise facial landmarks. The homologous cranial landmarks were placed on the skull as precise points.
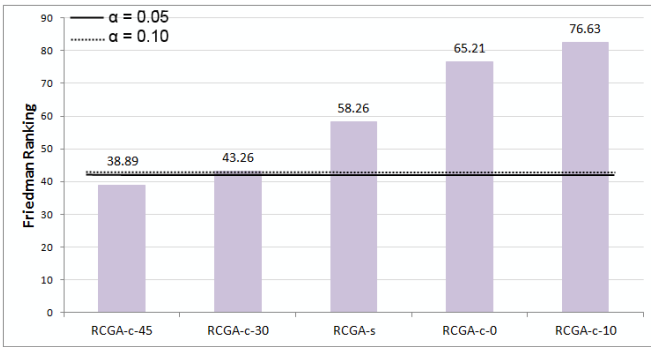

Fig. 8: Friedman ranking and Bonferroni-Dunn lines for the classification rate.

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 [43].

We 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 [44].

Table III details the pairwise comparisons considering RCGA-c-45 as the control approach. 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 at $\alpha = 0.05$.

RCGA-c-45 statistically outperforms the rest of the approaches with a confidence level of 95 % except for RCGA-s and RCGA-c-30 (Table III). These results corroborate the data obtained by the Bonferroni-Dunn test applied to the Friedman ranking where RCGA-c-45 and RCGA-c-30 present a better behavior than the other approaches (Fig. 8).

E. Analysis of the Influence of the Fuzzy Membership Function Type on the Landmark Matching Uncertainty Modeling

In order to finish studying the optimal design for our innovative approach, we analyzed the performance of the best automatic SFO method (RCGA-c-45) when considering two other types of fuzzy membership functions: piece-wise trapezoidal and Gaussian. Table IV presents the mean error obtained in each case and pose in the 30 runs performed for the new analyzed approaches, as well as the total average error per algorithm regarding the ground truth dataset.

The best performance is again obtained by RCGA-c-45 using a triangular fuzzy set function (total average = 5.281 mm). RCGA-c-45 with a Gaussian fuzzy set function is the best algorithm in cases 2 and 3 for frontal view. In lateral view cases 1 and 4, RCGA-c-45 with a trapezoidal function gives the best results.

TABLE IV: Mean error in mm regarding the ground truth obtained in 30 runs for each case, using different fuzzy membership functions for the best algorithm RCGA-c-45, f = frontal and l = lateral facial poses in the photograph

Case, pose	RCGA-c-45	RCGA-c-45 Gaussian	RCGA-c-45 trapezoidal
1,f	3.164	**2.960**	3.052
1,l	5.828	6.116	**5.465**
2,f	3.331	**3.071**	3.165
2,l	**3.688**	4.436	4.618
3,f	2.941	**2.774**	2.816
3,l	**7.001**	7.487	7.325
4,f	**3.045**	3.102	3.105
4,l	12.171	11.875	**10.747**
5,f	2.600	2.527	**2.524**
5,l	**2.825**	3.451	3.158
6,f	**2.921**	3.058	2.979
6,l	**10.626**	10.900	11.692
7,f	3.714	**3.653**	3.679
7,l	10.345	10.821	**10.279**
8,f	**2.882**	2.908	2.896
8,l	**5.142**	6.693	6.680
9,f	**4.887**	5.164	5.200
9,l	**9.509**	9.575	9.626
Average (mm)	**5.281**	5.490	5.406

TABLE V: Fuzzy mean error (FME) in pixels obtained in 30 runs for all the female cases using the best algorithm RCGA-c-45, f = frontal and l = lateral facial poses in the photograph. Positive cases are highlighted in bold, the rest are negative ones

Image,pose \ Skull	1	2	3	8
1,f	**0.020**	0.060	0.040	0.066
1,l	**0.022**	0.057	0.049	0.059
2,f	0.027	**0.061**	0.039	0.061
2,l	0.028	**0.047**	0.040	0.060
3,f	0.015	0.055	**0.034**	0.045
3,l	0.016	0.021	**0.034**	0.034
8,f	0.025	0.056	0.039	**0.047**
8,l	0.030	0.039	0.040	**0.045**

TABLE VI: Fuzzy mean error (FME) in pixels obtained in 30 runs for all the male cases using the best algorithm RCGA-c-45, f = frontal and l = lateral facial poses in the photograph. Positive cases are highlighted in bold, the rest are negative ones

Image,pose \ Skull	4	5	6	7	9
4,f	**0.035**	0.074	0.049	0.039	0.072
4,l	**0.035**	0.076	0.028	0.039	0.076
5,f	0.080	**0.043**	0.049	0.049	0.050
5,l	0.048	**0.034**	0.020	0.049	0.054
6,f	0.046	0.073	**0.027**	0.045	0.069
6,l	0.019	0.049	**0.023**	0.059	0.082
7,f	0.051	0.079	0.050	**0.034**	0.073
7,l	0.020	0.050	0.028	**0.051**	0.084
9,f	0.048	0.076	0.050	0.049	**0.047**
9,l	0.003	0.050	0.029	0.058	**0.058**

We also performed a Friedman test [41] following the previous experiment design. The aim was to test a null hypothesis stating that the mean total errors of all the methods are the same. We set the experiment level of significance at $\alpha = 0.05$. The statistic results are a Friedman χ^2 equal to 63.336 and a p value of 73.244e-11, thus rejecting the null hypothesis. A Bonferroni-Dunn test was performed due to the rejection of the null hypothesis using RCGA-c-45 triangular fuzzy function as the control algorithm. We obtained 3.330 and 3.106 as critical values using a level of confidence $\alpha = 0.05$ and $\alpha = 0.10$ respectively.

We also applied a paired t test with a Bonferroni and a Holm correction, as well as an unadjusted p value to discover the differences between approaches, considering RCGA-c-45 with a triangular fuzzy set function as the control approach. This analysis reveals that no significant difference is obtained between RCGA-c-45 (triangular membership function) and RCGA-c-45-Gaussian. Likewise, no significant difference is achieved between the RCGA-c-45 and RCGA-c-45-trapezoidal methods.

Therefore, we cannot assert that modeling the soft tissue depth between corresponding pairs of cranial and facial landmarks using trapezoidal and Gaussian fuzzy functions is significantly better than using triangular fuzzy functions for RCGA-c-45. Due to the fact that triangular fuzzy functions are quicker to handle, easier to implement and obtain the best total error regarding the ground truth (Table IV), we have chosen this type of function to model the uncertainty related to the

soft tissue depth between pairs of cranial and facial landmarks.

F. Negative Cases Study

In this final experiment the goal was two-fold. On the one hand, to show the behavior of our proposal using unrelated skull 3D models and facial photographs. On the other hand, to make clear that the aim of our method is to obtain the best possible SFO, without assuming any particular skull-face relationship *a priori*, rather than addressing the third CFS stage, i.e., the decision making.

Thus, our best algorithm was run over the same data set employed before (distinguishing between female and male cases) but this time superimposing each skull 3D model over the frontal and lateral images of the remaining subjects (i.e., each SFO involves a 3D skull and a photograph not belonging to the same person). We considered RCGA-c-45 using a triangular fuzzy set function and the same parameter set as in the previous experiments.
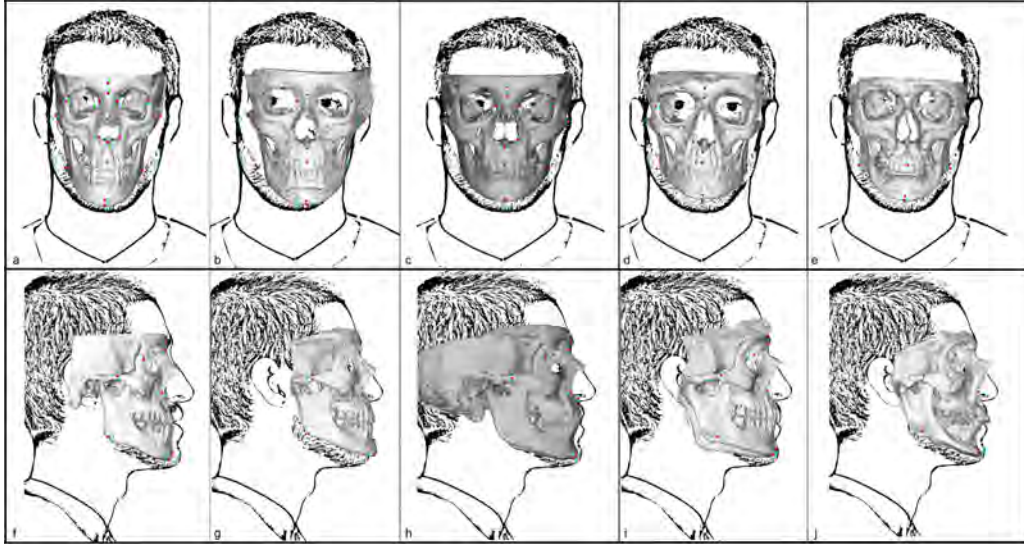
Fig. 9: Skull-face overlays achieved by the best algorithm RCGA-c-45 using the photographs (frontal and lateral poses) of case 5. (a) and (f) frontal and lateral SFOs (respectively) using the actual skull (positive case). From (b) to (e): SFOs using the frontal image of case 5 and the skulls of cases 4, 6, 7 and 9 respectively (negative cases). From (g) to (j) SFOs using the lateral image of case 5 and the skulls of cases 4, 6, 7 and 9 respectively

Tables V and VI present the fuzzy mean errors for the female and male negative cases respectively. Notice that the fuzzy mean error we used to evaluate each SFO within the EA-IR process is a measure of the quality of the overlay but cannot be used to directly determine if the skull and the face actually belong to the same person. In fact, as depicted in Tables V and VI, this error is smaller for some negative cases in comparison with their positive counterparts (for example, the skull in the first case with the frontal and lateral images of case 3, the skull in case 2 with the lateral images of cases 3 and 8, etc.). This simple experimentation allows us to illustrate that the purpose of our approach is to try to achieve the best possible skull-face overlay without assuming any particular skull-face correspondence. Decision making, carried out by forensic anthropologists, is a different and also challenging task where they have to consider different criteria such as analyzing the consistency of the bony and facial outlines/morphological curves, assessing anatomical consistency by positional relationship, marking and comparing lines to analyze anatomical consistency and evaluating the consistency of the soft tissue thickness among corresponding cranial and facial landmarks [3].

In order to illustrate the results of our proposal when dealing with negative cases, Fig. 9 shows the best superimpositions obtained by the RCGA-c-45 algorithm for the negative cases corresponding to the case 5 photographs (frontal and lateral poses) and the 3D skull of the male cases 4, 6, 7 and 9. The obtained overlays for the positive cases considering the actual subject of the fifth skull have also been included for comparison. Red points (dark gray in the black and white version of this manuscript) refer to the cranial landmarks after overlaying the skull 3D model on the photograph. Green points (light gray in the black and white version) are the facial landmarks marked by the forensic anthropologist in the photograph.

In particular, the fuzzy mean error of the lateral positive case number 5 is 0.034 while the error of the negative case using skull 6 and the image of case 5 is 0.020 (see Table VI). However, Fig. 9h shows that the latter negative superimposition is not correct because the nasal bone protrudes from the face, which can be quickly detected by an expert.

V. DISCUSSION AND CONCLUSIONS

The SFO process is a repetitive and tedious task in CFS. It requires several hours to overlay a skull on a facial photograph. The design of unbiased, systematic, automatic and quantifiable methods to perform SFO is a real need in forensic anthropology [8].

Within CFS, we have just focused on the SFO stage. The main goal of our methodology is to assist forensic anthropologists to obtain the best possible overlay, i.e., orienting and positioning the skull on the facial photograph, reducing the SFO processing time and simplifying their work.

Our method is aimed at automating the SFO task, the most time consuming one involving the projection of a skull model over a face picture. During the last century, this procedure has been tackled by manual or semi-automatic methods based on the use of either skull photographs (skull-photo superimposition) or a video mixing device (video superimposition). Recently, forensic anthropologists have determined that the use of a 3D skull model is a more informative representation, but the manual projection has remained. Our proposal involves the design of an automatic method to obtain accurate 3D skull-2D face overlays in a short time by considering the imprecision related to the matching of landmarks in the skull and face. However, the human identification decision, considered as the final output, is taken in the third stage, decision making. This decision is manually taken by a forensic anthropologist,

applying her/his knowledge and expertise to analyze the skull-face overlay obtained as the output of the second CFS stage. In addition, notice that the fuzzy mean error we defined to deal with SFO in the second stage is a measure of the quality of the skull-face overlay but cannot be used to directly determine if the skull and the face actually belong to the same person. Forensic anthropologists consider different criteria to take their final decision.

The correspondence between facial and cranial landmarks is not always symmetrical and perpendicular. We have defined two proposals to deal with this source of uncertainty by using fuzzy sets and taking into account the available information concerning soft tissue depths. In particular, we have considered the measurements of the facial soft tissue distances detailed in two studies for Mediterranean populations [23]–[25].

We have chosen a triangular fuzzy set by defining the masks for handling the landmark matching uncertainty related to the facial soft tissue distances. This choice is motivated by the large amount of literature available on the topic. Although different membership function shapes can be considered, both piece-wise (e.g. trapezoidal-shaped) and continuous (e.g. Gaussian), piece-wise triangular fuzzy membership functions are simpler and easier to handle, providing a quicker response. In addition, some works have proven that a linear piece-wise fuzzy membership function can approximate a continuous function to the desired degree, achieving similar results [38].

The performance of our automatic approaches and the two previous ones [9], [16], which do not include the landmark matching uncertainty treatment, have been objectively evaluated considering a ground truth dataset in 18 case studies. Our proposals achieve a competitive matching between pairs of corresponding cranial and facial landmarks because of the natural modeling of the landmark correspondences using SC techniques. In particular, RCGA-c-45 significantly outperforms the remaining automatic SFO methods tested. It obtains the best performance regarding the ground truth in the majority of the cases. The RCGA-c-30 variant also achieves a good ranking in the computed statistical analysis (in fact, it is the same method with a different parametrization). Hence, the new proposals that model the landmark matching uncertainty get better results than the previous approaches that do not consider the facial soft tissue thickness.

In order to extend our experiment, we analyzed the performance of the best automatic SFO method (RCGA-c-45) when considering another two types of fuzzy membership function: piece-wise trapezoidal and Gaussian, which model the soft tissue depth between pairs of corresponding cranial and facial landmarks. The performance of the SFO method considering these two kinds of membership functions was compared with the rest of the methods previously implemented (either based on the use of triangular fuzzy sets to model the soft tissue depth or not taking into account the soft tissue depths). The best performance is also obtained by RCGA-c-45 using a triangular fuzzy set function. RCGA-c-45 with Gaussian and trapezoidal fuzzy set functions achieve good results. However, statistical analysis shows that modeling the soft tissue depth between corresponding pairs of cranial and facial landmarks using RCGA-c-45 with trapezoidal and Gaussian fuzzy func-

tions is not significantly better than using RCGA-c-45 with triangular fuzzy functions. Furthermore, to better understand the behavior of our proposal, we have included a study based on negative cases. We have tested the best algorithm, i.e., RCGA-c-45 using a triangular fuzzy set function when a 3D skull and a photograph belong to different people. The fuzzy mean errors of negative cases are lower than the fuzzy mean error in the positive ones. This shows the proper behavior of our proposal, which only searches for the best possible SFO.

Although valuable results are obtained, it is important to note that a distance or residual error is still obtained when evaluating an SFO result against its corresponding ground truth overlay. A few sources of uncertainty (imprecision) are behind these errors. We list them together with possible solutions that could serve as future lines of research to improve SFO accuracy:

Our method makes use of soft tissue studies which provide population-based mean and standard deviation distances for corresponding landmarks. We neither consider the real cranial and facial landmark distances of each individual nor their exact spatial orientation (as would be the case in a real-world scenario when faced the identification of an unknown skull). This error could be mitigated having more information about the skull-face relationships. In particular, it could be extremely useful to measure the spatial orientations of the facial landmarks with respect to their corresponding cranial points in order to model the landmark matching uncertainty accordingly.

Our approach relies on the location of cranial and facial landmarks. As has already been explained in this article, placing landmarks on a photograph is a difficult task that presents a particular degree of imprecision [18], [19]. Thus, the imprecision in the location of landmarks has a negative effect on the final SFO achieved (best transformation). This problem could be minimized either including a final refinement stage relying only on these landmarks that are known to be more precisely located [18], [19] and/or guiding the overlay algorithm not only with landmarks but also with corresponding anatomical regions, i.e., teeth overlapping.

Finally, in our proposal we are trying to overlay a rigid 3D model (the whole skull) onto a photograph of the person where the underlying bony part is not rigid, i.e., the mandible in the photograph could have a different articulation than in the skull 3D model. To address this source of error, the mandible in the 3D model could be manually or automatically articulated accordingly to its position in the photograph. Note that this source of error could be considered insignificant in our study since all the patients were asked to have the same facial gesture while scanning and photographing [26].

REFERENCES

[1] K. R. Burns, *The forensic anthropology training manual*. Upper Saddle River, New Jersey: Pearson Education, 3rd ed., 2012.

[2] C. Cattaneo, "Forensic anthropology: development of a classical discipline in the new millennium," *Forensic Sci Int*, vol. 165, pp. 185–193, 2007.

[3] M. Yoshino, "Craniofacial superimposition," in *Craniofacial Identification* (C. Wilkinson and C. Rynn, eds.), pp. 238–253, Cambridge: University Press, 2012.

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

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

[6] S. Damas, O. Cordón, O. Ibáñez, J. Santamaría, I. Alemán, and M. Botella, "Forensic identification by computer-aided craniofacial superimposition: A survey," *ACM Computing Surveys (CSUR)*, vol. 43, p. 27, 2011.

[7] T. W. Fenton, A. N. Heard, and N. J. Sauer, "Skull-photo superimposition and border deaths: identification through exclusion and the failure to exclude," *J Forensic Sci*, vol. 53, pp. 34–40, 2008.

[8] D. H. Ubelaker, "A history of smithsonian-FBI collaboration in forensic anthropology, especially in regard to facial imagery [abstract]," *Forensic Sci Communications*, vol. 2, 2000.

[9] O. Ibáñez, O. Cordón, S. Damas, and J. Santamaría, "Modeling the skull–face overlay uncertainty using fuzzy sets," *IEEE Trans Fuzzy Syst*, vol. 16, pp. 946–959, 2011.

[10] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*. Toronto, Canada: Thomson-Engineering, 3rd ed., 2007.

[11] B. Zitovà and J. Flusser, "Image registration methods: a survey," *Image Vision Comput*, vol. 21, pp. 977–1000, 2003.

[12] P. P. Bonissone, "Soft computing: the convergence of emerging reasoning technologies," *Soft Comput*, vol. 1, pp. 6–18, 1997.

[13] L. A. Zadeh, "Soft computing and fuzzy logic," *IEEE Softw*, vol. 11, pp. 48–56, 1994.

[14] A. E. Eiben and J. E. Smith, *Introduction to Evolutionary Computing*. Heidelberg: SpringerVerlag, 2003.

[15] O. Ibáñez, O. Cordón, S. Damas, and J. Santamaría, "An experimental study on the applicability of evolutionary algorithms to craniofacial superimposition in forensic identification," *Inf Sci*, vol. 79, pp. 3998–4028, 2009.

[16] O. Ibáñez, O. Cordón, and S. Damas, "A cooperative coevolutionary approach dealing with the skull–face overlay uncertainty in forensic identification by craniofacial superimposition," *Soft Comput*, vol. 18, pp. 797–808, 2012.

[17] O. Ibáñez, O. Cordón, S. Damas, and J. Santamaría, "An advanced scatter search design for skull-face overlay in craniofacial superimposition," *Expert Syst Appl*, vol. 39, n. 1, pp. 1459–1473, 2012.

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

[19] B. R. Campomanes-Álvarez, O. Ibáñez, F. Navarro, I. Alemán, O. Cordón, and S. Damas, "Dispersion assessment in the location of facial landmarks on photographs," *Int J Legal Med.*, vol. 129, n. 1, pp. 227–236, 2015.

[20] G. M. Gordon and M. Steyn, "An investigation into the accuracy and reliability of skull-photo superimposition in a south african sample," *Forensic Sci Int*, vol. 216, pp. 198.e1–6, 2012.

[21] L. A. Zadeh, *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lofti A. Zadeh*. Singapore: World Scientific Pub Co Inc, 1996.

[22] C. Birngruber, K. Kreutz, F. Ramsthaler, J. Krähahn, and M. Verhoff, "Superimposition technique for skull identification with AFLOAT® software," *Int J Legal Med*, vol. 124, pp. 471–475, 2010.

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

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

[25] L. Valencia-Caballero, *Metodología para elaborar reconstrucciones faciales empleando gráficos computarizados tridimensionales [in Spanish]*. PhD thesis, Laboratorio de Antropología Forense, University of Granada, Spain, 2007.

[26] O. Ibáñez, F. Cavalli, B. R. Campomanes-Álvarez, C. Campomanes-Álvarez, A. Valsecchi, and M. I. Huete, "Ground truth data generation for skull-face overlay," *Int J Legal Med*, vol. 129, 3, pp. 569–581, 2015.

[27] J. Glaister and J. C. Brash, *The medico-legal aspects of the Ruxton case*. Edinburgh: Harcourt Brace/Churchill Livingstone, 1937.

[28] F. Galton, *The Bertillon System of Identification. [Review of Signaletic Instructions, Alphonse Bertillon]*. Chicago, IL: McClaughry RW, vol. 54, pp. 569–570, 1896.

[29] P. Broca, *Instructions craniologiques et craniométriques de la Societé d'Anthropologie de Paris [in French]*, Paris: Bulletin de la Societé d'Anthropologie, vol. 6, pp. 534–536, 1875.

[30] B. A. Nickerson, P. A. Fitzhorn, S. K. Koch, and M. Charney, "A methodology for near-optimal computational superimposition of two-dimensional digital facial photographs and three-dimensional cranial surface meshes," *J Forensic Sci*, vol. 36, pp. 480–500, 1991.

[31] M. I. Huete, T. Kahana, and O. Ibáñez, "Past, present, and future of craniofacial superimposition: literature and international surveys," *Leg Med(Tokyo)*. In press, doi= doi:10.1016/j.legalmed.2015.02.001, 2015.

[32] A. K. Ghosh and P. Sinha, "An economised craniofacial identification system," *Forensic Sci Int*, vol. 117, pp. 109–119, 2001.

[33] W. Jin, G. Geng, K. Li, and Y. Han, "Parameter estimation for perspective projection based on camera calibration in skull-face overlay," in *Virtual Reality and Visualization (ICVRV), 2013 International Conference on*, pp. 317–320, 2013.

[34] O. Faugeras, *Three-Dimensional Computer Vision. A Geometric Viewpoint*. Cambridge: The MIT Press, 1993.

[35] H. K. Park, J. W. Chung, and H. S. Kho, "Use of hand-held laser scanning in the assessment of craniometry," *Forensic Sci Int*, vol. 160, pp. 200–206, 2006.

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

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

[38] M. Delgado, M. A. Vila and W. Voxman, "On a canonical representation of fuzzy numbers," *Fuzzy Sets Systs*, vol. 93, 1, pp. 125–135, 1998.

[39] I. Bloch, "On fuzzy distances and their use in image processing under imprecision," *Pattern Recognit*, vol. 32, pp. 1873–1895, 1999.

[40] "Face2skull^{TM} software." http://www.face2skull.com/. [Online; accessed 15-Jun-2015].

[41] M. Friedman, "A comparison of alternative tests of significance for the problem of m rankings," *Ann Math Stat*, vol. 11, pp. 86–92, 1940.

[42] O. J. Dunn, "Multiple comparisons among means," *Ann Math Stat*, vol. 56, pp. 52–64, 1961.

[43] S. García, A. Fernández, J. Luengo, and F. Herrera, "A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability," *Soft Comp*, vol. 13, pp. 959–977, 2009.

[44] S. P. Wright, "Adjusted p-values for simultaneous inference," *Biometrics*, vol. 48, pp. 1005–1013, 1992.

B. Rosario Campomanes-Álvarez received the Ph.D. degree in Computer Science (2015) from the University of Oviedo, Spain. She is postdoctoral researcher at the European Centre for Soft Computing, Spain. Her main research interests include soft computing, computer vision and forensic identification.

Oscar Ibáñez received the Ph.D. degree in Computer Science (2010) from the University of Santiago de Compostela, Spain. He is postdoctoral researcher at the University of Granada. His research interests include evolutionary computation, fuzzy logic, computer vision, machine learning and their applications to tackle real-world problems with a particular interest in forensic identification.

Sergio Damas received the Ph.D. in Computer Science from the University of Granada, Spain. Since 2011, he is the principal researcher of the "Fuzzy-Evolutionary Applications" at the European Centre for Soft Computing, Spain. His current research interests are in the fields of soft computing for forensic anthropology and medical imaging, fuzzy logic, computer vision, evolutionary computation and their applications to different real-world problems.

Carmen Campomanes-Álvarez received the M.S. degrees in Telecommunication Engineering (2011) and Information Technologies and Communications in Mobile Networks (2012) from the University of Oviedo, Spain. She is currently working toward the Ph.D. degree in Information and Communications Technologies with the University of Granada, Spain. Her research interests are focused on forensic identification using soft computing.

Oscar Cordón received his Ph.D. (1997) in Computer Science from the University of Granada, Spain, where he is now Full Professor at the Dept. of Computer Science and Artificial Intelligence. He is Associate Editor of 11 international journals and a reviewer for more than 40. His current main research interests are in the fields of soft computing for forensic anthropology and medical imaging, fuzzy rule-based systems and genetic fuzzy systems, evolutionary computation, ant colony optimization and other single and multi-objective metaheuristics, and their applications to different real-world problems.

Universidad de Granada

Escuela Internacional de Posgrado

ACCEPTANCE AND RESIGNATION OF THE PUBLICATION'S CO-AUTHORS

Publication/article: ***MODELING FACIAL SOFT TISSUE THICKNESS FOR AUTOMATIC SKULL-FACE OVERLAY***

The co-authors:

Mr./Ms.	B. Rosario Campomanes-Álvarez
Mr./Ms.	Oscar Ibáñez
Mr./Ms.	Sergio Damas
Mr./Ms.	Oscar Cordón
Mr./Ms.	
Mr./Ms.	

Declare that they:

☒ Accept and authorize the use of the above mentioned publication/article as part of the documentation for the deposit and defence of the doctoral thesis of Mr./Ms. **Carmen Campomanes Álvarez** titled ***AUTOMATION OF THE ASSESSMENT OF CRANIOFACIAL SUPERIMPOSITION USING SOFT COMPUTING AND COMPUTER VISION***,

☒ Have not used the above mentioned publication/article as part of the documentation for the deposit and defence of another doctoral thesis and/or refuse to use it for a future doctoral thesis

Granada, 2 , mayo, 2017

Signed: B. Rosario Campomanes-Álvarez

Signed: Oscar Ibáñez

Signed: Sergio Damas

Universidad de Granada

Escuela
Internacional
de Posgrado

Signed: Oscar Cordón   Signed: Signed:

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

- C. Campomanes-Álvarez, B. R. Campomanes-Álvarez, S. Guadarrama, O. Ibáñez and O. Cordón.   An Experimental Study on Fuzzy Distances for Skull-Face Overlay in Craniofacial Superimposition, Fuzzy Sets and Systems, vol. 318, pp. 100-119, 2017. DOI: http://dx.doi.org/10.1016/j.fss.2016.06.015.

  – State: Published.
  – Impact Factor (JCR 2015): 2.098.
  – Category: COMPUTER SCIENCE, THEORY AND METHODS. Order: 13/105. Q1.

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

---

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

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.
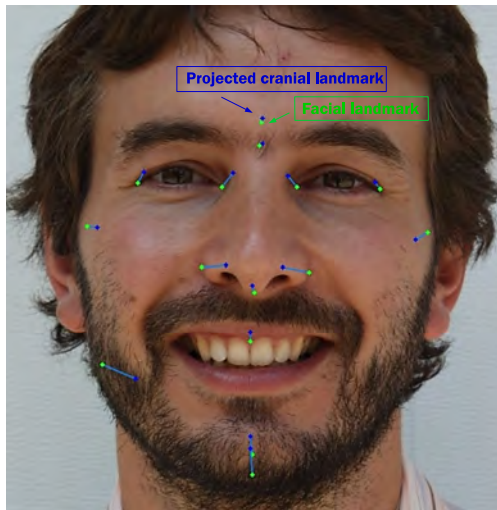
Figure 1: Scenario of the original genetic algorithm to search the best skull-face overlay: geometrical distance between crisp facial landmarks and projected crisp cranial landmarks.

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
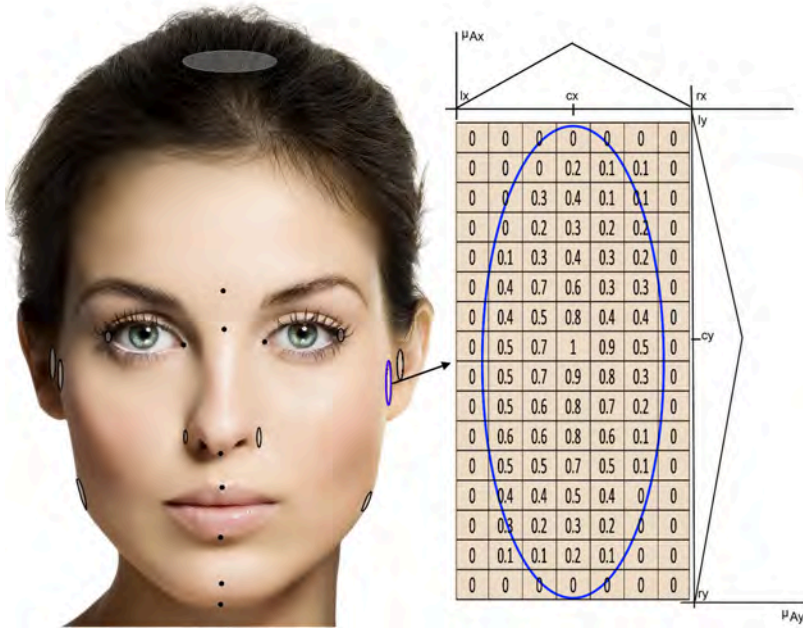
4

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:
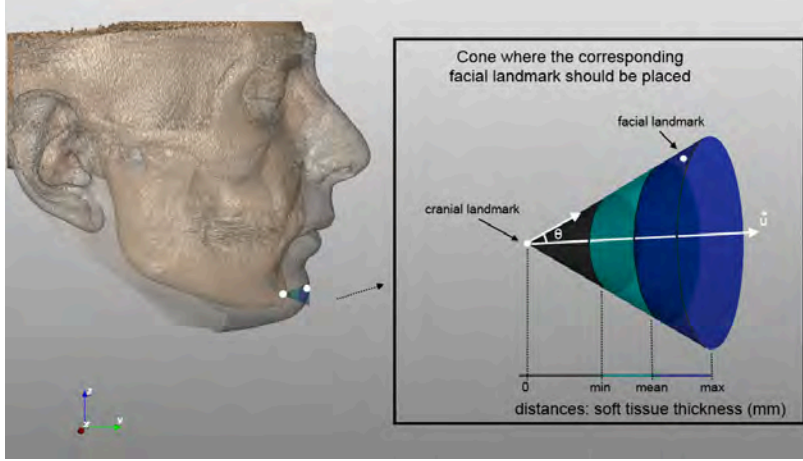
Figure 3: Facial landmark position from a cranial landmark using a cone. *min*, *mean*, and *max* are the soft tissue depths. $\vec{u}$ is the normal vector at the cranial landmark in the skull, θ is the rotation angle of $\vec{u}$.

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

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
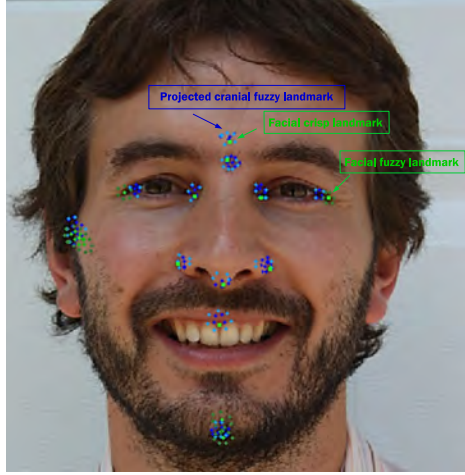
7

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.

$$d_5'(x, \tilde{B}) = \left\{ \begin{array}{ll} d_4'(x, \tilde{B}) & \text{if x } \notin \tilde{B} \\ 1 - \tilde{B}(x) & \text{if x } \in \tilde{B} \end{array} \right. \tag{12}$$

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].

$$d''_5(\tilde{A}, \tilde{B}) = \left\{ \begin{array}{ll} d''_4(\tilde{A}, \tilde{B}) & \text{if } \tilde{A} \cap \tilde{B} = \emptyset \\ 1 - \max \left\{ t[\tilde{A}(y), \tilde{B}(y)] \right\} & \text{if } \tilde{A} \cap \tilde{B} \neq \emptyset \end{array} \right. \tag{18}$$

   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].

$$d''_3(\tilde{A}, \tilde{B}) = \sum_{i=1}^{m} d(\tilde{A}_\alpha, \tilde{B}_\alpha) \cdot (\alpha_i - \alpha_{i-1}) \tag{19}$$

11

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:

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

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]:

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

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 $\tilde{D}_{FMD}$:

12

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

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
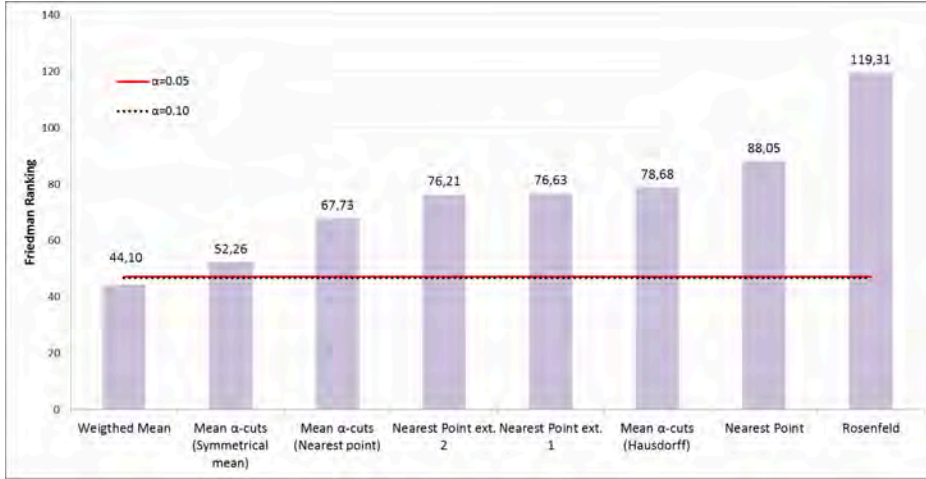
14

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
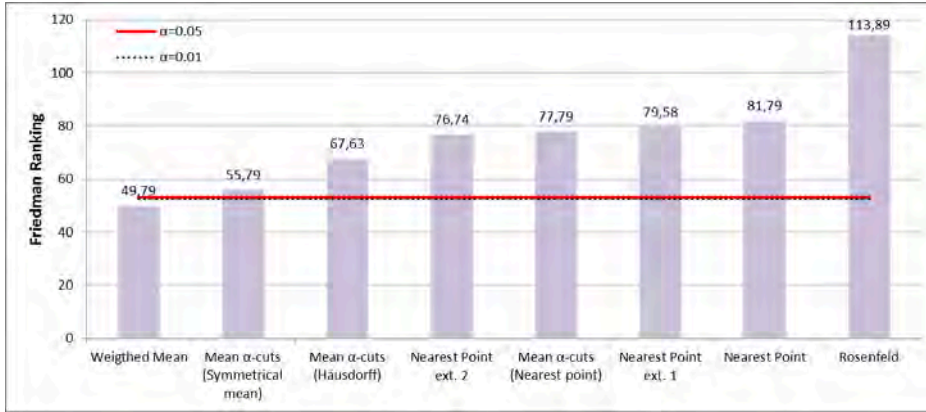

Figure 7: Friedman ranking and Bonferoni-Dunn lines for classification rate for the minimum fitness error in 30 runs.

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).

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
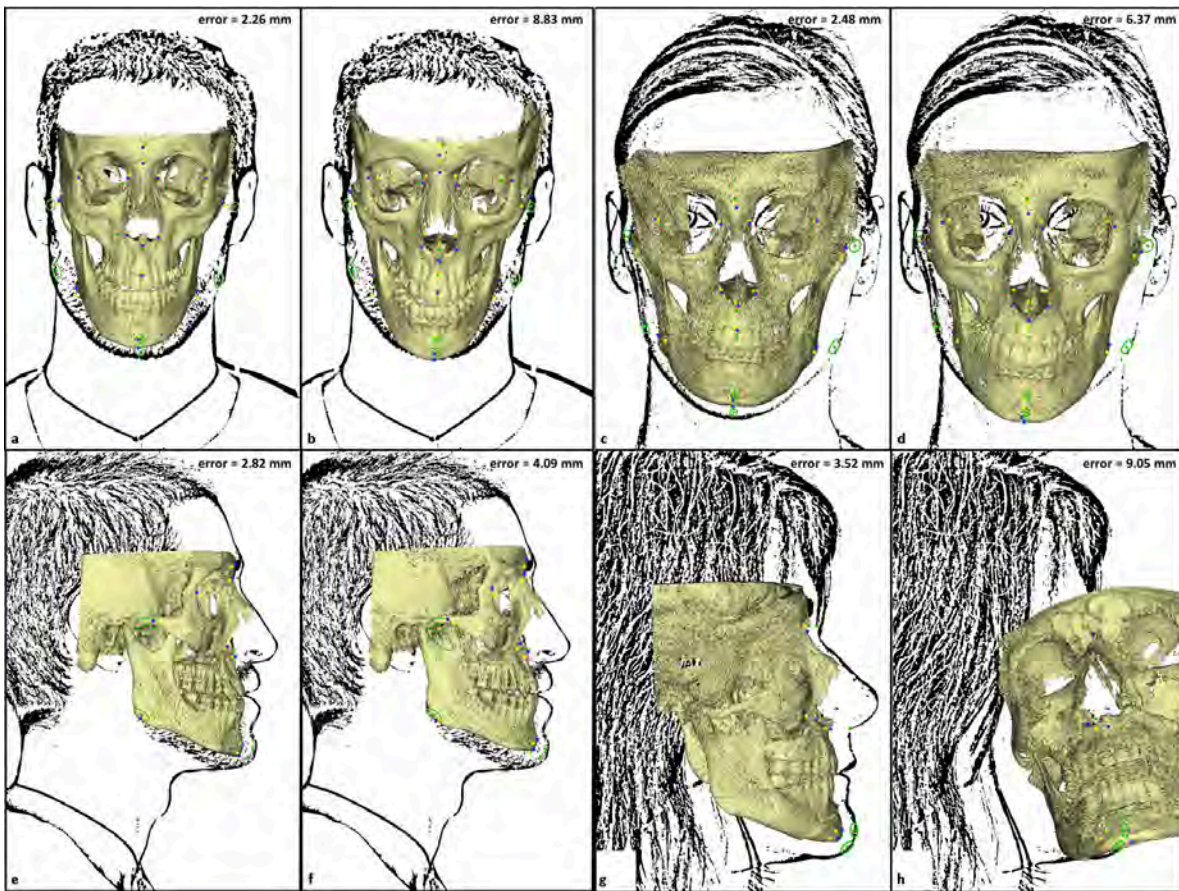
Figure 8: Visual results. Case 5, frontal view: (a) best result corresponds to the weighted mean distance, (b) worst result to the Rosenfeld distance. Case 8, frontal view: (c) best result with the weighted mean distance, (d) worst result with the Rosenfeld distance. Case 5, lateral view: (e) best result 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.

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 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

17

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.
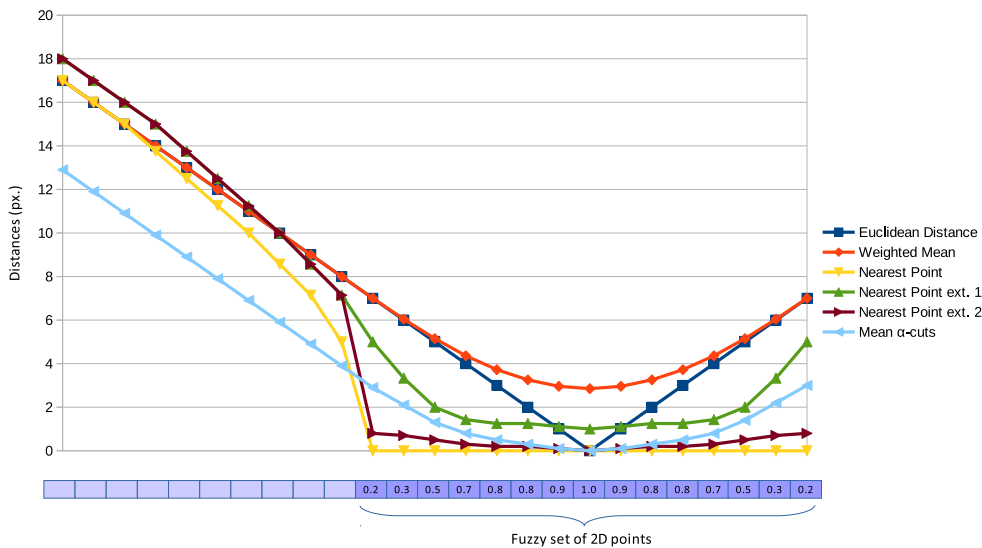

Figure 9: Performance of the different distances between a point and a fuzzy set. This example shows a simple case of an image, so the 2D points are pixels. X axis corresponds to the point from which the distance to the fuzzy set is calculated. Y axis represents the distance in pixels. The Euclidean distance is calculated from the point to the center of the set (point with membership value of one).

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.

18

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 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.

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.

[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.

ACCEPTANCE AND RESIGNATION OF THE PUBLICATION'S CO-AUTHORS

Publication/article: *AN EXPERIMENTAL STUDY ON FUZZY DISTANCES FOR SKULL-FACE OVERLAY IN CRANIOFACIAL SUPERIMPOSITION*

The co-authors:

Mr./Ms.	B. Rosario Campomanes-Álvarez
Mr./Ms.	Sergio Guadarrama
Mr./Ms.	Oscar Ibáñez
Mr./Ms.	Oscar Cordón
Mr./Ms.	
Mr./Ms.	

Declare that they:

☒ Accept and authorize the use of the above mentioned publication/article as part of the documentation for the deposit and defence of the doctoral thesis of Mr./Ms. **Carmen Campomanes Álvarez** titled *AUTOMATION OF THE ASSESSMENT OF CRANIOFACIAL SUPERIMPOSITION USING SOFT COMPUTING AND COMPUTER VISION*,

☒ Have not used the above mentioned publication/article as part of the documentation for the deposit and defence of another doctoral thesis and/or refuse to use it for a future doctoral thesis

Granada, 2 , mayo 2017

Signed: B. Rosario
Campomanes-Álvarez

Signed: Sergio Guadarrama

Signed: Oscar Ibáñez

Universidad de Granada

Escuela Internacional de Posgrado

Signed: Oscar Cordón

Signed:

Signed:

3. Design of Criteria to Asses Craniofacial Correspondence in Forensic Identification based on Computer Vision and Fuzzy Integrals

Vision and Fuzzy Integrals   *125*

3   Design of Criteria to Asses Craniofacial Correspondence in Forensic Identification based on Computer Vision and Fuzzy Integrals

- C. Campomanes-Álvarez, O. Ibáñez and O. Cordón. Design of Criteria to Asses Craniofacial Correspondence in Forensic Identification based on Computer Vision and Fuzzy Integrals, Applied Soft Computing, vol. 46, pp. 596 - 612, 2016. DOI: http://dx.doi.org/10.1016/j.asoc.2015.11.006.

  - State: Published.
  - Impact Factor (JCR 2015): 2.857.
  - Category: COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE. Order: 21/130. Q1.

Design of Criteria to Assess Craniofacial Correspondence in Forensic Identification based on Computer Vision and Fuzzy Integrals

Carmen Campomanes-Alvarez[a,*], Oscar Ibáñez[a], Oscar Cordón[a,b]

[a]*Department of Computer Science and Artificial Intelligent, 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.*

Abstract

Craniofacial superimposition is one of the most relevant skeleton-based identification techniques. Within this process, the skull-face overlay stage focuses on achieving the best possible overlay of a skull found and an *ante mortem* image of a candidate person. In previous work, we proposed an automatic skull-face overlay method, based on evolutionary algorithms and fuzzy sets. The following stage, decision making, consists of determining the degree of support of being the same person or not. This decision is based on the analysis of some criteria assessing the skull-face morphological correspondence through the resulting skull-face overlay. In this work, we take a first step to design a decision support system for craniofacial superimposition. To do so, we consider the modeling of two of the most discriminative criteria for assessing craniofacial correspondence: the morphological and spatial relationship between the bony and facial chin, and the relative position of the orbits and the eyeballs. For each criterion, different computer vision-based approaches have been studied. The accuracy of each method has been calculated as its capability to discriminate in a cross-comparison identification scenario. Sugeno integral has been used to aggregate the results of the different methods taking into account the corresponding individual accuracy index. This allows us to provide a single global output specifying the matching of each criterion while combining the capabilities of different methods. Finally, the performance of the designed criteria and methods have been tested on 172 skull-face overlay problem instances of positive and negative cases to illustrate the discriminative power of each criterion. It has been shown that thanks to the use of Sugeno integral for aggregating different methods, a more robust measurement output is achieved.

Keywords: Forensic identification, Craniofacial superimposition, Decision making, Computer vision, Soft computing, Spatial relations, Sugeno integral

1. Introduction

Skeleton-based identification methods have been under continuous investigation within the forensic anthropology and odontology communities [1]. Craniofacial superimposition (CFS) [2], one of the approaches in craniofacial identification [3, 4], is a representative technique of this kind. It involves superimposing a skull onto a number of ante-mortem (AM) images of a missing person and the analysis of their morphological correspondence to determine if they belong to the same subject.

Three consecutive stages have been distinguished for the whole CFS process in [5]. The first stage involves the acquisition and processing of the skull (or skull 3D model) and the AM facial images, followed by the location of the craniometric and facial landmark. The second stage is the skull-face overlay (SFO), which focuses on achieving the best possible superimposition of the skull and a single AM image of the missing

---

person. This process is repeated for each available photograph, obtaining different overlays. Thus, SFO corresponds to what traditionally has been known as the adjustment of the skull size and its orientation with respect to the facial photograph [2, 6]. Finally, the resulting superimpositions are analyzed in a third stage for decision making. It consists of determining the degree of support of being the same person or not (exclusion) by considering the different criteria studying the anatomical relationship (spatial and morphological) between the skull and the face. These criteria can vary depending on the region and the pose [7].

There is a strong interest in designing automatic methods to support the forensic anthropologist to put CFS into effect. In particular, the design of computer-aided CFS methods has experienced a boom over the past twenty years [8]. The most recent approaches use skull 3D models, which are employed in this contribution as well.

The works developed by authors such as [9, 10, 11, 12, 13, 14] serve as examples of how computer algorithms, specially computer vision [15] and soft computing techniques [16], can automate SFO and accommodate the uncertainty/fuzziness of some facial landmarks [17] and of the soft tissues [18]. These methods represent a clear step forward since they have managed to reduce time and subjectivity inherent to manual approaches applied by forensic anthropologists. However, the quality of the obtained overlays is influenced by several sources of uncertainty, as well as by partial and incomplete knowledge about skull-face anatomical correspondence. Thus, reaching an optimal accuracy is still an open field of research and manual refinement of SFO results is currently needed for such a purpose.

Once one or several appropriate skull-face overlays are obtained, forensic experts evaluate spatial and morphological skull-face relationships in the third stage. To do so, they focus on certain regions that demonstrated to be more discriminative. The final decision is provided in terms of strong, moderate or limited support to the assertion that the skull and the facial image belong to the same person or not [7]. This is a subjective process that relies on the forensic expert's skills and the quantity and quality of the used materials. Hence, a decision support system (DSS) is desirable to take their decision in a faster and more objective way. It would also open the door to the application of CFS to identification scenarios involving multiple comparisons. Our long-term, very complex goal is to design such a DSS based on the evaluation of the said spatial and morphological relations. This system will provide a numeric index as output, aiming to support to forensic anthropologist to take the CFS final decision.

Computational methods in the fields of computer vision (CV) and soft computing (SC) can be extremely useful for this aim. CV includes techniques for processing, analyzing, segmenting and registering image data in an automatic way [15]. Meanwhile, SC is aimed for the design of intelligent systems to process uncertain, imprecise and incomplete information [16]. SC methods applied to real-world problems often offer more robust and tractable solutions than those obtained by more conventional mathematical techniques. Two of the main SC techniques are fuzzy set theory and fuzzy logic. They extend classical logic to provide a conceptual framework for knowledge representation under imprecision and the consequent uncertainty [19]. Fuzzy integrals, in general, and Sugeno integrals, in particular, are well known to be one of the most powerful and flexible aggregation operators. They permit the aggregation of information under different assumptions on the independence of the information sources [20].

In this work, we take a first step to design a DSS for CFS. To do this, we model two of the most discriminative criteria for assessing craniofacial correspondence. Namely, the morphological and spatial relationship between the bony and facial chin, and the relative position of the orbits and the eyeballs. To model the former criterion, we have implemented some CV methods aimed to measure how the chin facial shape follows the skull shape given the delineation of these regions in a particular overlay. That process involves the proper extraction of the two chin curves (from the region given at hand) and the subsequent analysis of the relationship between them. Similarly, we have developed two methods to measure the relative position between the orbit and the center of the eyeball for the latter criterion. We have also implemented an adapted version of the state of the art methods in order to compare them with our proposal's performance.

Regardless the criterion type, we have performed a study to analyze different ways to aggregate the outputs of the measurement methods. The accuracy of each method is calculated as its capability to discriminate in the decision making process (ranking positive and identification negative cases). Sugeno integral [21] has been used to aggregate combinations of the different methods taking into account the

corresponding individual accuracy index. Thus, it serves to provide a global output specifying the matching of each criterion in the specific skull-face overlay. Finally, we have tested these methods on 172 skull-face overlay problem instances of positive and negative cases to illustrate the discrimination power of each criterion.

Notice that, the combination of the proposed DSS for the third CFS stage make up a complete hybrid intelligent system to support the forensic anthropology in the automation of the CFS task. It is based on the use of fuzzy integrals [21] and the two existing methods employed for the first and the second stages are based on evolutionary algorithms [29] and fuzzy sets [19].

The structure of this paper is organized as follows: in Section 2, we review previous proposals dealing with forensic anthropology based on CV and SC and introduce our automatic SFO approach. Section 3 outlines the main issues related to the final decision making stage in the CFS process. In Section 4, we explain our general methodological proposal for the development of a CFS DSS. In Section 5, we introduce the experimental setup, the corresponding results and their analysis. Finally, in Section 6 we remark the conclusions and the related future works.

2. Preliminaries

2.1. Computer Vision and Soft Computing Techniques in Forensic Anthropology

Computational methods as CV and SC can be extremely useful for the automation of the CFS decision making process. The guiding principle of these methods is perfectly adapted to the way in which reasoning and deduction have to be performed in forensic science. In fact, several successful applications of these techniques in forensic anthropology have been developed so far. They include age estimation [22], skull 3D modeling [23], facial soft thickness prediction [24], facial identification [25] and skull 3D model simplification [26]. Specifically, fuzzy integrals have been used for face recognition [27] and estimation of skeletal age-at-death [28]. Within CFS, evolutionary algorithms (EAs) [29] and fuzzy sets [19] have being used to tackle SFO in an automatic way [9, 10, 11, 12]. The following Section 2.2 summarizes the previous SFO system since the overlays analyzed in this work have been obtained using it.

2.2. Automatic Skull-Face Overlay

The SFO process requires positioning the skull in the same pose as the face in the photograph. From a CV point of view, the AM 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 in a given pose [30].

The most natural way to deal with the SFO problem is to replicate that original scenario. To do so, a 3D model of the skull must be used. Current 3D scanners provide skull 3D models with a precision of less than one millimeter in a few minutes [31]. The goal is to adjust its size and its orientation with respect to the head in the photograph [2]. In addition, the specific characteristics of the camera must also be replicated to reproduce the original situation as much as possible [30]. To do this, the skull 3D model is positioned in the camera coordinate system through geometric transformations, i.e. translation, rotation and scaling. The goal is to adjust the skull size and its orientation to be at the same angle as the face in the image [2]. Then, a perspective projection of the skull 3D model is performed onto the facial photograph.

Hence, a 3D-2D image registration process (IR) [32] where these unknown parameters are estimated seems to be the most appropriate formulation to automate SFO. In fact, that process directly replicates the original scenario in which the photograph was taken [9, 30].

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 on both the skull 3D model and the facial photograph (See Fig. 1). Once the location of these landmarks is provided by the forensic anthropologist, the SFO procedure is based on automatically 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
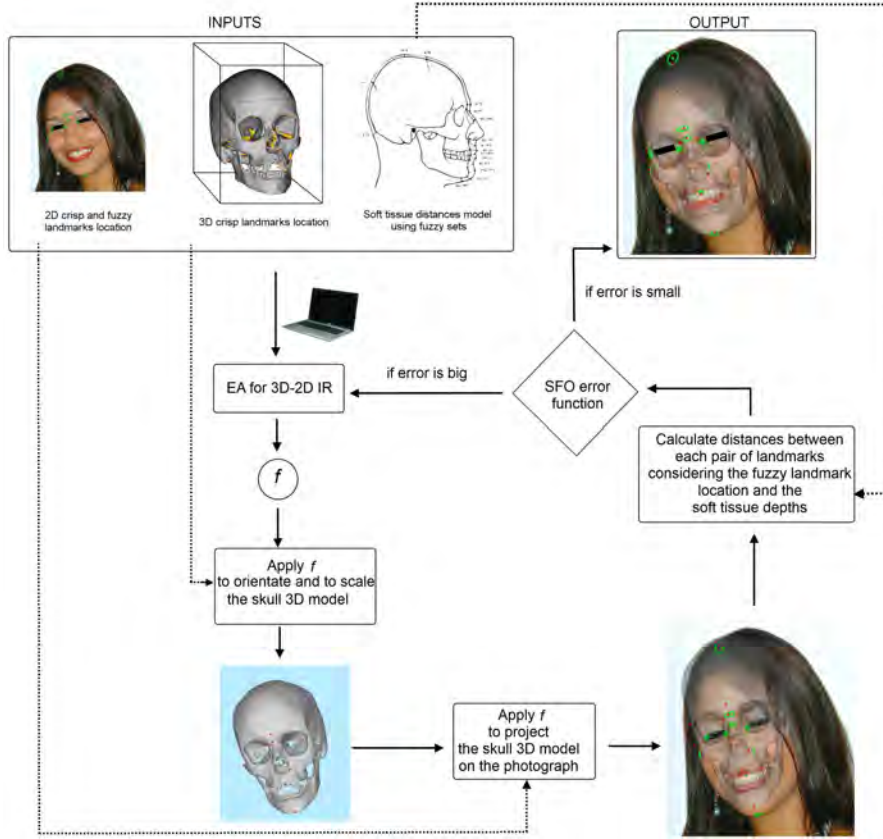
3

Figure 1: Scheme of the SFO procedure as a 3D-2D IR problem [30]

was modeled in [9] as a set of geometric operations. These operations involve 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: 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's 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:

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

The rotation matrix R turns the skull to the same pose as the head in the photograph. S, T, and P are scaling, translation and perspective projection matrices, respectively [9]. A complete description of the matrices of Eq. 1 is detailed in [33].

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 [9]:

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

4

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 onto the photograph, d is the 2D Euclidean distance, and N is the number of landmarks placed by the expert.

Fuzzy landmarks in photographs are used to jointly deal with the imprecise landmark location and the coplanarity problem [10]. Besides, since the correspondence between facial and cranial landmarks is not always symmetrical and perpendicular [34], it is also modeled by fuzzy sets in [12]. To do so, the available information concerning soft tissue depths [18] is considered.

Therefore, the original definition of our evolutionary SFO technique's fitness function in Eq. (2) was modified in [12] to take into account distances between two fuzzy sets, *Fuzzy Mean Error* (FME), as follows:

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

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 considered ($N = Ncrisp + Nfuzzy$), 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, 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. Decision Making in Craniofacial Superimposition

Once one or several skull-face overlays have been achieved for the same identification case, the main goal is to determine the degree of support that the skull and the face of the photograph(s) belong to the same person or not. This degree of support is based on the consistency of the matching between the face and the skull but it is also influenced by the quality and quantity of the materials used (photographs and skull). A scale for a craniofacial matching evaluation has been recently defined by some of the most representative experts in craniofacial identification in [7]. Accordingly, the final decision is provided in terms of strong, moderate or limited support.

This decision is guided by different criteria studying the anatomical relationship between the skull and the face. According to the literature [35], we can distinguish the following families of criteria for assessing the craniofacial correspondence:

1. Analysis of the consistency of the bony and facial outlines/morphological curves. Forensic experts confirm if two particular curves (of skull and face) are anatomically consistent. That is, if two curves follow the same shape or, in other words, if one curve is along the other.
2. Assessment of the anatomical consistency by positional relationship. These criteria consist of a positional relationship analysis in order to assess anatomical consistency. Thus, the goal is to check if the relative position of a skull region against a facial region is right with respect the anatomical reference.
3. Location and comparison of lines to analyze anatomical consistency. Experts analyze a set of marking lines (obtained by joining some reference landmarks) on the face and on the skull. In terms of CV, these lines have to be parallel in an image.
4. Evaluation of the consistency of the soft tissue thickness between corresponding cranial and facial landmarks. The last set of criteria consists of analyzing the consistency of the facial soft tissue thickness considering distances between pairs of homologous landmarks (located on the skull and the face). These distances can be checked in the skull-face overlay using existing studies about soft tissue thickness in different human populations [18].

Our long term goal is to automate the whole decision making process by modeling the previous four sets of criteria using CV and SC techniques. The resulting system would give as a result a global degree of support of a CFS identification to assist the forensic anthropologist to take her/his decision. In [36] we presented a very brief preliminary version of this task. In this contribution, we take a first step in the design of such system by modeling a criterion from the first family mentioned above.

There are just a few works tackling the automation of the analysis of craniofacial correspondences within the framework of CFS identification. This particular identification technique represents challenging problem, very linked to the forensic anthropologist's expertise and non-automated yet. The existing literature was published more than 18 years ago and the works are very basic and limited. In addition, they do not consider the use of either skull 3D models or computer techniques to perform the skull-face overlay. The techniques for the shape analysis also implies manual interaction. In [37], k-th-order polynomial functions and Fourier harmonic analysis are applied to assess the fit between the outline of the skull and the face. In [38], the authors present a skull identification system where the distance between the landmarks and the thickness of the soft tissue of the anthropometrical points are semi-automatically measured on the computer screen for the assessment of the anatomical consistency between the digitized skull and face. The software also includes polynomial functions and Fourier harmonic analysis for evaluating the match of outlines. To extract the outline, edge segmentation thresholding operations are used. In addition, the final evaluation parameter is the root of the average quadratic error and sum of the difference in amplitude value of the fists Fourier harmonics. This final value is not normalized and authors do not establish a threshold value on which to base positive or negative identification through appropriate procedures of discriminant analysis.

These previous works are semi-automatic methods. In order to adjust the correspondence of the two compared curves, facial image curves are manually shifted along the abscissa and ordinate. Their main pitfall is that they provide a value that does not take into account the actual spatial relation between skull and face since the methods employed are invariant to translation, scale and rotation. Besides, they rely on the 2D delineation of regions of the skull, what clearly involves a source of error and imprecision. Finally, these systems only implement a single group of the criteria to assess the craniofacial correspondence among the four existing ones. Our proposal is aimed to solve the latter drawbacks, properly modeling the underlying "along" relationship while considering the whole information provided by the delineation of regions on the 3D skull model.

A different approach is presented in [39] where an algorithm calculates the distance between crosses manually marked by the expert in both face and skull radiograph images. The mean value of the total distance between crosses represents the index of similarity between the given face and skull: the smaller the index value, the greater the similarity. Again, this index is not normalized, so this method is only valid for cross-comparisons but it does not provide an index of correspondence in a particular skull-face overlay. The algorithm matches a facial photograph to the correct skull in 100% of the cases following a completely different approach in comparison with ours: distances between anatomical points vs. the analysis of the consistency of morphological traits. However, the former is only valid under limited and unpractical conditions: the skull and face images have to be acquired in precise anterior position. Thus, the validity of the method has not been tested on real CFS identification scenarios and it serves more as a proof of craniofacial correspondence uniqueness. Besides, it is strongly sensitive to the SFO accuracy, a complex problem not addressed in that work since they superimpose images acquired under controlled conditions. Thus, it does not make sense to compare against our method since we work with real identification cases with no limitations on the *ante mortem-post mortem* photographs conditions.

As a consequence, our proposal is the first automated DSS for CFS. Our method includes the automatic segmentation of the contours. The analysis takes into account the relative position and the distance between the two objects as well as the shapes. There is no previous work in the specialized literature using CV methods to measure the craniofacial correspondence between chin shapes. In the same way, there are not studies measuring the correct position of the orbit and the eyeball center in a SFO. It will also integrate the accuracy and discriminative power of each criterion within the decision process. Thus, it will provide the expert with a more informed, objective, and accurate CFS DSS design.

4. Methodology

The decision making stage shows certain characteristics leading us to think on CV and fuzzy set theory as powerful tools to automate it by developing:

- The proper extraction of curves or contours of a facial region in an automatic way.

- The modeling of the criteria for assessing the skull-face relationship.

- The integration of different methods to evaluate a specific spatial skull-face relation.

- The aggregation of multiple degrees of support, each of them indicating the consistency between the skull and face guided by a particular criterion and from different facial regions.

- The integration of the quality and quantity of the materials within the decision process.

The need of fuzzy set-based approaches is justified by the imprecision on the manual delineation of facial regions. It is also needed to model the semantics of some anthropometrical relationships: "the prosthion lies posterior to the anterior edge of the upper lip", or "the nasion is higher than the nasal root", for example. In addition, they can provide a powerful tool for information fusion in order to take the final identification decision. In particular, fuzzy integral shows interesting capabilities to provide a support of the final decision. This technique combines the individual sources, the expected worth of subsets of these sources, and the experts' knowledge on them. Thus, we aim to use fuzzy integral as a multifactorial way to analyze methods taking into account any variable (known or observed) that can be quantified to reach a decision (as done in previous approaches like [28]). Thanks to its use, the uncertainties and errors quantified along the whole process (e.g. the uncertainty in contours location) can be accumulated in order to consider their influence in the final decision.

As described in Section 3, one of the most important criteria used by forensic experts to assess whether the skull belong to a specific person is the consistency of the bony and facial outlines or morphological curves. Another important criterion is the assessment of the anatomical consistency by checking the corresponding positional relationship. This consists of a positional relationship analysis in order to assess anatomical consistency. Thus, the goal is to measure the relative position of a skull region against a facial region and determine the degree of anatomical consistency of the resulting relative position. As expected, these two criteria depend on the facial pose within the photograph.

The regions under study in both the 3D skull and facial photograph have to be delineated in order to model the previous criteria using CV. Then, the 3D regions are projected into the 2D image (using the resulting geometric transformation obtained in the SFO stage) to analyze their consistency with the corresponding facial contour. Of course, this task is affected by imprecision and subjectivity in the location of the different regions involved.

The reached value to measure each criterion will be more robust and flexible if more than one method to model the same criterion is taken into account. A scheme of our current proposal for the decision making system is shown in Figure 2. We study several CV-based methods to evaluate each of the said criteria. Their description is explained in the following sections. After this, we introduce the process to calculate the accuracy index associated to each specific method. Then, we propose a way to aggregate their corresponding outputs to provide a global matching degree for each criterion. These outputs should have into consideration their reliability in the decision making process.

4.1. Modeling the Anatomical Consistency of the Bony and Facial Outlines (First family of Criteria)

By analyzing the existing studies of this area, we have identified the three most common relations between the bony curves (B) and the face curves (F) used by forensic experts as follows:
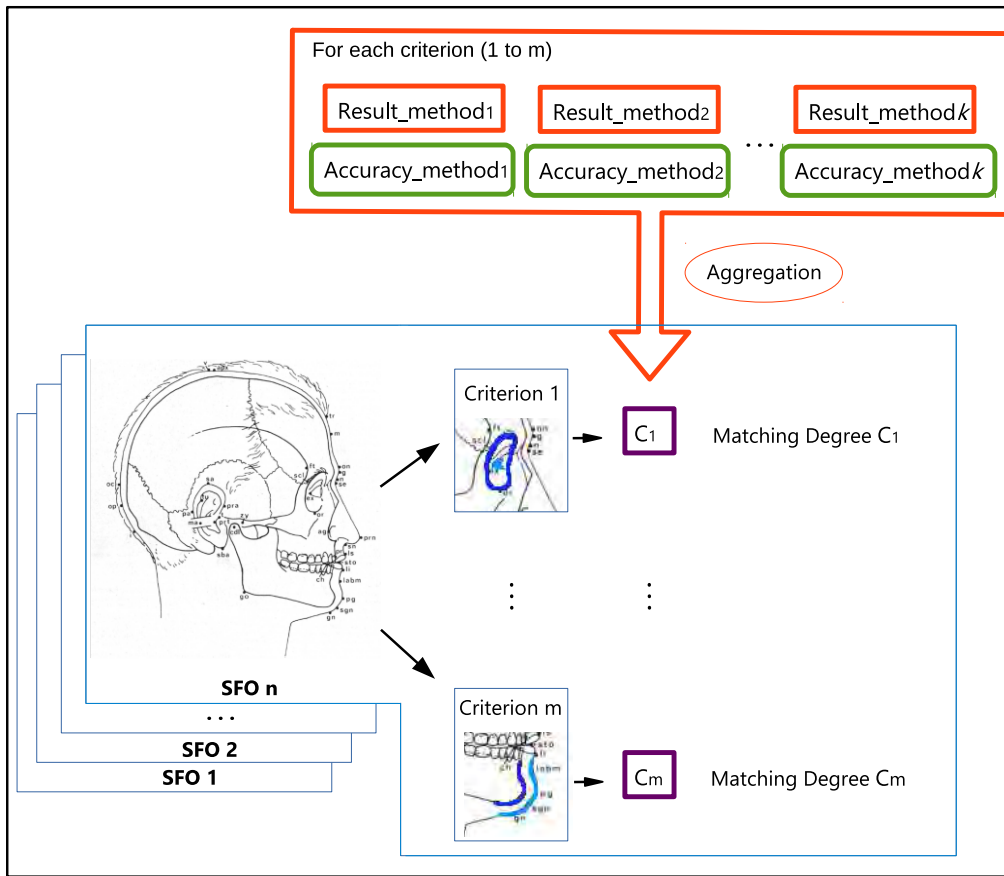
- F follows B.

- F is consistent with B.

Figure 2: Scheme of the present DSS for CFS.

- F is the mirror image of B.

Regarding to their modeling, the former two have similar effects (see Fig. 3 for a visual example). Notice that, these kinds of relations are established in different facial regions. For example, the outline of the frontal bone follows the forehead outline and the chin outline is consistent with the mental outline.

In order to compare both curves there is a need to extract the corresponding contour of the skull and facial regions (the bony outline) under study (see Fig. 3). Automating contour extraction using CV techniques is a difficult task when precision and robustness are demanded despite photographs and 3D models resolution, quality, noise or occlusions. In addition, it could require a different approach depending on the region at hand. Thus, manual delineation of both corresponding contours by the forensic expert have been preferred. The main idea is to select the corresponding curve of the skull that is located nearer to the facial curve of the photograph (See Fig 3). In this third stage (decision making), we just work with optimized overlays obtained in the previous step, so the skull will be (more or less) correctly superimposed onto the facial photograph. Thus, in the case when the bony outline crosses with the facial one, the studied criterion is evaluated as a zero matching since this fact is anatomically impossible.

As a proof of concepts of the proposed methodology two different kinds of methods to model the "F follows B" relation are considered: the spatial relation *along* and shape similarity measures.

4.1.1. Spatial relation "Along"

The spatial relation *along* gives us the degree to which an object A is along an object B. The considered approach is based on computing a degree of elongatedness of the region lying *between* A and B, i.e. the
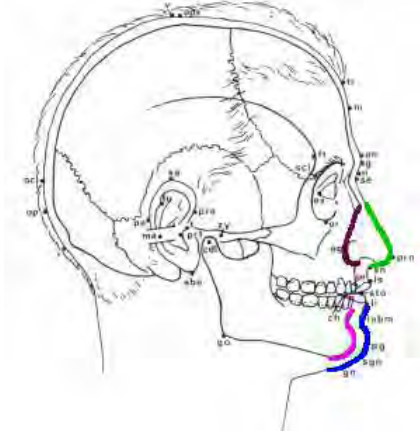
8

Figure 3: Example of two criteria used by forensic experts in lateral view to assess the consistency between bony and facial outlines. The chin facial curve follows the skull shape while the protruding part of the nose cartilage is the mirror image of the cartilage filling the pyriform aperture of the cranium.

between region [40]. Thus, it directly models the specific case of the "F follows B" relation by measuring whether the projected bone curve follows its counterpart facial curve in the 2D facial image.

Two main steps have to be performed: i) calculate the *between* region (noted by β), and ii) measure how elongated it is (thus defining the degree to which A is along B).

In [41] the authors present different methods to calculate the region between two objects. In this work we have decided to use the model based on the convex hull of the union of the two objects due to its lower computational complexity and similar performance.

Once the *between* region is obtained, the alongness measure may be computed as:

$$S_1 = f^a(P(\beta)^2/A(\beta)), \tag{4}$$

$$S_2 = f^a(P(C_{IN})^2/A(\beta)), \tag{5}$$

where P and A represent the perimeter and the area of β, and C_{IN} is the inner contour that corresponds to those contours portions which belong to β (in Fig 5 $P(C_{IN}) = P(C_B) + P(C_F)$). To normalize this measure between 0 and 1, we use a sigmoid function as $f_a(x) = (1 - exp(-ax))/(1 + exp(-ax))$, taken from [40]. These measures S_1 and S_2 tend towards 1 as β becomes more elongated. Absolute values can be changed by tuning parameter a to enhance the differences depending on the specific application.

Based on this idea, we propose another method to model the "F follows B" relationship. It consists of measuring the alongness of the *between* region using the classical definition of the circularity ratio of a shape [42]. The circularity ratio is the ratio of the area of a circle having the same perimeter:

$$C(\beta) = 4\pi \frac{A(\beta)}{P(\beta)^2} \tag{6}$$

where A is the area of the shape and P is the perimeter (in this case the area and the perimeter of the *between* region). This expression gives us a value in the interval [0,1] that tends towards 1 as the region β becomes more circular. So, the final formula for our model will be:

$$S_3 = 1 - C(\beta) \tag{7}$$

where S_3 tends toward 1 as β becomes more elongated, and accordingly when a curve is along the other.

These three versions of the spatial relation "along" have the problem that they are sensitive to the distance between the contours under comparison together with their length. First and second version (Eqs.

9

4 and 5 respectively) can be adapted to each case by tuning parameter a. However, this is not a desirable solution for an automatic system.

4.1.2. Similarity Measures

Similarity measures between shapes have been widely studied in the CV community. They are usually employed to identify objects in images despite the different location, size, rotations or deformations [43]. They are obtained from a distance metric and thus are affected by the properties of such distance. Hence, there is a need to select the most appropriate one for the problem at hand to achieve a good performance.

A common way to study the similarity between two shapes consists of comparing both shapes represented by their features (using shape parameters or shape description techniques). Shape parameters are simple geometric features that can be used to describe shapes in a general way (i.e.: Circularity ratio, eccentricity, Elliptic variance, Convexity, etc.) but they can only discriminate shapes with large differences. For our study, we propose to compare the circularity ratio of both shapes (Eq. 6). In order to determine the similarity between these two values (in the $[0, 1]$ interval), we use the following expression based on [44]:

$$S_4 = 1 - \frac{|C(F) - C(B)|}{C(F) + C(B)} \tag{8}$$

where F and B represent the facial and the bony chin shape, respectively.

Alternatively, there are many different image analysis methods, contour-based and region-based, to measure the representation and description of a shape [45]. We will only focus on contour-based methods due to the impossibility to precisely locate the bony region within a facial photograph. In addition, these measures are usually employed to compare objects of the same kind. So in general, the required features are scale, rotation, and translation invariance. As we have explained, that is not exactly the case of our problem.

In [42] the authors present a survey of the existing approaches for shape-based feature extraction, classifying them according to their processing approaches:

- One-dimensional function for the shape representation.

- Polygonal approximation.

- Spatial interrelation feature.

- Moments.

- Scale-space methods.

- Shape transforms domains.

Based on this overview, we have selected two methods belonging to the first group (one-dimensional functions) for our application. This choice has been made due to their low computational complexity and their interesting properties of being sensitive to rotation changes:

- Shape signatures. They represent a shape by a one-dimensional function derived from shape boundary points.

- Chain code representation. They describe an object by a sequence of unit-size line segments with a given orientation.

These methods are introduced as follows.

Shape Signatures

A shape signature represents a shape by a one dimensional function derived from shape boundary points. Complex coordinates, centroid distance function, tangent angle, curvature function, area function, triangle-area representation and Chord length function are the commonly used shape signatures [42]. For our application we will use the following shape signatures:

- Complex coordinates:

  A complex coordinates function is simply the complex number generated from the coordinates of boundary points, $P_n(x(n), y(n)), n \in [1, N]$:

  $$z(n) = [x(n) - g_x] + i[y(n) - g_y] \tag{9}$$

  where (g_x, g_y) is the centroid of the shape, given by:

  $$\begin{cases} g_x = \frac{1}{N} \sum_{i=0}^{N-1} x_i \\ g_y = \frac{1}{N} \sum_{i=0}^{N-1} y_i \end{cases} \tag{10}$$

- Centroid distance function:

  The centroid distance function is expressed by the distance of the boundary points from the centroid (g_x, g_y) (Eq. 10) of a shape

  $$r(n) = [(x(n) - g_x)^2 + (y(n) - g_y)^2]^{1/2} \tag{11}$$

- Area function:

  When the boundary points change along the shape boundary, the area of the triangle formed by two successive boundary points and the center of gravity also changes. This forms an area function which can be exploited as shape representation.
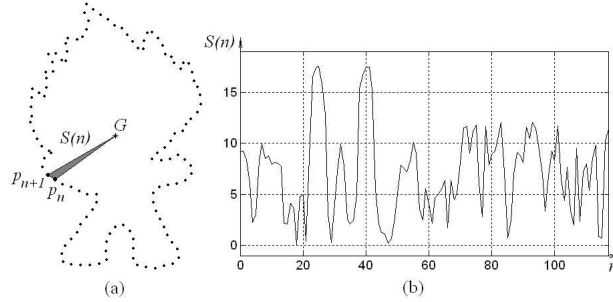

Figure 4: The original contour of a shape (a) and its area function (b) [42].

In order to measure the similarity between two curves we compute the Euclidean distance of their corresponding shape signatures. The following formula gives a similarity value between zero and one:

$$S_4 = 1 - \sqrt{\frac{1}{N} \sum_{i=1}^{N} (s_F(i) - s_B(i))^2} \tag{12}$$

with s_F and s_B being the shape signature of each contour.

11

Chain Code

Chain code is a common approach for representing contours. It describes an object by a sequence of unit-size line segments with a given orientation [45]. Freeman [46] introduced a chain code that describes the movement along a digital curve by using so-called 8-connectivity or 4-connectivity. The direction of each movement is encoded by the numbering scheme $\{i | i = 0, 1, 2, ..., 7\}$ or $\{i | i = 0, 1, 2, 3\}$ denoting a counter-clockwise angle of $45° \times i$ or $90° \times i$ regarding the positive x-axis.

In our implementation, we compute the similarity between boundaries by comparing their chain codes. Using 8-connectivity, the formula can be defined as follows:

$$S_5 = 1 - \sum_{i=1}^{N} d_i / N * 4 \tag{13}$$

where

$$d_i = \begin{cases} d_i^* & \text{if } d_i^* \leq 4 \\ 8 - d_i^* & \text{if } d_i^* > 4 \end{cases}, \tag{14}$$

$$d_i^* = |ch(F) - ch(B)| \tag{15}$$

with $ch(X)$ being the chain code representation of an object, d_i the distance between two chain code representations, and N the number of points of the objects. Notice that we normalize the final measure with respect to $N * 4$ since 4 is the largest distance between two chain codes using 8-connectivity (i.e., the opposite direction).

This method is very sensitive to noise or to small variations. Once we have the chain code representation of each contour, we calculate the mean of a number of consecutive elements (w) and we compare the final 'smoothed' chain code representations to reduce this sensibility.

These approaches (shape signatures and chain code comparison) imply both contours must have the same number of points. To solve this problem a sampling method is applied in one of them.

4.1.3. Case of Study: Mental and Chin Outlines Comparison

Within the group of criteria that study the consistency of the bony and facial morphological curves, experts have stated that the chin outline has to be consistent with the mental outline [35]. This relationship between bony and facial contours can be modeled using the spatial relation described in Section 4.1.

In our case, the goal is to analyze if both shapes are similar, but not identical, since they belong to different objects (the skull and face have a different size). At the same time, the distance between them (according to soft tissue depth), their orientation and relative position (anatomical correspondence) need to be taken into account. Hence, different from most common applications of similarity measures (mainly in the field of image retrieval), we need some degree of translation, scale, and rotation sensitivity while analyzing shape similarity. Rotation sensitivity is critical in order to model "F follows B" spatial relation. Translation and scale differences are expected due to the anatomical distance and different sizes of the regions under comparison.

To do so, we use a method that involves the convex hull of the union of the two objects in order to segment the contours. The first step involves obtaining the region between both objects defined as [41]:

$$\beta_{CH}(F, B) = CH(F \cup B) \cap F^C \cap B^C \tag{16}$$

where CH(X) denotes the convex hull of X and X^C its complement.

The basic idea of this method is that the contour of the *between* region is composed of an alternate sequence of line segments of the convex hull and the boundary of the two objects. The intersection points are the pairs of subsequent vertexes belonging to different objects. These points divide each contour into two portions (inner and outer contours). The inner contours correspond to the skull and face contours that we want to analyze (refer to Fig. 5 for an illustrative example). In some cases this method has a wrong behavior under some circumstances as depicted in Fig. 5 b. That problem is solved setting the intersection points of
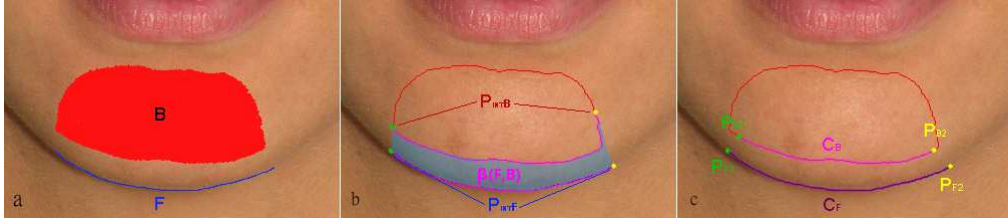
Figure 5: Chin contours segmentation. (a) Original regions, facial (F) and bony (B). (b) Convex hull $CH(F \cup B)$ with intersection points between $\beta(F, B)$ and CH. (c) Final contours with corresponding extreme points.

the bony contour as the nearest points to the corresponding intersection points of the facial contour (Fig. 5 c). By using this segmentation method, the section of the contours under comparison changes according to their relative position and size. Thus, despite the method employed to model the "F follows B" relation, the resulting approach will be sensitive to translation and scale.

Once the contour segmentation has been accomplished, we have to study whether the facial curve follows the bony curve. With this segmentation method, when both curves cross, it is not possible to extract the contours, so the result value will be zero. This result concurs with the fact that the crossing of these two regions is morphologically impossible.

4.2. Measuring the Relative Position of Two Regions for Studying Anatomical Consistency (Second Family of Criteria)

The second group of criteria consists of a positional relationship analysis in order to assess anatomical consistency. Thus, the goal is to measure the relative position of a skull region against a facial region and determine the degree of anatomical consistency of the resulting relative position. For instance, the lateral angle of the eye must lie within the lateral wall of the orbit in both lateral and oblique views (see Fig. 6).
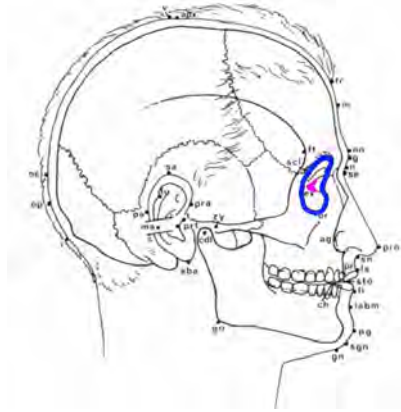

Figure 6: Consistency between the lateral angle of the eye and the orbit.

In a 2D image, any positional relation between two objects A and B can be obtained easily. The main directional spatial relationships are shown in Fig. 7. To model these relationships we take the relative position between two points as a base [47]. The angle between the segment joining two points a and b and the x-axis of the coordinate frame is denoted by $\vartheta(a, b)$ (see Fig. 8). This angle takes values in the range $[-180, +180]$.

Relations "left", "right", "above" and "below" can be modeled using $cos^2\vartheta$ and $sin^2\vartheta$ functions [47], as shown in Fig. 8. The former functions can be directly used as fuzzy membership functions to establish a membership degree in $[0, 1]$ for each relation. To model "diagonally", we combine two of these relations. For the case of the Fig. 8, a is 0.5 to the left of b, 0.5 above b, 0.0 to the right of b and 0.0 below b.
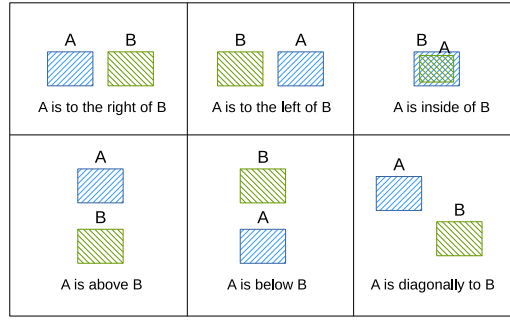
13

Figure 7: Main directional relative position between two objects.

The positional relationship between two 2D objects can be computed in two different ways [48]:

1. Aggregation method. It consists of computing the relationship between each pair of points in A and B and calculate the "A relation B" as the aggregated value of all of them. We choose the average mean for this application.

2. Centroid method. It is based on calculating the positional relation between both centroids.

Thus, the positional relation between two objects A and B is expressed by a degree $\delta(A, B) \in [0, 1]$ for each position.
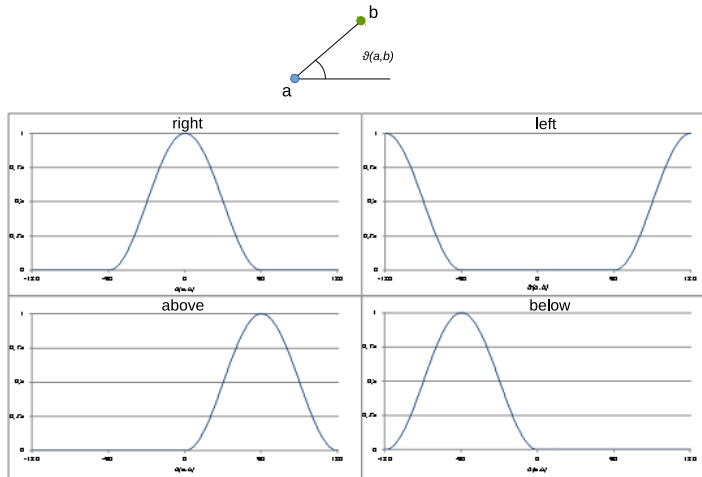

Figure 8: Functions to calculate the relative position between two points a and b.

Once the relative position between two regions is obtained, we need to assess if that position is anatomically consistent. One choice is comparing the current relative position against one reference. Specialized forensic anthropology literature [35] normally refers to the relative position of anatomical regions according to lateral and frontal facial pose within the photograph. However, these criteria always depend on the photograph view and in most of the cases the facial pose is not exactly frontal or lateral. Alternatively, we propose to compare the actual relative position against the relative position given by a 3D reference model.

From a given set of 3D head models we could create statistical mean facial and skull 3D models. Then, we can apply the resulting geometric transformation obtained in the SFO stage to the reference 3D models (skull and face) onto a white image with the same size as the photograph. Proceeding in this way, the positional relationship between skull and face regions in the reference case is obtained and can be compared with the positional relationship between the same regions in the given case (See Fig. 9).

14

We should note that this option is only applicable when the positional relationships are not significantly affected by population factors like ancestry, age, sex, etc.

In order to compare two different relative positions between two objects, we compute the similarity using the following expression based on [44]:

$$S_6 = 1 - \frac{|\delta_a(A,B) - \delta_a(C,D)| + |\delta_b(A,B) - \delta_b(C,D)| + |\delta_r(A,B) - \delta_r(C,D)| + |\delta_l(A,B) - \delta_l(C,D)|}{\delta_a(A,B) + \delta_a(C,D) + \delta_b(A,B) + \delta_b(C,D) + \delta_r(A,B) + \delta_r(C,D) + \delta_l(A,B) + \delta_l(C,D)} \quad (17)$$

where δ_a, δ_b, δ_r and δ_l are the degrees of the position relation "above", "below", "right", and "left", respectively.
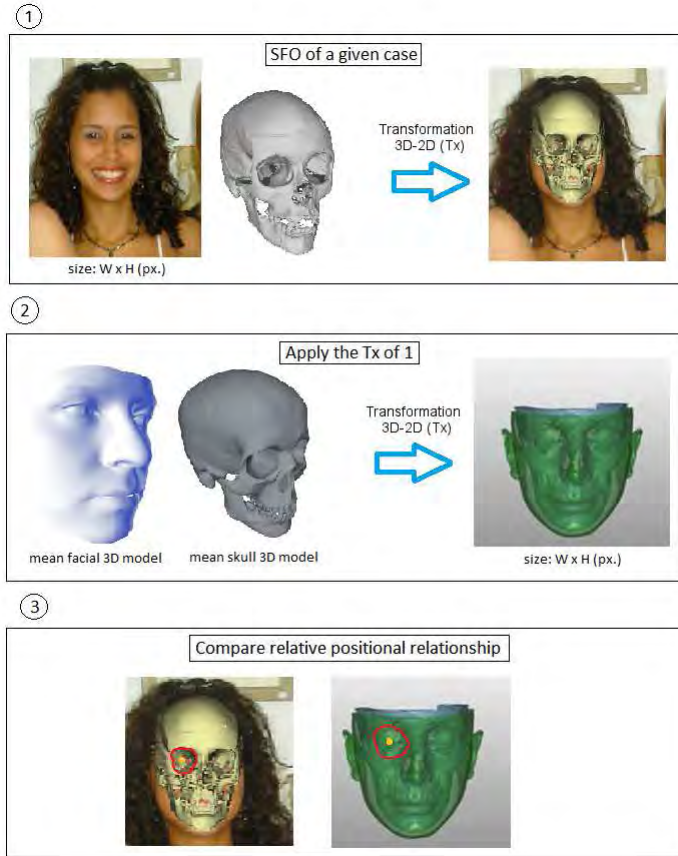

Figure 9: Scheme of the solution to solve the dependency on the photograph view to check relative position between skull and face.

4.2.1. Case of Study: Orbits and Eyeball Center

One of the criteria that focuses on analyzing the relative position of two regions involves studying the anatomical consistency in the position of the orbits and the center of the eyeball. This relationship is independent of populations factors like race, age, sex, etc. [49]. For this reason, we decided to use a reference model to compare with each specific case, as explained in the previous section.

To evaluate this criterion, we need to extract the appropriate contours both for the orbits (bony region) and eyeball center (facial region). The desired orbital contour is the interior contour, selected as the smaller one, while the facial region is simply the contour of the marked zone on the photograph (see Fig. 10).
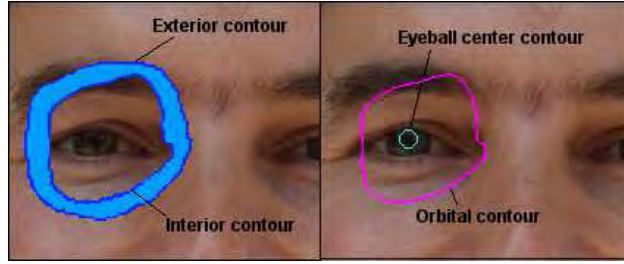
Figure 10: Orbital contour segmentation.

4.3. Definition of an Accuracy Index for the Computer Vision Methods Associated to the Different Criteria

Different CV methods can be employed to automatically evaluate the degree of fulfillment of each specific anatomical relationship. For example, "F follows B" relation can be evaluated by using different matching methods, as explained in Section 4.1. Each of these methods provides a different value to measure the same criterion. The aggregation of these different values is expected to give a more robust and reliable final output. Fuzzy integrals, which will be explained in the next subsection, are very good tools for this aim.

This kind of aggregation process implies that each method has a specific weight associated. In order to set these weights in a proper way, we define the accuracy of a method as its capability to discriminate in the decision making process. Hence, this accuracy is specified as how well the pertinent relationship between skull and face is modeled by that method for a specific criterion.

The accuracy is calculated from a performance study on positive and negative identification cases. A similar procedure is implemented in [28] for the case of a skeletal age estimation system based on the aggregation of the outcomes from different methods. In our approach, positive cases are skull-face overlays where the skull and the face belong to the same person, and the 3D model skull is perfectly located onto the facial image. Hence, we use a ground truth dataset as positive cases, whose skull-face overlays can be assumed as optimal. These overlays were created from living subjects submitted to cone-beam computed tomography (CBCT) for clinical purposes [50]. Skull-face overlays of the negative cases have been obtained by our automatic method in [12] using parameter values reported in that contribution. We use 172 cases of which 18 are positive and 154 negative.

The accuracy index of a particular method for a specific criterion is calculated as follows. First, we obtain the value that corresponds to apply each method for a specific criterion over the database of cases. Next, the matching values reported are used to rank the candidates based on their chance to be the actual subject. Then, a value between 0 and 1 is assigned to each positive case taking into account this ranking: if the method reported the highest value to a positive case (first position of the ranking), 1 is assigned. On the contrary, if the position of a positive case is the last of the ranking, 0 is assigned. In general, we propose to use the following formula to assign the accuracy of the method x_i in the instance j:

$$Acc(x_i)_j = 1 - \frac{r-1}{M_j - 1} \tag{18}$$

where r is the position of the positive case in the ranking and M_j is the lower value of the ranking for the instance j (all cases getting the same criterion-method value are supposed to have a draw, that is, they are assigned to the same ranking). Finally, we calculate the average of all these accuracy values over all cases. As a result, an accuracy index in the interval $[0, 1]$ is achieved for each method.

Table 1 shows an illustrative example of the accuracy index calculation for two different methods using an instance composed of a positive case and four negative cases.

4.4. Aggregation of the Outputs of the Individual Methods to Get the Criterion Matching Degree

As explained in Section 4.1, there is a need to aggregate multiple degrees of support. Each of these degrees correspond to a specific criterion guided by a particular method with an associated accuracy index. For instance, different methods to evaluate the anatomical consistency of the bony and facial curves are

16

Cases	Results Method 1	Rank	Accuracy index	Results Method 2	Rank	Accuracy index
Case 1 (positive)	0.9	1	$1 - \frac{1-1}{5-1} = 1$	0.7	3	$1 - \frac{3-1}{4-1} = 0.33$
Case 2 (negative)	0.5	4		0.8	2	
Case 3 (negative)	0.7	3		0.9	1	
Case 4 (negative)	0.8	2		0.7	3	
Case 5 (negative)	0.3	5		0.5	4	

Table 1: Example of the accuracy index calculation

shown in Section 4.1. The reached final decision will be more robust if multiple methods are taken into account, considering their reliability for the combination of results.

Aggregation operators are the basic tools for achieving such objective. From a general point of view, these operators combine data from different sources to build an aggregated value. Fuzzy integrals are well known to be one of the most powerful and flexible functions aggregation operators. They can be used in a great variety of decision making applications. In particular, they can be used to model situations in which sources are independent as well as others where such independence cannot be assured [51].

One of the most well-known integrals is the Sugeno integral [20]. The flexibility of this operator is due to the fact that it generalizes several of the most widely-known and used aggregation functions. In particular, it generalizes the arithmetic and weighted mean, as well as the median and linear combination of order statistics.

Fuzzy integrals combine the data supplied by several information sources according to a fuzzy measure. This fuzzy measure, that represents the background knowledge on the information sources, is a set function from the set of information sources into an appropriate domain (e.g. the $[0,1]$ interval). This fuzzy measure can represent the importance or relevance of the sources when computing aggregation, as well as the reliability or the confidence, or similar concepts like satisfaction.

The fuzzy measure is defined as follows:

Definition 1. Let $X = \{x_1, ..., x_p\}$ be a finite set of criteria or attributes, and let F be the family of all subsets of X. A fuzzy measure g on the finite set X is a set function $g : F \rightarrow [0,1]$ satisfying the following conditions:

1. $g(\phi) = 0$, $g(X) = 1$ (boundary condition)
2. $A \subseteq B$ implies $g(A) \leq g(B)$ for all $A, B \in F$ (monotonicity)

To determine a fuzzy measure on X, we must identify $2^p - 2$ coefficients satisfying $p2^{p-1}$ conditions. To solve this drawback, some approaches have been proposed to reduce the number of parameters to be determined [21]. In this paper, we use a Sugeno λ-measure defined as in [52]:

Definition 2. Let g be a fuzzy measure, then g_λ is a Sugeno λ-measure if there exists $\lambda > -1$ such that

$$g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A)\mu(B) \tag{19}$$

holds for all $A, B \in F$.

It is to be noted that, for a Sugeno λ-measure,

$$\prod_{j=1}^{p}(1 + \lambda g_\lambda(\{x_j\})) = 1 + \lambda \tag{20}$$

holds because of the boundary condition [52].

Once λ is found, the Sugeno integral can be calculated.

Definition 3. Let g be a fuzzy measure on X, then a Sugeno integral of a function $f : X \rightarrow [0,1]$ with respect to the fuzzy measure g_λ is defined by

17

$$\oint f \circ g = \vee_{j=1}^{p}(f(x_{(j)}) \wedge g(x_{(1)}, ..., x_{(j)})) \tag{21}$$

where $f(x_{(i)})$ indicates that the index have been permuted so that $0 \leq f(x_{s(1)}) \leq ... \leq f(x_{s(p)}) \leq 1$, $A_{s(j)} = x_{s(j)}, ..., x_{s(p)}$ and $f(x_{s(0)}) = 0$.

If λ equals zero, then μ shows that μ is a probability measure. The values $g^1, g^2, ..., g^p$, are called the fuzzy densities. The densities are interpreted as the importance of the individual sources. The measure of a set A of information sources is interpreted as the importance of that subset of sources toward answering a particular question (such a class membership).

Using the above definitions $g_\lambda(A)$ can be constructed from the fuzzy densities of the elements of A for any subset A of X. Given the set of densities, the value λ can be easily found as the unique root greater than -1 of a simple polynomial [53]. Thus, estimating the densities is a core problem when using the Sugeno measures [20].

The $\vee$ in the Sugeno integral is a t-conorm and the $\wedge$ is a t-norm [19]. These are generally chosen to be the maximum and minimum, respectively, for the Sugeno integral.

The fuzzy integral differs from other aggregation methods in that objective evidence from the individual sources and the expected worth of subsets of these sources is considered within the fusion procedure. The result of the fuzzy integral is the support in the hypothesis being tested. Support, as used here, is the strength of the hypothesis acquired by fusing these two distinct different sources of information [28].

In our application, x_i represents a specific method and $f(x_i)$ is the result of applying this method to a specific criterion. Accordingly, $g(x_i)$ corresponds to the accuracy index of the x_i method, computed as described in Sec. 4.3.

5. Experimentation

Some experiments have been developed to analyze the ability of the automatic methods to measure the morphological correspondence between bony and facial chin shapes, and the relative position of the orbits and the eyeball centers. Both positive and negative identification cases have been considered in order to study the performance of our methods in real cases.

First, we test and compare eight different methods for measuring the mental and chin outlines spatial correspondence and the aggregation of the best three. In particular, a proposal based on the circularity of the region between, and three different variants of the spatial relation along: version 1 corresponds to Eq. 4, version 2 corresponds to Eq. 5, and version 3 is our proposal (Eq. 8). In addition, five shape similarity methods have been employed to measure the degree in which the facial chin curve follows the bony mental outline: i) the comparison between the circularity ratio of both shapes; ii) the complex coordinates signatures; iii) the centroid distance function signatures; iv) the area function signatures; v) and the Chain code representations (as introduced in Section 4.1). Secondly, we study the criterion of the relative positions between the orbits and the eyeball centers using the two methods introduced in Section 4.2: the aggregation method and the centroid method.

In next sections we explain the experimental setup and we depict the obtained results and the corresponding analysis for the specific criterion.

5.1. Experimental Setup

The experimental design involves the sex-based cross-comparison of nine CBCT skull models of living individuals from Italy and Spain (four females and five males) against 38 photographs (18 females and 20 males). Two positive photographs (lateral and frontal views of the same person) are available for each skull and 16 (corresponding to seven different persons) and 18 (corresponding to 10 different persons) negative ones for females and males, respectively. The SFOs have been obtained by our automatic method in [12] using the parameter values reported in that contribution. For the positive cases, we use a ground truth dataset, whose SFO are assumed as optimal [50]. Table 2 summarizes the composition of the dataset employed along the experiments explained later.

18

The experimental procedure consists of two parts. In the first one, we calculate the accuracy of each individual method taking into account all the cases (positive and negative instances, 172 in total). This process is detailed in Section 4.3. This step is only executed once in order to obtain the accuracy index of each method. Then, for each particular SFO case, we aggregate three of the best methods using all the possible combinations for the chin region and the two methods for the eyeball region (using the accuracy index calculated previously) in order to compare the output results.

Finally, we compare our proposal with the semi-automatic methods existing in the literature [37, 38]. Both works are based on polynomial functions and Fourier harmonic analysis for evaluating the matching of the bony and facial outlines. Hence, the use of these techniques can be considered for the first group of criteria. For this reason, we can only apply this method for our case of study regarding the mental and chin outlines comparison.

In order to establish an objective comparison, the contours of the skull and the face are extracted using our method to segment the chin contours (explained in Section 4.1.3). Then, we follow the same criterion of the authors in [37, 38] and the two contours are shifted along the abscissa and ordinate. On the one hand, the polynomial function coefficients are calculated by an interpolation-regression algorithm with the least-squares method. The order of the polynomial degree is 10. The distance between the two polynomial curves is measured on pixels. This distance reveals the reciprocal point-to-point difference (10 points) between the two curves. Then, the sum of this point-to-point difference is calculated. The authors transform this pixel distance to millimeters. They do it by converting the measurement between two specific pair of landmarks into the number of pixels of the same measurements on the image. We keep the distance in pixels due to the unreliability of this conversion despite being sensitive to the image's resolution. We refer to this result as the Pesce-Yoshino's distance. On the other hand, the amplitude of the Fourier harmonics are calculated for both contours. The output parameter is the sum of the harmonic amplitude differences between the skull and the face. For this reason, this method is called Pesce-Yoshino's harmonic in this manuscript.

As we have mentioned in Section 3, these two output values are not normalized and authors do not establish a critical threshold on which to base a degree of consistency between the skull and the face. In this case, in order to compare with our results, we calculate the accuracy following our definition (Section 4.3). We use each output results (Pesce-Yoshino's distance and Pesce-Yoshino's harmonic) to rank the candidates. Then, a value between 0 and 1 is assigned to each positive case taking into account this ranking, as we do for the other methods. Proceeding in this way, we ensure obtaining a fair comparison procedure.

	Females		Males		Total	
Skulls	4		5		9	
	Positives	Negatives	Positives	Negatives	Positives	Negatives
Per skull	2	16	2	18		
Total	8	64	10	90	18	154

Table 2: Experimentation dataset summary

5.2. Results and Discussion

The morphological correspondence between chin shapes was firstly evaluated individually applying the eight proposed methods over each photograph. In this experiment, we have not taken into account one of the positive cases (case 3.1) since that person had her chin operated on after the photograph was taken. The average accuracy index for each individual method and for different combinations of methods as well as the corresponding λ parameter of the Sugeno integral are shown in Table 3.

In view of these results, the best individual performance is achieved with the area function comparison method, with an accuracy degree of 0.834. The complex coordinates signatures performs the second best with 0.814. Finally, the third best result is obtained with centroid distance signatures comparison, having a 0.721 of accuracy. The chain code comparison achieves an accuracy equal to 0.612. The worst performance is obtained by the comparison between the circularity ratios and the three variants of the spatial relation

Methods	Parameter λ	Accuracy index
Along v1 (a=0.05) (M1)		0.230
Along v2 (a=0.05) (M2)		0.275
Along v3 (M3)		0.235
Circularity ratios (M4)		0.433
Complex coordinates (M5)		0.814
Centroid distance (M6)		0.721
Area function (M7)		0.834
Chain codes (M8)		0.612
Pesce-Yoshino's distance [37, 38]		0.475
Pesce-Yoshino's harmonic [37, 38]		0.494
M5 $\oplus$ M7	-0.954525	**0.843**
M7 $\oplus$ M8	-0.874205	0.809
M5 $\oplus$ M8	-0.855186	0.814
M5 $\oplus$ M7 $\oplus$ M8	-0.986125	0.812

Table 3: Accuracy degrees obtained for the proposed methods to model the morphological correspondence between chin contour and their aggregations using Sugeno Integral.

"along". These methods achieve an accuracy of 0.433, 0.230, 0.275 and 0.235, respectively. The semi-automatic methods from the state of the art [37, 38] show worse performance than most of our proposed alternatives. They get an intermediate position in the accuracy index ranking. As we can see the use of the Pesce-Yoshino's distance to rank the candidates achieves an accuracy of 0.475 and the use of the Pesce-Yoshino's harmonic achieves 0.494 of accuracy.

	M5	M7	M8	M5 $\oplus$ M7	M7 $\oplus$ M8	M5 $\oplus$ M8	M5 $\oplus$ M7 $\oplus$ M8
Case 1.1	0.907	0.823	0.903	0.823	0.823	0.903	0.903
Case 1.2	0.820	0.845	0.935	0.834	0.845	0.820	0.845
Case 2.1	0.793	0.835	0.880	0.834	0.835	0.793	0.835
Case 2.2	0.872	0.839	0.925	0.839	0.839	0.872	0.872
Case 3.2	0.828	0.793	0.943	0.814	0.793	0.828	0.828
Case 4.1	0.847	0.833	0.750	0.833	0.833	0.814	0.833
Case 4.2	0.836	0.895	0.900	0.836	0.895	0.836	0.895
Case 5.1	0.838	0.785	0.908	0.814	0.785	0.838	0.838
Case 5.2	0.769	0.780	0.885	0.780	0.780	0.769	0.780
Case 6.1	0.810	0.789	0.887	0.810	0.789	0.810	0.810
Case 6.2	0.895	0.892	0.903	0.892	0.892	0.895	0.895
Case 7.1	0.749	0.770	0.908	0.770	0.770	0.749	0.770
Case 7.2	0.816	0.831	0.830	0.831	0.831	0.816	0.831
Case 8.1	0.729	0.674	0.910	0.729	0.674	0.729	0.729
Case 8.2	0.808	0.820	0.868	0.820	0.820	0.808	0.820
Case 10.1	0.862	0.859	0.913	0.859	0.859	0.862	0.862
Case 10.2	0.614	0.779	0.830	0.779	0.779	0.614	0.779

Table 4: Obtained results for the best performing proposed methods and their combinations for all the positive cases in the chin shapes analysis.

We have combined three of the best methods (area function signature, complex coordinates signatures and chain codes) using the four possible combinations, i.e., combining the elements by pairs (three aggregations) and the three together using the Sugeno Integral. The methods with an accuracy lower than 0.5 were discarded for their poor performance. We also discarded the centroid distances method since it has a similar behavior to the area function signature (correlated but lower values) so the aggregation of these two does not contribute to obtain better results. Table 3 shows the corresponding parameter λ calculated for each

aggregation and the resulting accuracy degree. The higher accuracy index is obtained aggregating the area function and the complex coordinates comparison (0.848), improving the performance of the individual methods. The other combinations do not achieve better results than their individual counterparts: the accuracy of the area function and the chain codes comparison aggregation is 0.809, the combination of the complex coordinates and the chain code achieves 0.814, and the aggregation of the three gets 0.812. For each method and method combinations, we show the output results (Table 4), the ranking and the accuracy index (Table 5).

	M5	M7	M8	M5 ⊕ M7	M7 ⊕ M8	M5 ⊕ M8	M5 ⊕ M7 ⊕ M8
Case 1.1	1.000 (1)	1.000 (1)	0.857 (2)	1.000 (1)	1.000 (1)	1.000 (1)	1.000 (1)
Case 1.2	0.857 (2)	1.000 (1)	1.000 (2)	1.000 (1)	1.000 (1)	0.857 (2)	1.000 (1)
Case 2.1	1.000 (1)	1.000 (1)	0.500 (2)	1.000 (1)	1.000 (1)	1.000 (1)	1.000 (1)
Case 2.2	1.000 (1)	1.000 (1)	1.000 (1)	1.000 (1)	1.000 (1)	1.000 (1)	1.000 (1)
Case 3.2	0.857 (2)	0.857 (2)	1.000 (1)	1.000 (1)	0.857 (2)	0.857 (2)	0.857 (2)
Case 4.1	0.833 (2)	1.000 (1)	0.167 (6)	1.000 (1)	1.000 (1)	0.833 (1)	0.833 (1)
Case 4.2	0.833 (2)	1.000 (1)	0.667 (3)	1.000 (1)	1.000 (1)	0.833 (2)	1.000 (1)
Case 5.1	1.000 (1)	0.455 (7)	0.818 (3)	0.909 (2)	0.455 (7)	1.000 (1)	1.000 (1)
Case 5.2	0.818 (3)	0.455 (7)	0.182 (10)	0.364 (8)	0.455 (7)	0.818 (3)	0.364 (8)
Case 6.1	1.000 (1)	0.875 (2)	0.250 (7)	1.000 (1)	0.875 (2)	1.000 (1)	1.000 (1)
Case 6.2	1.000 (1)	1.000 (1)	0.750 (3)	1.000 (1)	1.000 (1)	1.000 (1)	1.000 (1)
Case 7.1	0.500 (8)	0.786 (4)	0.857 (3)	0.286 (11)	0.786 (4)	0.500 (8)	0.286 (11)
Case 7.2	0.714 (5)	1.000 (1)	0.214 (12)	1.000 (1)	1.000 (1)	0.714 (5)	0.857 (3)
Case 8.1	0.308 (10)	0.154 (12)	0.923 (2)	0.231 (11)	0.154 (12)	0.308 (10)	0.231 (11)
Case 8.2	0.846 (3)	1.000 (1)	0.154 (12)	1.000 (1)	1.000 (1)	846 (3)	0.846 (3)
Case 9.1	1.000 (1)	1.000 (1)	1.000 (1)	0.933 (2)	1.000 (1)	1.000 (1)	1.000 (1)
Case 9.2	0.267 (12)	0.600 (7)	0.133 (14)	0.533 (8)	0.600 (7)	0.267 (12)	0.533 (8)

Table 5: Accuracy index and ranking positions (in brackets) for the best performing proposed methods and their combinations for all the positive cases in the chin shapes analysis.

Tables 6 and 7 show the analogous results for the modeling of the orbits and the eyeball center criterion. For these experiments, there are 18 positive instances, nine of which correspond to frontal view and 9 to lateral view. Hence, we have 18 positive cases in frontal view (for each eye) and 9 positives cases in lateral view.In view of the figures in Table 8, the combination of the two methods tested achieves the highest accuracy (0.922), improving the behavior of the individual methods (0.894 for the aggregation method and 0.861 for the centroid method). According to Tables 3 and 8, the criterion based on analyzing the relative position of the orbits and the eyeball centers is able to discriminate better than the criterion of the chin contours analysis.

The time consumed for obtaining the accuracy index of each method and criterion was 8 min 15.75 sec for the eight methods of the chin region and 58 min 32.91 sec for the two methods of the eyeball region overall. It is important to remark that this process is only executed once. The average real times taken by execution of the best methods are shown in Table 9. These times were obtained using an Intel Corel^{TM} 2 Quad CPU Q8400 2.66 GHz, with 4GB RAM, running Windows 7 Professional^{TM}. They are the mean times for an overlay of an identification case for each criterion. The chin method corresponds to execute the comparison of the area function signatures and the comparison of the complex coordinates signatures, and then the calculation of the aggregation of both using the Sugeno integral. In the case of the eyeball region, the execution time corresponds to the execution of the two methods (aggregation and centroid) and the subsequent aggregation with Sugeno integral.

21

	M'1 Right	M'2 Right	M'1 ⊕ M'2 Right	M'1 Left	M'2 Left	M'1 ⊕ M'2 Left
Case 1.1	0.885	0.945	0.885	0.934	0.968	0.934
Case 1.2				0.883	0.925	0.883
Case 2.1	0.915	0.966	0.915	0.982	0.7	0.894
Case 2.2	0.966	0.888	0.894			
Case 3.1	0.946	0.719	0.894	0.979	0.832	0.894
Case 3.2	0.927	0.974	0.927			
Case 4.1	0.907	0.460	0.894	0.886	0.543	0.886
Case 4.2	0.897	0.949	0.897			
Case 5.1	0.970	0.994	0.970	0.971	0.977	0.971
Case 5.2	0.904	0.963	0.904			
Case 6.1	0.951	0.754	0.894	0.965	0.973	0.965
Case 6.2	0.915	0.846	0.894			
Case 7.1	0.971	0.990	0.971	0.958	0.992	0.958
Case 7.2	0.782	0.728	0.782			
Case 8.1	0.958	0.000	0.894	0.993	1.000	0.993
Case 8.2	0.926	0.951	0.926			
Case 9.1	0.983	0.939	0.939	0.982	0.999	0.982
Case 9.2	0.937	0.912	0.912			

Table 6: Obtained results for the proposed methods and their combinations for all the positive cases in eyeball center and orbit analysis.

	M'1 Right	M'2 Right	M'1 ⊕ M'2 Right	M'1 Left	M'2 Left	M'1 ⊕ M'2 Left
Case 1.1	0.800 (4)	0.933 (2)	0.800 (4)	1.000 (1)	1.000 (1)	1.000 (1)
Case 1.2				0.923 (2)	1.000 (1)	0.923 (2)
Case 2.1	0.500 (8)	0.857 (3)	0.857 (3)	1.000 (1)	0.875 (3)	1.000 (1)
Case 2.2	0.857 (3)	0.643 (6)	0.714 (5)			
Case 3.1	1.000 (1)	1.000 (1)	1.000 (1)	1.000 (1)	1.000 (1)	1.000 (1)
Case 3.2	1.000 (1)	1.000 (1)	1.000 (1)			
Case 4.1	1.000 (1)	0.714 (5)	1.000 (1)	1.000 (1)	0.875 (3)	1.000 (1)
Case 4.2	1.000 (1)	1.000 (1)	1.000 (1)			
Case 5.1	0.941 (2)	1.000 (1)	0.941 (2)	0.974 (2)	0.974 (2)	1.000 (1)
Case 5.2	0.588 (8)	0.941 (2)	0.882 (3)			
Case 6.1	0.941 (2)	0.588 (8)	0.882 (3)	1.000 (1)	0.842 (4)	1.000 (1)
Case 6.2	0.824 (4)	0.706 (6)	0.882 (3)			
Case 7.1	1.000 (1)	1.000 (1)	1.000 (1)	1.000 (1)	0.941 (2)	1.000 (1)
Case 7.2	0.294 (13)	0.667 (6)	0.294 (13)			
Case 8.1	0.882 (3)	0.000 (12)	0.882 (3)	1.000 (1)	1.000 (1)	1.000 (1)
Case 8.2	0.824 (4)	0.909 (2)	1.000 (1)			
Case 9.1	1.000 (1)	0.941 (2)	0.941 (2)	1.000 (1)	1.000 (1)	1.000 (1)
Case 9.2	0.824 (4)	0.882 (3)	0.882 (3)			

Table 7: Accuracy index and ranking position (in brackets) for the proposed methods and their combinations for all the positive cases in eyeball center and orbit analysis.

6. Conclusions and Future Works

In this work, we have introduced a new methodology to model spatial and morphological relations between skull and face to support the forensic anthropologist in the identification by CFS task. On the one hand, we have proposed eight different methods to measure the consistency between bony and facial

Methods	Parameter λ	Accuracy index
Aggregation method (M'1)		0.894
Centroid method (M'2)		0.861
M'1 $\oplus$ M'2	-0.981014	**0.922**

Table 8: Accuracy degrees obtained for the two proposed methods to model the positional relationship between the orbit and the eyeball center and their combinations computed by Sugeno Integral.

Criterion	Mean execution time
Chin	2.147 sec
Eyeballs	12.332 sec

Table 9: Mean run time of the best combination of methods for each criterion over one SFO.

chin shapes. On the other hand, we have implemented two different approaches to model the positional relationship between the orbit and the eyeball center. We have also proposed the methods to segment the proper contours in both cases. All these implementations are based on CV techniques. We have studied the behavior of these different methods separately and we have obtained the accuracy index of each of them using a dataset of positive and negative identification cases. Then, we have analyzed the performance of their combinations, obtaining the accuracy index that allow us to aggregate their outputs using Sugeno integral. It has been proven how four of the methods proposed clearly outperformed the state-of-the-art methods.

It has been shown that the Sugeno integral is a perfect tool to these kinds of applications. In the case at hand, we have achieved a more precise method. This proposed DSS, based on the use of fuzzy integral, together with our previous contributions for the first and second stages based on evolutionary algorithms and fuzzy sets, make up a complete hybrid intelligent system in the automation of the CFS task.

In future works, we aim to complete the DSS process by modeling the other two families of criteria for assessing the craniofacial correspondence. The next step will be the integration of multiple degrees of support, each of them indicating the consistency between the skull and face guided by a particular criterion in different facial regions to obtain a unique degree for a SFO. Finally, other assessment methodologies including the quality and quantity of the used materials will also be studied for this goal.

Acknowledgments

References

[1] C. Cattaneo, Forensic anthropology: developments of a classical discipline in the new millennium, Forensic Sci Int 165 (2) (2007) 185–193.

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

[3] C. Wilkinson, C. Rynn, Craniofacial identification, Cambridge University Press, 2012.

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

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

[6] P. T. Jayaprakash, Conceptual transitions in methods of skull-photo superimposition that impact the reliability of identification: A review, Forensic Sci Int 246 (2015) 110–121.

[7] S. Damas, C. Wilkinson, T. Kahana, E. Veselovskaya, A. Abramov, R. Jankauskas, P. Jayaprakash, E. Ruiz, F. Navarro, M. Huete, E. Cunha, F. Cavalli, J. Clement, P. Leston, F. Molinero, T. Briers, F. Viegas, K. Imaizumi, D. Humpire, O. Ibáñez, Study on the performance of different craniofacial superimposition approaches (ii): best practices proposal, Forensic Sci Int. In Press, doi:10.1016/j.forsciint.2015.07.045.

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

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] J. Huang, M. Zhou, F. Duan, Q. Deng, Z. Wu, Y. Tian, The weighted landmark-based algorithm for skull identification, in: Computer Analysis of Images and Patterns, Springer, 2011, pp. 42–48.

[14] W. Jin, G. Geng, K. Li, Y. Han, Parameter estimation for perspective projection based on camera calibration in skull-face overlay, in: Virtual Reality and Visualization (ICVRV), 2013 International Conference on, IEEE, 2013, pp. 317–320.

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

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

[17] 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 (1) (2015) 227–236.

[18] 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.

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

[20] J. M. Keller, P. Gader, H. Tahani, J.-H. Chiang, M. Mohamed, Advances in fuzzy integration for pattern recognition, Fuzzy Set Syst 65 (2) (1994) 273–283.

[21] M. Sugeno, Theory of fuzzy integrals and its applications, Tokyo Institute of Technology, 1974.

[22] S. Aja-Fernández, R. de Luis-Garcıa, M. A. Martın-Fernández, C. Alberola-López, A computational $TW3$ classifier for skeletal maturity assessment. A computing with words approach, J Biomed Inform 37 (2) (2004) 99–107.

[23] J. Santamaría, O. Cordón, S. Damas, J. M. García-Torres, A. Quirin, Performance evaluation of memetic approaches in 3D reconstruction of forensic objects, Soft Comput 13 (2009) 883–904.

[24] D. Navega, R. Vicente, D. N. Vieira, A. H. Ross, E. Cunha, Sex estimation from the tarsal bones in a portuguese sample: a machine learning approach, Int journal of legal medicine (2014) 1–9.

[25] D. D. Angelis, R. Sala, A. Cantatore, M. Grandi, C. Cattaneo, A new computer-assisted technique to aid personal identification, Int J Legal Med 123 (4) (2009) 351–356.

[26] B. R. Campomanes-Álvarez, O. Cordón, S. Damas, Evolutionary multi-objective optimization for mesh simplification of $3d$ open models, Integr Comput-Aid E 20 (4) (2013) 375–390.

[27] P. Karczmarek, W. Pedrycz, M. Reformat, E. Akhoundi, A study in facial regions saliency: a fuzzy measure approach, Soft Comput 18 (2) (2014) 379–391.

[28] M. F. Anderson, D. T. Anderson, D. J. Wescott, Estimation of adult skeletal age-at-death using the sugeno fuzzy integral, Am J Phys Anthropol 142 (1) (2010) 30–41.

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

[30] 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.

[31] 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.

[32] A. Goshtasby, 2D and 3D Image Registration, Wiley Interscience, 2005.

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

[34] 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.e1–6.

[35] P. Jayaprakash, G. Srinivasan, M. Amravaneswaran, Cranio-facial morphanalysis: a new method for enhancing reliability while identifying skulls by photo superimposition, Forensic Sci Int 117 (1) (2001) 121–143.

[36] C. Campomanes-Alvarez, O. Ibanez, O. Cordon, Modeling the consistency between the bony and facial chin outline in craniofacial superimposition, in: 16th World Congress of the International Fuzzy Systems Association (IFSA), 2015, pp. 1612–19.

[37] D. V. Pesce, E. Vacca, F. Potente, T. Lettini, M. Colonna, Shape analytical morphometry in computer-aided skull identification via video superimposition, Iscan MY, Helmer RP. Forensic analysis of the skull: craniofacial analysis, reconstruction, and identification. New York: Wiley-Liss, 1993.

[38] M. Yoshino, H. Matsuda, S. Kubota, K. Imaizumi, S. Miyasaka, S. Seta, Computer-assisted skull identification system using video superimposition, Forensic Sci Int 90 (3) (1997) 231–244.

[39] A. Ricci, G. L. Marella, M. A. Apostol, A new experimental approach to computer-aided face/skull identification in

forensic anthropology, Am J Foren Med Path 27 (1) (2006) 46–49.

[40] C. M. Takemura, R. Cesar, I. Bloch, Fuzzy modeling and evaluation of the spatial relation "Along", in: Progress in Pattern Recognition, Image Analysis and Applications, Springer, 2005, pp. 837–848.

[41] I. Bloch, O. Colliot, R. M. Cesar, On the ternary spatial relation "between", IEEE T Syst Man Cy B 36 (2) (2006) 312–327.

[42] M. Yang, K. Kpalma, J. Ronsin, et al., A survey of shape feature extraction techniques, Pattern Recogn (2008) 43–90.

[43] R. C. Veltkamp, Shape matching: Similarity measures and algorithms, in: Shape Modeling and Applications, SMI 2001 International Conference on., IEEE, 2001, pp. 188–197.

[44] C. P. Pappis, N. I. Karacapilidis, A comparative assessment of measures of similarity of fuzzy values, Fuzzy Set Syst 56 (2) (1993) 171–174.

[45] D. Zhang, G. Lu, Review of shape representation and description techniques, Pattern Recogn 37 (1) (2004) 1–19.

[46] H. Freeman, On the encoding of arbitrary geometric configurations, Electronic Computers, IRE Transactions on (2) (1961) 260–268.

[47] I. Bloch, A. Ralescu, Directional relative position between objects in image processing: a comparison between fuzzy approaches, Pattern Recogn 36 (7) (2003) 1563–1582.

[48] J. M. Keller, X. Wang, Comparison of spatial relation definitions in computer vision, in: Uncertainty Modeling and Analysis, 1995, and Annual Conference of the North American Fuzzy Information Processing Society. Proceedings of ISUMA-NAFIPS'95., Third International Symposium on, IEEE, 1995, pp. 679–684.

[49] C. M. Wilkinson, S. A. Mautner, Measurement of eyeball protrusion and its application in facial reconstruction, J Forensic Sci 48 (1) (2003) 12–16.

[50] 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 (3) (2015) 569–81.

[51] V. Torra, Y. Narukawa, The interpretation of fuzzy integrals and their application to fuzzy systems, Int J Approx Reason 41 (1) (2006) 43–58.

[52] H. Imai, V. Torra, On a modeling of decision making with a twofold integral., in: EUSFLAT Conf., 2003, pp. 714–717.

[53] M. Sugeno, Fuzzy measures and fuzzy integrals: a survey, Fuzzy automata and decision processes 78 (33) (1977) 89–102.

Carmen Campomanes-Alvarez received her M.S. degree (2011) in Telecommunication Engineering from the University of Oviedo (Oviedo, Spain). In August 2012, she received the M.S. degree in Information Technologies and Communications in Mobile Networks from the same university. Since April 2013 she has a research grant by the Spanish Ministry of Science, Culture and Sport to perform her PhD studies. She is currently working toward the Ph.D. degree in Information and Communications Technologies with the University of Granada (Granada, Spain). Her main research interests are focused on forensic identification using Soft Computing.

Oscar Ibáñez received the Ph.D. degree in Computer Science from the University of Santiago de Compostela, Spain, in 2010. That work led to receive the IFSA Award for Outstanding Applications of Fuzzy Technology and the EUSFLAT Best Ph.D. Thesis Award. It is also the core part of a patent and the corresponding commercial software developed Face2Skull by the European Centre for Soft Computing and the University of Granada. He was working at the former for six years and he is currently a Postdoc researcher at the latter institution (*Juan de la Cierva* fellowship). His main research interests include evolutionary computation, fuzzy logic, computer vision, machine learning, and their applications to tackle real world problems with a particular interest in forensic identification.

Oscar Cordón received his M.S. degree (1994) and his Ph.D. (1997) both in Computer Science from the University of Granada, Spain, where he is now Full Professor at the Dept. of Computer Science and Artificial Intelligence. He was the founder and leader of this University's Virtual Learning Center between 2001 and 2005, and was awarded with the Young Researcher Career Award in 2004. From 2006 to 2011 he was one of the founding researchers of the European Centre for Soft Computing, Mieres, Spain, in his role of Principal Researcher, and he is still contracted as Distinguished Affiliated Researcher. In 2014, he was awarded with the Spanish National Award on Computer Science ARITMEL by the Spanish Computer Science Scientific Society. He is Associate Editor of 11 international journals and a reviewer for more than 40. His current main research interests are in the fields of soft computing for forensic anthropology and medical imaging, fuzzy rule-based systems and genetic fuzzy systems, evolutionary computation, ant colony optimization and other single and multi-objective metaheuristics, and their applications to different real-world problems.

ACCEPTANCE AND RESIGNATION OF THE PUBLICATION'S CO-AUTHORS

Publication/article: ***DESIGN OF CRITERIA TO ASSESS CRANIOFACIAL CORRESPONDENCE IN FORENSIC IDENTIFICATION BASED ON COMPUTER VISION AND FUZZY INTEGRALS***

The co-authors:

Mr./Ms.	Oscar Ibáñez
Mr./Ms.	Oscar Cordón
Mr./Ms.	
Mr./Ms.	
Mr./Ms.	
Mr./Ms.	

Declare that they:

☒ Accept and authorize the use of the above mentioned publication/article as part of the documentation for the deposit and defence of the doctoral thesis of Mr./Ms. **Carmen Campomanes Álvarez** titled ***Automation of the assessment of craniofacial superimposition using soft computing and computer vision***,

☒ Have not used the above mentioned publication/article as part of the documentation for the deposit and defence of another doctoral thesis and/or refuse to use it for a future doctoral thesis

Granada, 2 , mayo 2017

Signed: Oscar Ibáñez   Signed: Oscar Cordón   Signed:

4 Hierarchical Information Fusion for Decision Making in Craniofacial Superimposition

- C. Campomanes-Álvarez, O. Ibáñez, O. Cordón and C. Wilkinson. Hierarchical Information Fusion for Decision Making in Craniofacial Superimposition, Information Fusion. DOI: http://dx.doi.org/10.1016/j.inffus.2017.03.004.

  - State: Accepted manuscript, in press.
  - Impact Factor (JCR 2015): 4.353.
  - Category: COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE. Order: 9/130. Q1.

Hierarchical Information Fusion for Decision Making in Craniofacial Superimposition

Carmen Campomanes-Alvarez[a,*], Oscar Ibáñez[a,*], Oscar Cordón[a,*], Caroline Wilkinson[b,*]

[a]*Department of Computer Science and Artificial Intelligence, University of Granada,*
C/ Daniel Saucedo Aranda, s/n, 18071 Granada, Granada, Spain.
[b]*School of Art and Design, Liverpool John Moores University, Liverpool L3 5TF, UK*

Abstract

Craniofacial superimposition is one of the most important skeleton-based identification methods. The process studies the possible correspondence between a found skull and a candidate (missing person) through the superimposition of the former over a variable number of images of the face of the latter. Within craniofacial superimposition we identified three different stages, namely: 1) image acquisition-processing and landmark location; 2) skull-face overlay; and 3) decision making. While we have already proposed and validated an automatic skull-face overlay technique in previous works, the final identification stage, decision making, is still performed manually by the expert. This consists of the determination of the degree of support for the assertion that the skull and the ante-mortem image belong to the same person. This decision is made through the analysis of several criteria assessing the skull-face anatomical correspondence based on the resulting skull-face overlay. In this contribution, we present a hierarchical framework for information fusion to support the anthropologist expert in the decision making stage. The main goal is the automation of this stage based on the use of several skull-face anatomical criteria combined at different levels by means of fuzzy aggregation functions. We have implemented two different experiments for our framework. The first aims to obtain the most suitable aggregation functions for the system and the second validates the proposed framework as an identification system. We tested the framework with a dataset of 33 positive and 411 negative identification instances. The present proposal is the first automatic craniofacial superimposition decision support system evaluated in an objective and statistically meaningful way.

Keywords: forensic anthropology, craniofacial superimposition, decision making, information fusion, fuzzy aggregation operators, computer vision.

1. Introduction

Craniofacial superimposition (CFS) [1] is the most representative technique within craniofacial identification [2]. It involves superimposing a skull onto a one or more ante-mortem (AM) photographs of a missing person. The consequent analysis of their morphological correspondence determines if they belong to the same subject.

The whole CFS process can be divided into three consecutive stages [3] (Fig. 1): 1) The acquisition and processing of the materials, i.e, skull and AM facial images, and the location of somatometric landmarks on both; 2) Skull-face overlay (SFO), which deals with accomplishing the best possible superimposition of the skull and a single AM photograph of a missing person. This procedure is iteratively executed for each photograph, thus getting different overlays. 3) Decision making process aims to determine the degree of support for a match based on the SFOs achieved in the previous step. The final decision is managed by

---

different criteria based on the anatomical relationship between the face and the skull. These criteria can vary depending on the region and the pose [4].

Designing automatic methods to address CFS and support the forensic anthropologist remains a challenge and dreamed milestone within the anthropology community. In fact, the development of computer-aided CFS methods has increased over the past twenty years [5]. Recent approaches use skull 3D models and soft computing (SC) methods for the first two identification stages. These methods allow us to both automate the task and handle the inherent uncertainty [6, 7, 8, 9].
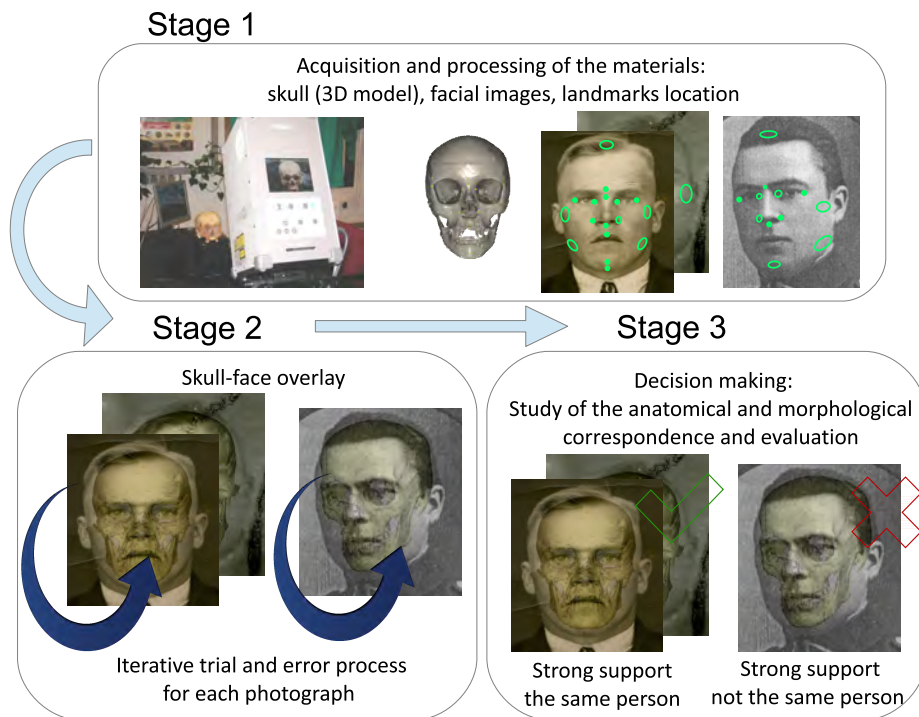
Figure 1: CFS procedure scheme.

In the third stage, once one or more SFOs are obtained, experts evaluate morphological and spatial skull-face relations, focusing on certain regions demonstrated to be more discriminative. The final decision is provided from an aggregation of the partial decisions. It is taken in terms of limited, moderate or strong support to the assumption that the skull and the facial image belong to the same individual or not [4]. This process is subjective and it relies on the skills of the forensic expert while influenced by the quantity and quality of the used materials. Therefore, there is a need to design a decision support system (DSS) to help practitioners to make their decision based on the fusion of the available information sources in a faster and more objective way. It would also lead to the application of CFS in identification scenarios with multiple comparisons, a possibility not explored yet due to the unaffordable time lapse needed to analyse all possible cross-comparisons. Our long-term and complex goal is the implementation of a DSS by evaluating the spatial and morphological relationships. The system will return a numeric index as output, in order to support the forensic anthropologist to make the final CFS decision while automatically filtering a number of cases in multiple comparison scenarios.

In previous works [10, 11, 12], we presented a first preliminary proposal to design a DSS for CFS using computer vision (CV) and fuzzy integrals. We implemented two of the most discriminative criteria to assess craniofacial correspondence, namely, the spatial and the morphological relation between the bony and facial chin, and the relative position of the orbits and the eyeballs.

In this work we present a complete hierarchical DSS for CFS with three connected levels of decision. The previous studies only tackled what we currently identified as the third and simplest level. We only

implemented some CV methods aimed to measure two criteria to assess craniofacial correspondence in the corresponding two isolated regions. Thus, previous developments cannot be used for the identification task. Here, for the first time, we propose a complete framework that allows forensic experts to automatically address the final decision making stage. The presented fuzzy DSS develops information fusion concerning skull-face anatomical correspondence at different levels: criterion evaluation, SFO evaluation, and CFS evaluation. Additionally, in this study, we provide an implementation of the SFO evaluation level of the DSS (as explained above, we have already provided an implementation for the criterion evaluation level [10]). Within this level, we distinguish three sublevels with different conditions of aggregation. In each of them, the different sources of uncertainty are modeled, and different aggregation mechanisms account for information fusion and propagation. These sources of uncertainty also provide a mechanism to propagate information and uncertainty from criterion evaluation to SFO evaluation levels.

The uncertainty sources and degrees of confidence involved in the information fusion process are classified into bone, image, SFOs, morphological aspects, and computational methods used to model the criteria. The bone uncertainty refers to the quality of the skull, and the uncertainty of the photograph considers the visibility of each region and the resolution of the image. Morphological aspects can vary the degree of confidence of a criterion, depending on the sex, age, body mass index (BMI), or ancestry. Finally, the accuracy of the used methods is also taken into account, as well as the quality of the SFO achieved in the previous step.

We perform an experiment with positive and negative identification cases. In total, we analyze 33 positive SFOs against 411 negatives. We test 24 different combinations of aggregation functions within the proposed fuzzy DSS. We both analyze the results studying the mean accuracy of each approach and its capability of identification.

This manuscript is organized as follows. In Section 2 we introduce the relevant previous work and state the characteristics of the problem we aim to tackle. Section 3 describes our proposed DSS and Section 4 the corresponding implementation. Section 5 shows the experiments, and Section 6 details the discussion and conclusions.

2. Background

2.1. Craniofacial Superimposition

CFS approaches evolved as new technology was available although their foundations were laid more than 100 years ago [13, 14]. Three families of approaches have been developed along this time: photo superimposition (appeared in the mid 1930s), video superimposition (widely common since 1975), and computer-aided superimposition (developed in the second half of the 1980s) [1, 15].

Computerized systems in CFS are very transcendent [3, 5]. Some publications [6, 7, 8, 9, 16, 17] serve as examples of how computer techniques, specially CV and SC, can automate SFO and tackle the uncertainty/fuzziness of several cephalometric landmarks [18] and of the soft tissue distances [19]. These proposals are based either on photograph to photograph comparison [20] or on skull 3D model to photograph comparison [6, 7, 8, 15, 17]. Computerized CFS methods play an important role since they have managed to reduce time and subjectivity inherent to manual approaches followed by forensic experts. However, the resulting overlays' quality is influenced by several sources of uncertainty, as well as by partial and incomplete knowledge about skull-face anatomical correspondence. Accordingly, it is very difficult to achieve an optimal accuracy in an automatic way and a later manual refinement of SFO results is currently needed for such a purpose. Related to this, it is important to note that the forensic expert selects the materials (i.e. skull and photographs) in a previous phase of the process. This is a forensic technique frequently employed in multiple labs around the world, so the expert filters the materials and if they are not reliable, the SFO cannot be carried out. Best practices to follow by forensic experts in CFS were recently agreed with the framework of the EU project MEPROCS and they are described in [4].

Our system consists of the automation of the CFS process based on SC and CV techniques. In order to automate the SFO stage, we attempt to replicate the original scene in which the photograph was taken [21]. From the CV perspective, this involves a 3D-2D image registration problem. This process is guided

by incomplete and vague information (matching of two different objects, face and skull), and it involves an optimization task within a challenging search space with many local minima to establish the parameters of the registration transformation. For these reasons, advanced SC techniques have been designed to face this complex optimization problem [6, 7, 8, 9]. The resulting automatic overlays generated by our system are the inputs to the CFS DSS proposed in this paper.

2.2. Decision Making in Craniofacial Superimposition

Once one or several skull-face overlays have been achieved for the same identification case[1], the main goal is to determine the degree of support for the assertion that the skull and the face of the photograph(s) belong to the same person or not. This degree of support is based on the consistency of the matching between the face and the skull but it is also influenced by the quality and quantity of the materials used (photographs and skull). A scale for a craniofacial matching evaluation has been recently defined by some of the most representative experts in craniofacial identification within the MEPROCS project framework [4]. Accordingly, the final decision is provided in terms of strong, moderate or limited support.

This decision is guided by different criteria studying the anatomical relationship between the skull and the face. According to the literature [23], we can distinguish the following families of criteria for assessing the craniofacial correspondence:

1. Analysis of the consistency of the bony and facial outlines/morphological curves. Forensic experts confirm if two particular curves (of skull and face) are anatomically consistent. That is, if two curves follow the same shape or, in other words, if one curve mirrors the other. An example of this criteria can see it in Fig. 2.a where the forehead curve of the face follows the forehead curve of the skull.

2. Assessment of the anatomical consistency by positional relationship. These criteria consist of a positional relationship analysis in order to assess anatomical consistency. Thus, the goal is to check if the relative position of a skull region against a facial region is similar in respect to anatomical reference. Fig. 2.b shows a case of this family: the consistency between the lateral angle of the eye and the cranial orbit.

3. Location and comparison lines to analyze anatomical consistency. Experts analyze a set of marking lines (obtained by joining some reference landmarks) on the face and on the skull. In terms of CV, these lines have to be parallel in an image. For example, the ectocanthion lines marked in the skull and the face (Fig. 2.c).

4. Evaluation of the consistency of the soft tissue thickness between corresponding cranial and facial landmarks. The last set of criteria consists of analyzing the consistency of the facial soft tissue thickness considering distances between pairs of homologous landmarks (located on the skull and the face). Fig. 2.d shows an example of how the facial landmarks positions can be estimated from cranial landmarks using cones to model the soft tissue thickness. These distances can be checked using the skull-face overlay in existing studies relating to soft tissue thickness in different human populations [19].

MEPROCS work group also discussed and quantitatively analyzed these criteria for the evaluation of the morphological skull-face correspondence, providing a set of the most discriminative and easy to assess criteria [24].

Our long term goal is to automate the whole decision making process by modeling the most relevant criteria within the previous four families using CV and SC techniques. The resulting system would give as a result a global degree of support of a CFS identification to assist the forensic anthropologist to make her/his decision.

---

[1]The number of SFOs to be used for the identification depends on the available valid photos. The reliability of the technique relies on having more than one photo, at different poses, etc. [22]
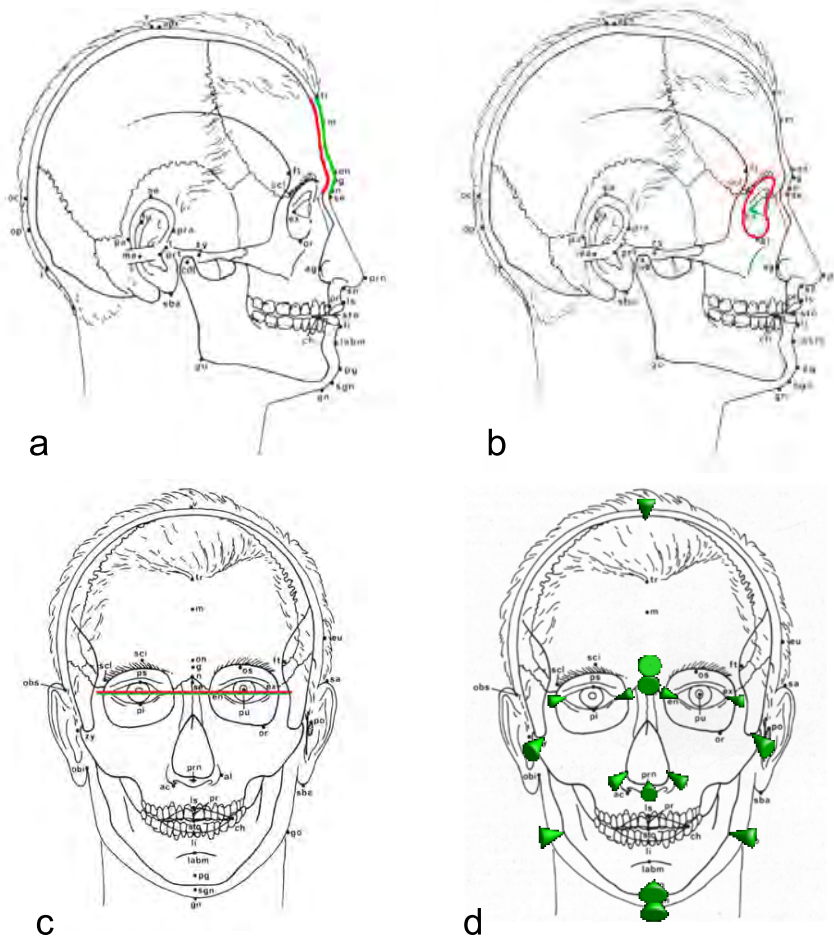
Figure 2: Examples of the four families of criteria for assessing the craniofacial correspondence. a) Consistency between the outline of the frontal bone and the forehead outline in lateral view; b) Consistency between the lateral angle of the eye and the orbit; c) Ectocanthion lines marked in the skull and the face; d) Facial landmark position from a cranial landmark using cones to model the soft tissue thickness.

There are just a few works tackling the automation of the analysis of craniofacial correspondences within the framework of CFS identification [25, 26, 27]. Most of the existing literature was published more than 18 years ago and the works are very basic and limited. In addition, they do not consider the use of either skull 3D models or computer techniques to perform the skull-face overlay. Besides, the employed technique for the shape analysis implies manual interaction. They provide a value that does not take into account the actual spatial relation between skull and face since the methods employed are invariant to translation, scale and rotation. Finally, these systems only implement a single group of the criteria to assess the craniofacial correspondence. For further information see [10].

Recently, we presented a simple and preliminary version of the DSS in [10, 11, 12]. In these works, we considered two of the most discriminative criteria to assess craniofacial correspondence: the morphological and spatial relationship between the facial and bony chin and the relative position of the eyeballs and the orbits. We developed several CV-based methods to assess the degree of matching of each of these two criteria, and aggregated their results in a single value (using different aggregation functions) to obtain more robust and accurate results (see Fig. 3).

To model the former criterion, we implemented several CV methods aimed to measure how the chin facial shape follows the skull shape given the delineation of these regions in a SFO. Our method includes

the automatic segmentation of the contours based on the region "between" them. Thanks to this process, this analysis takes into account the relative position and the distance between the two objects as well as the shapes. In [10], we proved that the best performance to model the relation "a curve follows another curve" is obtained using shape similarity measures with the area function comparison method and the complex coordinates' signature. In order to measure this similarity the Euclidean distance of their corresponding shape signatures is computed. The following formula gives a similarity value between zero and one:

$$S_4 = 1 - \sqrt{\frac{1}{N} \sum_{i=1}^{N} (s_F(i) - s_B(i))^2} \tag{1}$$

with s_F and s_B being the shape signature of each contour.

Similarly, we developed a method to measure the relative position between the orbit and the center of the eyeball. For this aim, we implemented two different ways to compute the positional relationship between two objects in an image: the aggregation method and the centroid method. Once the relative position is obtained, we need to compare this position with a reference model in order to assess if there is anatomical consistency. In order to compare two different relative positions between two objects, we compute the similarity using the following expression based on [28]:

$$S_6 = 1 - \frac{|\delta_a - \delta_a'| + |\delta_b - \delta_b'| + |\delta_r - \delta_r'| + |\delta_l - \delta_l'|}{\delta_a + \delta_a' + \delta_b + \delta_b' + \delta_r + \delta_r' + \delta_l + \delta_l'} \tag{2}$$

where δ_a, δ_b, δ_r and δ_l are the degrees of the position relation "above", "below", "right", and "left", respectively, and δ and δ' are the relative position of two different pairs of objects.

To evaluate this criterion, we need to extract the appropriate contours both for the orbits (bony region) and eyeball center (facial region). The desired orbital contour is the interior contour, selected as the smaller one, while the facial region is simply the contour of the marked zone on the photograph.
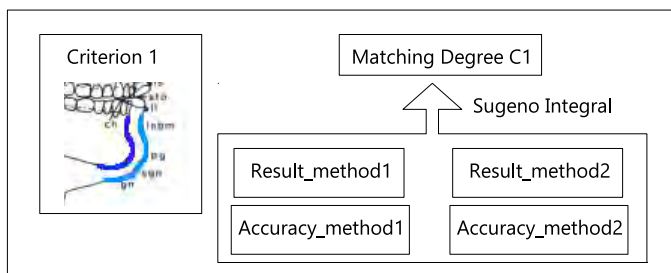

Figure 3: Scheme of the matching degree calculation for a criterion presented in [10].

All of these performed methods return a value between zero and one, which indicates how well the relation is achieved. Then, we aggregate the degrees of support of the individual methods to strengthen the final result. To do this, it is needed a measurement of importance of each one (the accuracy). We review the calculation of this value as follows since it is employed along the current proposal described in Sections 3 and 4. Regardless the criterion type, this accuracy of each method is calculated as its capability to discriminate in the decision making process (ranking positive and identification negative cases). Hence, a database composed of real positive identification cases (including the 3D skull model and one or more subject photographs) and negative cases is required. These negatives cases are obtained combining real 3D skull models and non corresponding photos of similar subjects (same gender, age, ethnic group, ...). First, the corresponding value to apply each method for a specific criterion over the database of cases is achieved. Next, the matching values reported are used to rank the candidates based on their chance to be the actual subject. Then, a value between 0 and 1 is assigned to each positive case taking into account this ranking: if the method reported the highest value to a positive case (first position of the ranking), 1 is assigned. On

the contrary, if the position of a positive case is the last of the ranking, 0 is assigned. The formula to assign the accuracy of the method x_i in the instance j is:

$$Acc(x_i)_j = 1 - \frac{r-1}{M_j - 1} \tag{3}$$

where r is the position of the positive case in the ranking and M_j is the worst (highest) value of the ranking for the instance j (all cases getting the same criterion-method value are supposed to have a draw, that is, they are assigned to the same ranking).

Then, the average of all these accuracy values over all cases is calculated. As a result, an accuracy index in the interval $[0, 1]$ is achieved for each method. The final step is to aggregate the degrees of support of the best individual methods taking into account their accuracy. To do that, Sugeno integral [29] is used. For the two cases of study, the accuracy index of the aggregation overcame the individual results. The scheme of this procedure is shown in Fig. 3. In [12] we performed a deeper study of the behavior of different aggregation functions for the same aim. The obtained results show that Sugeno integral ranks better than the Choquet Integral and the Weighted Arithmetic Mean although no significant conclusions can be delivered regarding the performance.

3. Hierarchical Decision Support Framework for Craniofacial Superimposition

The whole CFS process is affected by several uncertainty sources and degrees of confidence that must be considered for decision making. In particular, we have distinguished the following sources of uncertainty from the forensic experts' experience.

Bone quality: the quality of the skeletal remains is an important issue during the CFS process. The condition of the bones depends on environmental factors, its preservation state has a direct influence on the confidence on the evaluation of face and skull anatomical correspondence.

Image quality: photographic quality is an additional criterion that has to be taken into consideration. The uncertainty inherent to the location of landmarks and regions in an image, already described in [4], can be greatly affected by the quality of the image. In particular, the location and evaluation of each single region/landmark is affected by the following sources of imprecision: 1) the variation in the distribution of shadows which depend on the light; 2) unsuitable focus, especially when the plane of focus is not enough depth and hence the critical objects are not sharp; 3) the image resolution. For optimal examination, experts recommend using photographs in which the facial image resolution is at least 180 pixels corresponding to the width of the head, or 90 pixels between the eyeballs (for full frontal images); 4) the pose of the face in the image, i.e. angle of view (frontal, lateral or oblique) and facial expression; 5) complete or partial occlusion of a region due to the presence of elements such as glasses, clothes or hair [4].

Skull-face overlay accuracy: the confidence degree of the SFO accomplished in the previous stage is another important factor to be considered. This process focuses on achieving the best possible superimposition of the skull and a facial image and it can be influenced by different sources of uncertainty, as described in [7].

Morphological aspects: the degree of confidence of each particular criterion analyzing anatomical correspondence can be affected by several factors. Firstly, the expected craniofacial relationship of a specific region is affected (in a lower or higher degree) by the age, sex, ancestry (biological profile), and/or BMI [30]. Thus, these factors have to be considered for each particular criterion during the decision making process. For example, the chin shape of the skull follows the facial shape in young people, but after 45 years of age this relation is increasingly unreliable. This relation can also be distorted due to being overweight. Secondly, each isolated region can have a different discriminative identification power. This is considered as the rate of being able to make a positive identification taking into account only that region.

Automatic method modeling the spatial/morphological relationship accuracy: different computer methods can be considered to automatically evaluate the degree of matching for each specific criterion. For example, the chin correspondence can be evaluated by considering different shape extraction and matching methods (see Sec. 2.2). The accuracy of a method is defined as how well the specific craniofacial relationship is modeled by that method for an actual criterion.

7

The DSS proposed in this work considers the evaluation of the skull-face anatomical correspondence at different levels. In each of them, the various sources of uncertainty introduced above are modeled, and different aggregation mechanisms account for information fusion and propagation. In particular, we have defined the following three levels or decision hierarchies (see Fig. 4):

- Level 1: CFS evaluation

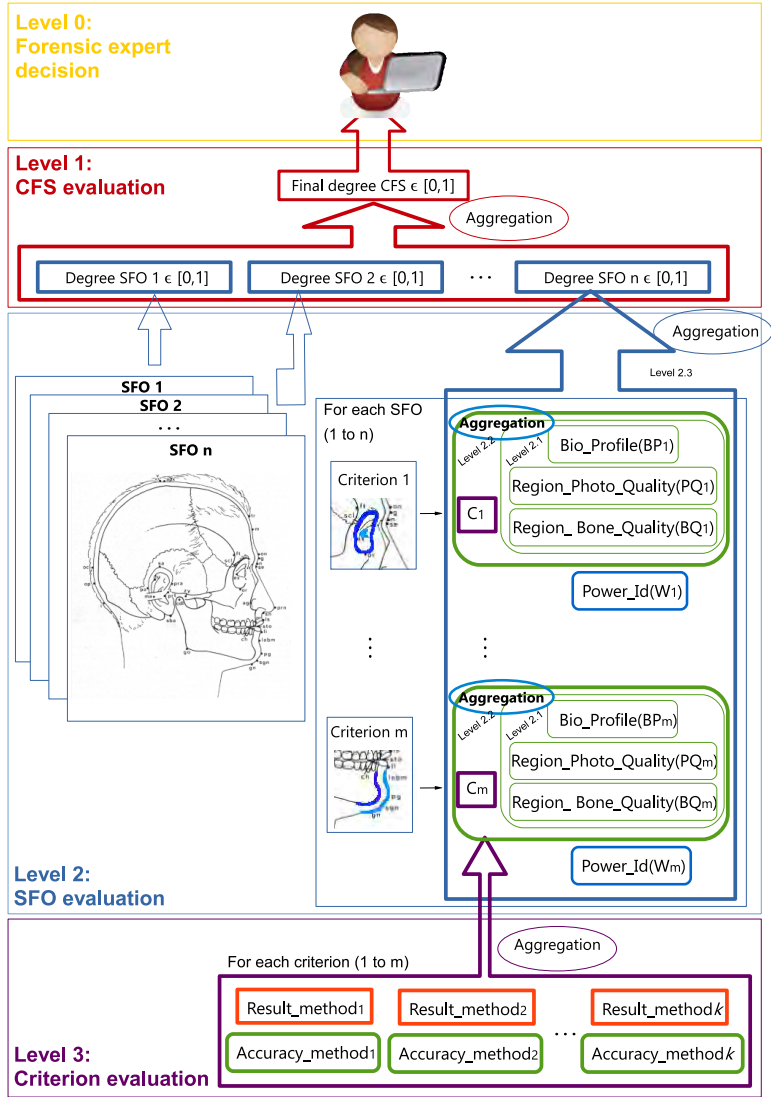- Level 2: SFO evaluation

- Level 3: Criterion evaluation (introduced in [10, 11, 12], see Sec. 2.2)

Figure 4: Hierarchical scheme of the DSS for CFS.

In this DSS scheme, the final degree of the CFS identification is obtained by aggregating all the SFO degrees. This corresponds with the highest level (**Level 1**) in the hierarchical system. In the next level (**Level 2**), for each SFO achieved in the previous stage, we aim to analyze the degree of fulfillment of

8

several criteria studying skull-face anatomical correspondence. For each criterion, facial and cranial regions are located (in the facial photograph and the 3D model, respectively) and a specific method to evaluate the skull-face correspondence is applied (the kind(s) of method(s) to be considered depends on the nature of the specific criterion). The degree of craniofacial correspondence of a SFO is computed by aggregating the matching degree of each single criterion taking into consideration the confidence of that criterion. Thus, the skull-face consistency in a region is expressed by a value between 0 and 1, obtained in the previous level (C_m). This value is complemented by the region/criterion confidence based on the quality of the photo (PQ_m), the quality of the bone (BQ_m), the biological profile variability of the criterion (BP_m), and the discriminative power of the isolated region. At this level, we set three different aggregation sublevels. The first one consists of aggregating the first three sources of uncertainty to get a single uncertainty value associated to the criterion sample quality and biological variability. The second sublevel integrates this aggregation with the matching degree (C_m). Finally, at the third one, we obtain the degree of the SFO craniofacial correspondence by aggregating the different previous values for all the regions taking into account the discriminative power of the isolated region as weight. We denote these sublevels as level 2.1, level 2.2, and level 2.3, respectively. If there are more than one CV-based method to evaluate a criterion, there is a need to aggregate the results of all of them (**Level 3**). To do so, we take into account the accuracy of each method. Fig. 4 graphically summarizes the proposed hierarchical DSS for CFS.

In addition, but not included within our DSS, there is a Level 0 of evaluation, the one carried out by the human expert. At this higher level the input is the final degree of CFS matching provided by the DSS. Then, the forensic experts will make the final identification decision. In this sense, the MEPROCS international consortium agreed a set of possible decision degrees according to the quality and quantity of materials [4]. The matching degree provided by our DSS could be directly incorporated within this scale, to any other, or considered in its own.

4. Framework Implementation

Together with the hierarchical DSS framework proposal, in this contribution we focus on the design of the SFO evaluation level. In this level 2 we aggregate the degree of the craniofacial matching in a region with the associated criterion uncertainty (see Fig. 5).

We have the problem of how to choose an aggregation function for each sublevel within the vast variety of aggregation functions available in the literature. In [31] the authors give some advices to select the most appropriate aggregation function for a specific application. First, the function must be consistent with the semantic of the aggregation process, i.e, if one is a disjunction, conjunctive or averaging aggregation functions are not suitable. Other important aspects to take into account are if the aggregation function should be symmetric, idempotent, or have a neutral or absorbing element. If the input number is always the same and what is the interpretation of the input values are also important to make a good choice of the suitable family or class. The second criterion is to select the appropriate member of that family or class, in order to produce adequate outputs for the given inputs. In our case, to address this second criterion we decided to perform an experimental study to analyze which aggregation function provides the best results in each case. To do that, we use a data set with positive and negative identification cases that helps us to make the good decision about the most suitable aggregation functions. In this sense, the SFO craniofacial correspondence of the positive cases have to be ranked before the negatives.

Based on the latter guidelines and the definitions of the Appendix, we justify the choice of the analyzed aggregation function for each sublevel as follows:

4.1. Aggregation function to combine material quality assessments and biological profile (level 2.1)

First, we have to aggregate the quality of the photo at the region (PQ_m), the quality of the bone at the region (BQ_m), and the biological profile variability of the criterion (BP_m) (see Fig. 6). These three aspects have a direct influence on the confidence of the matching degree of each particular region. Thus, we have decided to aggregate them in a single uncertainty value using an aggregation function denoted by $O_{Level2.1}(PQ_m, BQ_m, BP_m)$.
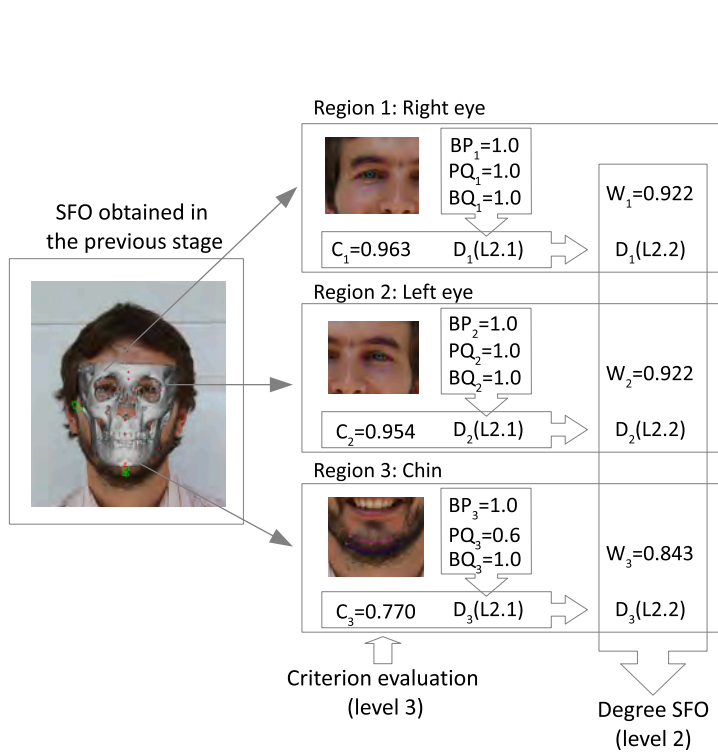
Figure 5: Graphical example of the level 2 of the fuzzy DSS framework (SFO evaluation).

Biological profile: the way biological profile affects the degree of confidence of each particular criterion can be easily modeled using fuzzy sets. According to the expert knowledge (task developed by Prof. Caroline Wilkinson, one of the most recognized experts in craniofacial identification), we have defined one or more fuzzy sets to model in which way the biological profile parameters (age, sex and ancestry) and the BMI modify the degree of confidence of each particular criterion.

Bone quality: the variations seen on bones can be described according to the bone's surface texture using the weathering stages [32]. They consider the taphonomic processes that may have affected the bones of a subject. For our application, the accuracy of the 3D model will have to be taken into account at the same time. Values of quality are set, based on the weathering stages, and are specifically associated with each region of the skull. If a specific region is deteriorated, the method to analyze the skull-face correspondence for each criterion can use it but the confidence in that criterion is reduced, so it will consequently be associated with a lower support value. Similarly to [33], where weathering stages were employed to modify the confidence of age estimation methods, we established quality indexes as ordinal numbers ranging in [0, 1]. They are assigned by a forensic expert according to the analysis of the state of the available skull. The assignation in this manner indicates the least weathering as being a perfect skull region (1) and the most weathering as being a faulty or not present region (0). We use a six-stage system in which stage 0 is determined to be a quality of 1.0, stage 1 of 0.8, stage 2 of 0.6, stage 3 of 0.4, stage 4 of 0.2, stage 5 of 0.1, and stage 6 is assigned a value of 0.0.

Photo quality: we also use a six-stage system to establish the quality of each facial region. A facial region belonging to the highest stage means that is clearly identified on the photograph and that is the ideal situation to apply the corresponding method to analyze the skull-face correspondence. On the contrary, a region of the lowest stage implies the impossibility to view that region in the photograph.

To aggregate all these values, a conjunctive behavior seems to be the best choice since it does not allow for compensation. Thus, low scores for some criteria (in this quality or biological aspects) cannot be compensated by other scores. If the quality of the bone is very bad, no matter how well the other two sources are, it applies that the matching between the skull and the face will be less reliable. However, averaging mean could be a more conservative choice. On the other hand, we consider that these three aspects affect in the same way
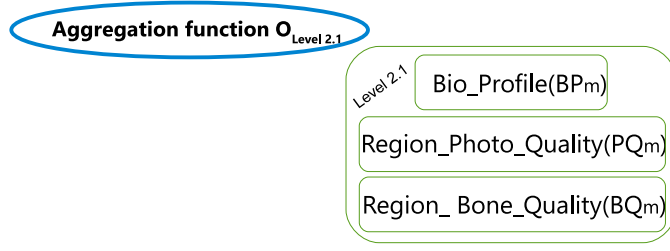
10

Figure 6: Aggregation scheme at level 2.1 of the DSS proposed framework.

to the region criterion so the aggregation function has to be symmetric. For these reasons, we decide to analyze the behavior of our system using the aggregation functions Minimum, Product, and Arithmetic Mean. Thus, at this level we can state the aggregation function as $O_{Level2.1}(PQ_m, BQ_m, BP_m)\{min, prod, mean\}$.

4.2. Aggregation function to combine the matching degree and the uncertainty value of level 2.1 (level 2.2)

Secondly, we have to aggregate the previous uncertainty sources with the matching degree of the skull and the face at the corresponding region as we can see in Fig. 7. For this application, an averaging procedure is required. The basic rule of this class of aggregation functions is that the total score cannot be above or below any of the inputs. The aggregated value is seen as some sort of representative value of all the inputs. In addition, we consider that the aggregated inputs do not have the same contribution to the total output, so a not symmetric weighted function is needed.
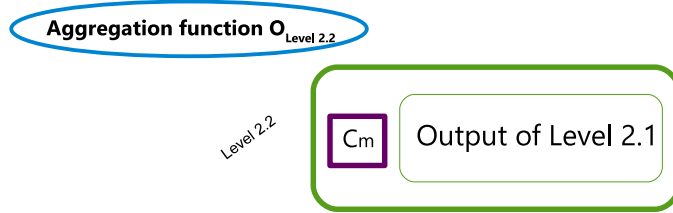
Figure 7: Aggregation scheme at level 2.2 of the DSS proposed framework.

Two of the most common weighted aggregation functions are the Weighted Arithmetic Mean and the Weighted Geometric Mean. Both functions are not symmetric. We establish different weights to each input based on the expert knowledge. This weight, a number $w_i \in [0, 1]$, represents the importance of each one. These functions are abbreviated as *wam* and *wgm*, respectively.

Since the values of the weighting vector for the *wam* and for the *wgm* must sum up to 1, we apply a simple normalization of the accuracy index with respect to their sum:

$$w_i = \frac{W_i}{\sum_{i=1}^{n} W_i} \qquad (4)$$

where W_i is the identification power of the i-region.

This aggregation function can be denoted as $O_{Level2.2}(C_m, Output_{level2.1})\{wam, wgm\}$

4.3. Aggregation function to combine the identification power and the degree of level 2.2 (level 2.3)

The final step in level 2 is to obtain the SFO evaluation degree. As explained before, there is a need to aggregate multiple degrees of support with an associated weight. Each of these degrees corresponds to the skull-face matching degree in a specific region with the corresponding uncertainty integrated (Fig. 8). The weight in this case will be the identification power of each isolated part of the face.
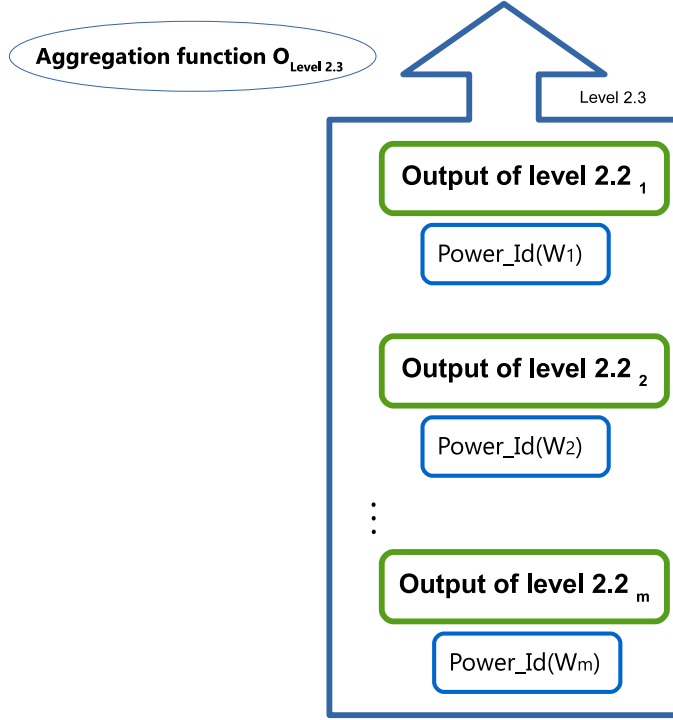
11

Figure 8: Aggregation scheme at level 2.3 of the DSS proposed framework.

The requirement for this aggregation function is that it has to be non symmetric. The identification power of a region (the weight) reflects the relative contribution of each input to the final output: the degree of a SFO evaluation.

As in the previous sublevel, we consider the most common weighted aggregation functions: the *wam* and the *wgm*. In this case, the weight which represents the importance of each input is the power identification of each region. These weigths are realistic and they are computed using a dataset of positive and negative identification cases in a similar way of the calculation of the accuracy of each method in level 3 (See Sec. 2.2 and Eq. 3). Again, the values of the weighting vector for these functions have to sum up to 1, so we apply a simple normalization of the power identification with respect to their sum (Eq. 4).

Apart from the weighted aggregations, we also use fuzzy integrals as aggregation functions. These combine the data supplied by several information sources according to a fuzzy measure, that represents the background knowledge on the information sources. In this study, we use the Choquet and the Sugeno integral and we use a Sugeno λ-measure to determine the fuzzy measure. A fuzzy measure, g, is a real valued function defined on the power set of X (the universe of discourse), 2^X, with range $[0, 1]$, satisfying the following properties: Let A and B be two subsets from X.

1. $g(\phi) = 0$, $g(X) = 1$ (boundary condition)

2. $A \subseteq B$ implies $g(A) \leq g(B)$ for all A, $B \in F$ (monotonicity)

A fuzzy measure specifies the opinion of the 'worth' or 'goodness' of each subset of information sources in evaluating a particular hypothesis. Each information source gives a belief or confidence in the hypothesis and the measure lets you know how to weight that belief or confidence, in this case the identification power or each facial region. To determine a fuzzy measure on X, we must identify $2^p - 2$ coefficients satisfying $p2^{p-1}$ conditions. To solve this drawback, some approaches have been proposed to reduce the number of parameters to be determined [29]. In this paper, we use a Sugeno λ-measure defined as in [33]:

12

Definition 1. *Let g be a fuzzy measure, then g_λ is a Sugeno λ-measure if there exists $\lambda > -1$ such that*

$$g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A)\mu(B) \tag{5}$$

holds for all $A, B \in F$.

It is to be noted that, for a Sugeno λ-measure,

$$\prod_{j=1}^{p}(1 + \lambda g_\lambda(\{x_j\})) = 1 + \lambda \tag{6}$$

holds because of the boundary condition [34]. Using the above definitions $g_\lambda(X)$ can be constructed from the fuzzy densities of the elements of X. Given the set of densities, the value λ can be easily found as the unique root greater than -1 of a simple polynomial [35].

Once λ is found, the fuzzy integral can be calculated.

- *The discrete Choquet Integral*

  The discrete Choquet integral with respect to a λ-fuzzy measure is given by

  $$C_g(\mathbf{x}) = \sum_{i=1}^{n}[x_{(i)} - x_{(i-1)}]g(H_{(i)}) \tag{7}$$

  where $\mathbf{x}_\searrow = (x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ is a non-decreasing permutation of the input $\mathbf{x}$, $x_{(0)}) = 0$ by convention, and $H_i = (i), \ldots, (n)$ is the subset of indices of $n - i + 1$ largest components of $\mathbf{x}$.

- *The Sugeno Integral*

  The Sugeno integral with respect to a λ-fuzzy measure is given by

  $$S_g(\mathbf{x}) = \max_{i=1,\ldots,n} \min(x_{(i)}, g(H_{(i)})), \tag{8}$$

  where $\mathbf{x}_\searrow = (x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ is a non-decreasing permutation of the input $\mathbf{x}$, and $H_i = (i), \ldots, (n)$.

In the following section we refer to these aggregation functions as *choq* and *sug*, respectively. Hence, at this level we can state the aggregation function as $O_{Level2.3}(Output_{Level2.2_i}, Power\_ID(W_i))\ \{wam, wgm, choq, sug\}$.

5. Experiments

The main contribution of this work is the proposal of the fuzzy hierarchical DSS for CFS framework. Hence, the experiments' aim is directly related to its validation. To do so, we have designed and developed two different experiments.

The objective of the first experimental set-up is to study the performance of the different aggregation functions within our framework. In order to analyze which are the most appropriate functions for our system, we obtain the accuracy degree for identifying positive cases in each case at the SFO evaluation level. We also perform a statistical test in order to analyze whether significant differences exist among the results of the different aggregation functions.

A second experiment has been performed with the specific focus on validating the proposed DSS framework for CFS. This considers both positive and negative identification cases, and make use of Cumulative Match Characteristic (CMC) curves to study the identification capabilities of the current implementation of the proposed DSS framework.

13

5.1. Experimental Design

The experimental design involves the sex-based cross-comparison of nine Cone Bean Computed Tomography (CBCT) models of living individuals and seven 3D skull models (acquired using a 3D structure light scanner, the Artec MHT [36]) of deceased people against a variable number of candidates and photographs. In each case there is one positive candidate with one or more photographs available. Forensic experts have made a previous filter based on sex and age, so there is not the same number of negative cases for each skull. In total, there are 16 3D skull models and 66 photographs of candidates, resulting in 33 positive and 411 negative SFOs. Table 1 summarizes the composition of the dataset employed. The SFOs have been obtained by our automatic method in [9] using the parameter values reported in that contribution. For the CBCTs positive cases, we use a ground truth dataset, whose overlays are assumed as optimal according to [37].

Table 1: Experimentation dataset summary

Skull Model	Positive SFOs	Negative SFOs
CBCT 1	2	20
CBCT 2	2	20
CBCT 3	2	20
CBCT 4	2	20
CBCT 5	2	33
CBCT 6	2	33
CBCT 7	2	33
CBCT 8	2	33
CBCT 9	2	33
3D Model 10	3	19
3D Model 11	1	21
3D Model 12	1	27
3D Model 13	4	24
3D Model 14	1	26
3D Model 15	2	25
3D Model 16	3	24
Total	33	411

For each available photograph, experts set the age and the BMI. They also marked visible regions in each photograph and the related quality of each one is established in a scale between 0 and 1 (see Sec. 4.1). In each image, experts delineated from one to four regions (depending on the visibility of them): chin contour, cranial contour, eyeball center right, and eyeball center left. These four regions were also marked on the skull 3D models. Again, experts set the quality of the bone region based on the weathering stages (see Sec. 4.1). A graphical example of this process is shown in Fig. 9.

The discriminative power of each region is reported in [10], as well as the corresponding methodology to obtain it. These values are computed using Eq. 3 of Sec. 2.2, and it is understood as the capability to discriminate in the decision making process after aggregating the best methods for modeling each region at level 3. The cranial contour follows the same criterion as the chin contour. Thus, given the inability to calculate identification power from CBCT ground truth cases (CBCT does not include the upper part or the skull), the value of the identification power is taken the same as the chin case. Table 2 summarizes these values for each region used in this work.

For these implemented criteria the influence related with biological profile was defined by Prof. Wilkinson according to her expert knowledge and represented using fuzzy sets in Fig. 10:

As can be seen in Fig. 10, the chin criterion is less reliable after 60 years old, decreasing to 0.25 from 75. The same criterion is unreliable with BMI values above 35 (changes in fat will alter the shape of the chin) [30]. Neither the eyeball position nor the cranial contour are affected by any morphological aspect.

Marking skull regions	Marking photograph regions
Setting skull regions qualities	**Setting photo qualities and profile**
Cranial bone quality: 1.0 Orbit right bone quality: 0.8 Orbit left bone quality: 0.8 Chin bone quality: 1.0	Age: 47 BMI: 22 Cranial contour quality: 0.8 Eyeball center right quality: 1.0 Eyeball center left quality: 0.8 Chin contour quality: 0.6

Figure 9: Example of marking regions and setting qualities in both skull and photograph.

Table 2: Identification power of each isolated region [10].

Region	Power of identification
Chin contour	0.843
Cranial contour	0.843
Eyeball (left and right) center position	0.922
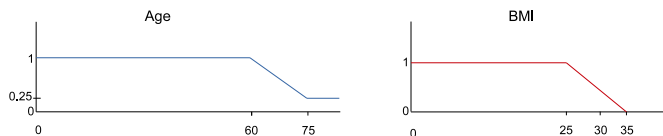
Figure 10: Defined fuzzy sets to model morphological aspects for the chin outline. At the left the age related confidence and at the right the BMI related confidence.

For the level 2.2, the experts in Prof. Wilkinson lab established that the uncertainty sources have a third of influence and the matching degree have two thirds of influence. So the weighted vector used in this case is $\mathbf{w} = (\frac{1}{3}, \frac{2}{3})$.

Once the SFOs are achieved, all the previous data are taken into account to obtain the final SFO degree as explained in Section 4. We tested the different aggregation functions mentioned in this section for each sublevel, that is: level 2.1: *min*, *prod* and *mean*; level 2.2: *wam* and *wgm*; level 2.3: *wam*, *wgm*, *choq* and *sug*. Accordingly, 24 different combinations are analyzed.

The final values reported for each skull are used to rank the candidates based on their chance of being the actual subject.

5.2. Study of aggregation functions

This first study consists in analyzing the behavior of the different combinations of the aggregation functions. To do that, we calculate the accuracy of the 24 different fuzzy DSSs (each of them represented by a different combination of aggregation functions) over all cases as explained in Section II.B and Equation 3.

Table 3 shows the mean accuracy for the system using each combination of the selected aggregation functions. As can be seen, the combination *mean-wam-wam* achieves the highest value, with 0.8550 of mean accuracy. *Mean-wgm-wam*, *prod-wam-wam*, *min-wam-wam*, *prod-wgm-wam* and *min-wgm-wam* present similar results, with 0.8516, 0.8461, 0.8394, 0.8378 and 0.8137, respectively. The next methods show a bigger gap with respect to the previous ones, *mean-wam-choq* and *mean-wgm-choq* with 0.7742 and 0.7733 of

mean accuracy. The wost performances are achieved by combinations *min-wam-wgm* and *min-wgm-wgm*, obtaining a mean accuracy of 0.6520 and 0.6494, respectively.

Table 3: Mean accuracy of each combination method and ranking obtained through Friedman's test.

Combination method	Mean accuracy	Ranking
mean-wam-wam	**0.8550**	**7.561**
mean-wgm-wam	0.8516	8.030
prod-wam-wam	0.8461	7.970
min-wam-wam	0.8394	8.485
prod-wgm-wam	0.8378	8.955
min-wgm-wam	0.8137	9.758
mean-wam-choq	0.7742	11.030
mean-wgm-choq	0.7733	11.167
mean-wam-sug	0.7460	11.848
prod-wam-choq	0.7330	11.788
min-wam-choq	0.7248	12.379
prod-wam-sug	0.7014	13.379
mean-wgm-sug	0.6914	14.348
prod-wgm-choq	0.6895	14.303
min-wgm-sug	0.6893	14.091
min-wam-sug	0.6882	14.076
min-wgm-choq	0.6797	14.939
mean-wgm-wgm	0.6795	14.106
mean-wam-wgm	0.6757	14.212
prod-wam-wgm	0.6633	14.955
prod-wgm-wgm	0.6593	15.242
prod-wgm-sug	0.6525	15.909
min-wam-wgm	0.6520	15.561
min-wgm-wgm	0.6494	15.909

Friedman test [38], a nonparametric test for analysis of variance, aims to test a null hypothesis stating that the mean total accuracy of all the methods are the same. We have set the experiment level of significance in $\alpha = 0.05$.

Table 3 summarizes the ranking obtained by Friedman's test for the studied methods. The result of applying Friedman's test is $\chi^2_F = 116.21$ and a p-value of 1.99×10^{-14}. Given that the p-value of Friedman are lower than the level of significance considered, $\alpha = 0.05$, there are significant differences among the observed results. Attending to these results, a post-hoc statistical analysis could help us to detect concrete differences among methods.

In particular, Bonferroni-Dunn test [39] is accomplished to detect significant differences among a control approach and the rest. The control method in this case is the combination of *mean-wam-wam*. Fig. 11 displays a graphical representation, including the rankings obtained for each method and the critical difference for each value of α. A Bonferroni-Dunn's graphic illustrates the difference among rankings obtained for each approach. The horizontal line represents the level of significance considered in the study at height equal to the sum of the ranking of the control method and the corresponding critical difference computed by the Bonferroni-Dunn method. Those bars which exceed this line are the associated to an approach with worse performance than the control method.

The Bonferroni-Dunn's test shows us the following significant differences with *mean-wam-wam* as control method: *mean-wam-wam* is better than every method except *prod-wam-wam*, *mean-wgm-wam*, *min-wam-wam*, *prod-wgm-wam*, *min-wgm-wam*, *mean-wam-choq*, *mean-wgm-choq*, *prod-wam-choq*, *mean-wam-sug* and *min-wam-choq* with $\alpha = 0.05$ (13/23 methods). Although *mean-wam-wam* obtains the lowest error and ranking rates, the Bonferroni-Dunn's test is not able to distinguish it as better than all the following 10
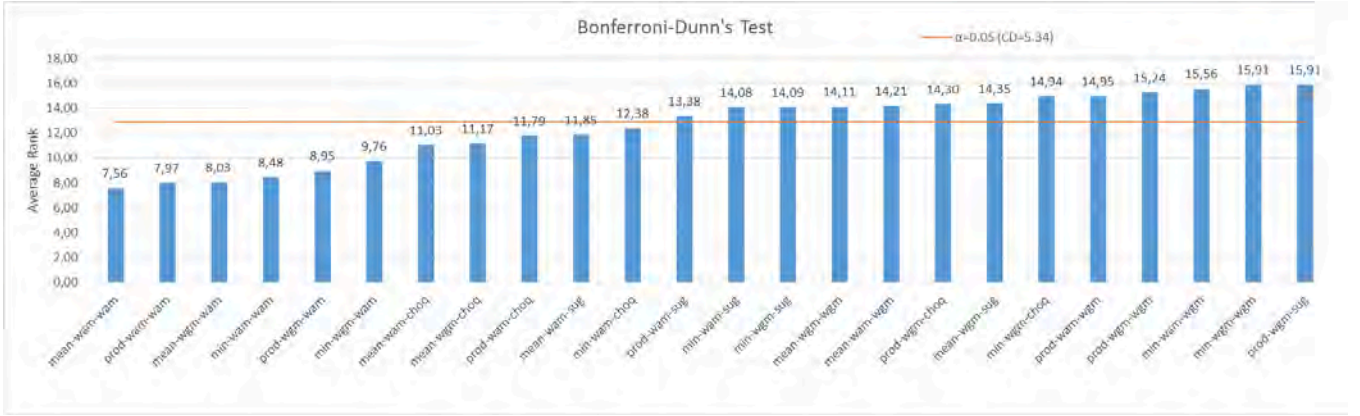
16

Figure 11: Bonferroni-Dunn's graphic corresponding to the results.

approaches.

We applied more powerful procedures, such as Holm's and Hochberg's, for comparing the control method with the others. However, the same conclusions are achieved. The results are shown by computing p-values for each comparison. Table 4 show the p-value obtained for Bonferroni-Dunn', Holm's and Hochberg's procedures.

Table 4: p-values on the results (*mean-wam-wam* is the control method)

mean-wam-wam vs.	Bonferroni-Dunn p	Holm p	Hochberg p
prod-wam-wam	1.0	1.0	0.8142
mean-wgm-wam	1.0	1.0	0.8142
min-wam-wam	1.0	1.0	0.8142
prod-wgm-wam	1.0	1.0	0.8142
min-wgm-wam	1.0	1.0	0.8142
mean-wam-choq	1.0	0.2774	0.2774
mean-wgm-choq	0.8811	0.2682	0.2682
prod-wam-choq	0.3488	0.1239	0.1213
mean-wam-sug	0.3167	0.1239	0.1213
min-wam-choq	0.1298	0.0564	0.0564
prod-wam-sug	0.0191	0.0091	0.0091
min-wam-sug	0.0042	0.0024	0.0022
min-wgm-sug	0.0040	0.0024	0.0022
mean-wgm-wgm	0.0039	0.0024	0.0022
mean-wam-wgm	0.0031	0.0020	0.0020
prod-wgm-choq	0.0025	0.0017	0.0017
mean-wgm-sug	0.0022	0.0016	0.0016
min-wgm-choq	0.0005	0.0004	0.0004
prod-wam-wgm	0.0005	0.0004	0.0004
prod-wgm-wgm	0.0002	0.0002	0.0002
min-wam-wgm	$9.9223.10^{-5}$	$9.0594.10^{-5}$	$9.0594.10^{-5}$
min-wgm-wgm	$3.7259.10^{-5}$	$3.7259.10^{-5}$	$3.5639.10^{-5}$
prod-wgm-sug	$3.7259.10^{-5}$	$3.7259.10^{-5}$	$3.5639.10^{-5}$

5.3. DSS framework validation

Our second performed experiment studies the capability of our system to identify the correct individual. The results are reported using CMC curves [40] along with average rank 10 identification accuracy. A CMC curve captures the percentage (or probability) that the correct match for a case is present in a candidate

17

list of the r best matches, where r denotes the rank. Therefore, rank 10 identification accuracy denotes the probability that the correct match is one of the subjects in a list of the top 10 matches provided by the system.

There is no previous work in the literature with a similar proposal to that presented in this manuscript. For this reason, we cannot compare the obtained results with existing CFS computer-based proposals. We could focus instead on the performance demonstrated by human experts. The reliability of CFS in human identification has been assessed by different authors [41, 42, 43, 23, 44] achieving confronted conclusions about the technique's reliability, with positive matches ranging from 70% to 100%, and false negatives from 0% to 20%.

The reliability studies reported in the forensic literature are fraught with limitations, such as the absence of an objective measurement of the craniofacial superimposition match, limitations of the technical equipment, imprecision in landmark location while performing landmark-based methods, absence of soft tissue data for the tested population, deficient quality of the skull 3D models, postmortem photographs, limited samples, lack of appropriate statistical analysis, and the absence of inter and intra observer studies.

The most recent work [45], a multiple-lab study with 26 participants from 17 different institutions involving 60 CFS cases showed a global average performance of 79% of correct identifications. This percentage scaled up to 84% in a similar study [46] where the participants were asked to follow MEPROCS best practices [4]. We should notice that these two studies were laid out with the intention of identifying good and bad practices and to validate MEPROCS standards, respectively, rather than to test CFS reliability.

The last possibility to provide a reference for the identification performance can be obtained through the examination of other automatic or semi-automatic identification methods that have been proposed to model different forensic anthropology techniques. In [47], a computerized clavicle identification system was presented. The method quantifies the clavicle outline shape from the skeletons and postero-anterior AM chest radiographs to rank individuals. The results show that a positive predictive value of 78% is achieved when considering the 21 first classified bones (rank-21), increasing to 90% around rank-100.

In our case we compare the six best aggregation function combinations of the previous study, i.e. *mean-wam-wam*, *prod-wam-wam*, *mean-wgm-wam*, *min-wam-wam*, *prod-wgm-wam* and *min-wgm-wam*. Although they all obtained similar results without significant differences with respect to the mean accuracy (see in Sec. 5.2), CMC curves allow us to differentiate among them (see Fig. 12).

CMC curves show us that *mean-wam-wam* and *mean-wgm-wam* have the best performance to identify the actual subject. Although they do not have the highest value for identifying the correct match in rank 1 and 8, they have the best performance in the remaining scenarios (Fig. 12). As it can be seen, when our system uses these aggregation functions it is able to rank the correct individual within the five first positions with more than a 70% of accuracy. This rate increases if we consider the 10 first positions, reaching more than 90% of identification accuracy and for the most of the combinations except *min-wgm-wam* that presents a worse behavior. Although these results are not good enough to consider the current proposal for identification purposes on its own, our fuzzy DSS has demonstrated promising capabilities to filter (shortlist) cases. In fact, it has showed a significant improvement in comparison with the only 'comparable' method [47] (dealing with a simpler identification technique since it involves a bone to bone comparison, a 3D clavicle and the same bone in a radiograph) which needed to look at around 100 cases to reach a 90% identification rate.

Finally, in Fig. 13 the eight first visual results for one case of the ranking using the *mean-wam-wam* combination are depicted. In the figure, positive cases have the highest SFO degrees, so they are ranked at the firsts positions.

6. Discussion and Conclusions

In the present contribution, we propose a complete framework for a DSS in CFS. The proposal develops information fusion concerning skull-face anatomical correspondence at three different levels: criterion evaluation, SFO evaluation, and CFS evaluation. We classify the uncertainty sources and degrees of confidence involved in this process as related to bone, image, skull-face overlays, morphological aspects and used methods. In this study, we focus on the SFO evaluation level. Within this stage, we distinguish three
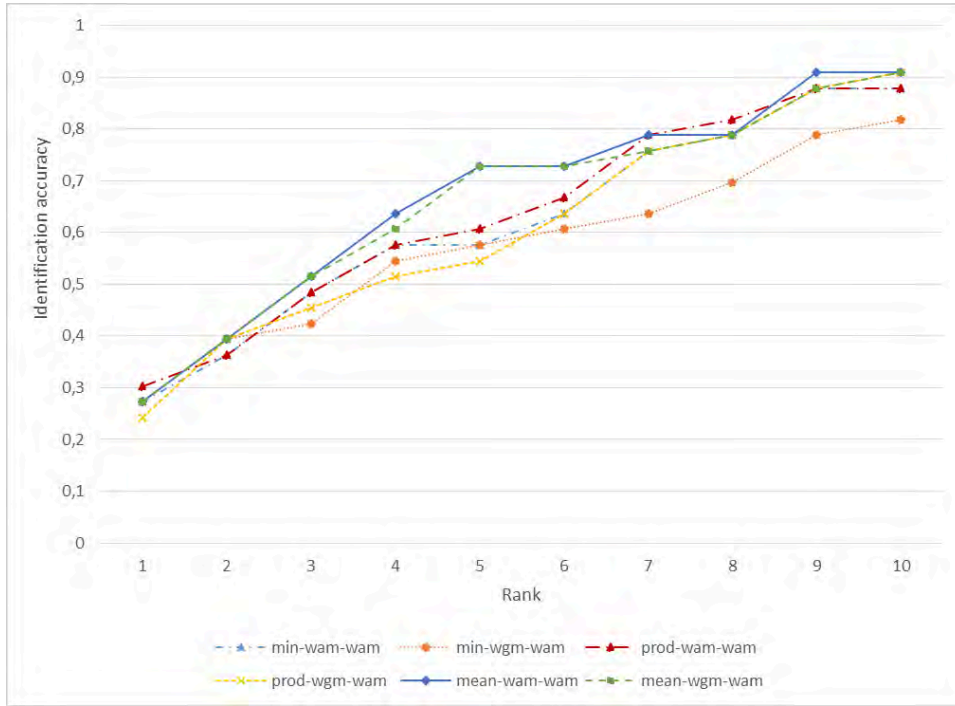
18

Figure 12: CMC curves corresponding to the six best combinations of the proposed fuzzy DSS.

sublevels with different conditions of information fusion. For this reason we perform an experimental study to analyze which aggregation function provides the best results. The first sublevel aggregates the sources of uncertainty of the bone and the image, and the biological profile variability. For this sublevel, we propose to use the minimum (min), the product ($prod$), and the arithmetic mean ($mean$) as aggregation functions. The second sublevel consists of integrating this aggregation with the matching degree of the skull and the face. In this case, we study the use of the weighted arithmetic mean (wam) and the weighted geometric mean (wgm). Finally, at the third sublevel, we obtain the degree of the SFO craniofacial correspondence by aggregating the different previous values for all the regions taking into account the discriminative power of the isolated region as weight. To study that, we test the wam, the wgm, the Sugeno (sug) integral, and the Choquet ($choq$) integral. We perform an experiment with positive and negative identification cases. We both analyze the results studying the mean accuracy of each approach and its capability of identification.

With respect to the mean accuracy, the combination $mean$-wam-wam shows the best results in our system. It also presents the first position at the Friedman ranking. Statistical tests show this combination of the aggregation functions is significant better than 13 of the 24 studied methods, but we can not confirm that there are significant differences between this method and the remaining 10 first approaches. However, according to these results, we can conclude that the best aggregation function for the sublevel 2.3 is the weighted arithmetic mean, since the six highest accuracies are obtained with this function (always more than 0.8).

Finally, we validate the DSS framework as an identification system. At this point the identification accuracy is insufficient for running independently as an automatic identification tool. However, it can be already used as a powerful shortlisting tool capable of successfully filtering out a large number of candidates (20 out of 30) in 90% of the identification cases we tested.

We identify the following sources of uncertainty/error that in our opinion are limiting the accuracy of the automatic DSS for CFS:

- The quality of the used SFOs: as explained in the introduction section, reaching an optimal SFO accuracy is still an open field of research and manual refinement of SFO results is currently needed
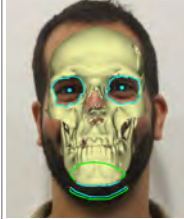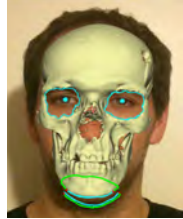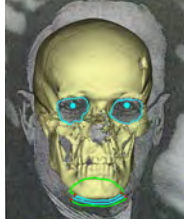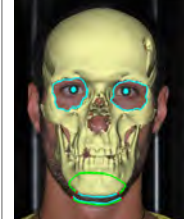
Ranking position	1 (positive case)	2 (positive case)	3 (negative case)	4 (negative case)
SFO				
Eyeball reference position				
Chin Criterion	0.7753	0.6906	0.8140	0.7841
Eye Right Criterion	0.8943	0.8581	0.8147	0.8112
Eye Left Criterion	0.8112	0.8657	0.8129	0.7626
Cranial Criterion	-	-	-	-
Degree SFO	0.8787	0.8721	0.8549	0.8434
Ranking position	5 (negative case)	6 (negatitve case)	7 (negative case)	8 (negative case)
SFO				
Eyeball reference position				
Chin Criterion	0.8049	0.8140	0.8140	0.8037
Eye Right Criterion	0.7077	0.8057	0.6961	0.7516
Eye Left Criterion	0.7185	0.6679	0.6521	0.8019
Cranial Criterion	-	-	-	0.0000
Degree SFO	0.8139	0.8115	0.8050	0.6363

Figure 13: Visual results of the eight firsts positions of the ranking for the combination *mean-wam-wam* for one cranial case.

over an important number of cases. One of the issues negatively affecting the accuracy of the automatic SFO method is the consideration of the mandible as a rigid part of the skull.

- The number of SFOs considered: our experimental set-up does not take into consideration the level 1, still to be developed, in which CFS instances are evaluated (using several AM images of the same person). It was demonstrated in [44] that CFS is more reliable when two or more facial photos of the same individual taken from different points of view are jointly used in the examination.

- The number of regions evaluated: we have only implemented four regional criteria, namely, the morphological and spatial relation between the face and the skull chin shape, the relative position of the orbits and the eyeballs (two regions, one for each eye), and the morphological correspondence between the cranial contour in the skull and the face. In addition, the last criterion could be only evaluated in the few cases where the candidate is bald, and the one related with the orbits-eyeball is partially evaluated in many photographs, those with a profile pose.

- 3D and 2D regions delimitation: this task is subjective and it is expected that two different experts, or even the same expert in different times, will mark a different slightly region. Among all the limitation factors we think this is the least important.

In any case, it is important to remark that this fuzzy hierarchical DSS framework is the first automatic CFS system and the obtained results are promising taking into account the actual scope of improvement.

In this sense, in future works, we aim to include more families of criteria for assessing the craniofacial correspondence and test the DSS with a larger dataset. We also plan to perform inter and intra expert studies to measure inter-intra variability marking those criteria (regions) on the facial photograph and the 3D skull model. Besides, in order to calculate the accuracy of each method and the power identification of each criterion we can take into account the variety of the instances used, separating into groups of ages, sex, ethnical, etc. Thus, different overall values for different kind of populations can be used to strengthen the final output. To do that, we need a bigger dataset of cases since the current is still insufficient. In addition, we could study the use of regression techniques to build aggregation functions from our specific data following [48]. Another aspect to take into consideration is to aggregate the different SFOs of the same person in order to obtain a final degree of a CFS case (i.e., the design of level 1). Besides, we keep on achieving more accurate SFOs through the study of new camera parametrizations, optimization strategies, and the modelization of the mandible rotation and translation movement and its inclusion with the SFO optimization process. Our final objective is to reach a higher identification accuracy. The need of objective assessment and automation has been justified along this manuscript, although and in depth justification was recently provided in [5]. The performance of newly developed CFS methods should be compared with a usual forensic dataset of known case studies. The methods proposed could be applied to solve these cases in order to validate them. Then, the obtained results would be compared with the identification previously determined by forensic experts. Despite there are promising works available in the specialized literature in this line, automatic methods are not implemented yet due to the inability to test their behavior in an objective manner. In this manuscript, authors emphasize the requirement of statistically significant reliability studies that tackle these challenges to obtain a more solid picture on the reliability of CFS. Accordingly, the identification values reported in the present work will be used as reference in the future.

Acknowledgments

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

[2] C. Wilkinson, C. Rynn, Craniofacial identification, Cambridge University Press, 2012.

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

[4] S. Damas, C. Wilkinson, T. Kahana, E. Veselovskaya, A. Abramov, R. Jankauskas, P. Jayaprakash, E. Ruiz, F. Navarro, M. Huete, E. Cunha, F. Cavalli, J. Clement, P. Leston, F. Molinero, T. Briers, F. Viegas, K. Imaizumi, D. Humpire, O. Ibáñez, Study on the performance of different craniofacial superimposition approaches (ii): best practices proposal, Forensic Sci Int 257 (2015) 504–508.

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

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] C. Campomanes-Alvarez, O. Ibáñez, O. Cordón, Design of criteria to assess craniofacial correspondence in forensic identification based on computer vision and fuzzy integrals, Applied Soft Computing 46 (2016) 596–612.

[11] C. Campomanes-Alvarez, O. Ibáñez, O. Cordón, Modeling the consistency between the bony and facial chin outline in craniofacial superimposition, in: 16th World Congress of the International Fuzzy Systems Association (IFSA), 2015, pp. 1612–19.

[12] C. Campomanes-Alvarez, O. Ibáñez, O. Cordón, Experimental study of different aggregation functions for modeling craniofacial correspondence in craniofacial superimposition, in: the 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2016), 2016, pp. 437–444.

[13] P. Broca, Instructions craniologiques et craniométriques de la Société d'anthropologie de Paris, Vol. 2, G. Masson, 1875.

[14] A. Bertillon, The bertillon system of identification, McClaughry, Ed., Chicago, IL.

[15] B. A. Nickerson, P. A. Fitzhorn, S. K. Koch, M. Charney, A methodology for near-optimal computational superimposition of two-dimensional digital facial photographs and three-dimensional cranial surface meshes, J Forensic Sci 36 (1991) 480–500.

[16] J. Huang, M. Zhou, F. Duan, Q. Deng, Z. Wu, Y. Tian, The weighted landmark-based algorithm for skull identification, in: Computer Analysis of Images and Patterns, Springer, 2011, pp. 42–48.

[17] W. Jin, G. Geng, K. Li, Y. Han, Parameter estimation for perspective projection based on camera calibration in skull-face overlay, in: Virtual Reality and Visualization (ICVRV), 2013 International Conference on, IEEE, 2013, pp. 317–320.

[18] 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 (1) (2015) 227–236.

[19] 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.

[20] A. K. Ghosh, P. Sinha, An economised craniofacial identification system, Forensic Science International 117 (1) (2001) 109–119.

[21] 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.

[22] 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 (1) (2008) 34–40.

[23] P. T. Jayaprakash, G. J. Srinivasan, M. G. Amravaneswaran, Cranio-facial morphanalysis: a new method for enhancing reliability while identifying skulls by photo superimposition, Forensic Sci Int 117 (1) (2001) 121–143.

[24] O. Ibáñez, A. Valsecchi, F. Cavalli, M. I. Huete, B. R. Campomanes-Alvarez, C. Campomanes-Alvarez, R. Vicente, D. Navega, A. Ross, C. Wilkinson, et al., Study on the criteria for assessing skull-face correspondence in craniofacial superimposition, Legal Medicine 23 (2016) 59–70.

[25] D. V. Pesce, E. Vacca, F. Potente, T. Lettini, M. Colonna, Shape analytical morphometry in computer-aided skull identification via video superimposition, Iscan MY, Helmer RP. Forensic analysis of the skull: craniofacial analysis, reconstruction, and identification. New York: Wiley-Liss, 1993.

[26] M. Yoshino, H. Matsuda, S. Kubota, K. Imaizumi, S. Miyasaka, S. Seta, Computer-assisted skull identification system using video superimposition, Forensic Sci Int 90 (3) (1997) 231–244.

[27] A. Ricci, G. L. Marella, M. A. Apostol, A new experimental approach to computer-aided face/skull identification in forensic anthropology, Am J Foren Med Path 27 (1) (2006) 46–49.

[28] C. P. Pappis, N. I. Karacapilidis, A comparative assessment of measures of similarity of fuzzy values, Fuzzy Set Syst 56 (2) (1993) 171–174.

[29] M. Sugeno, Theory of fuzzy integrals and its applications, Tokyo Institute of Technology, 1974.

[30] J. G. Clement, Craniofacial identification in forensic medicine, Arnold, 1998.

[31] G. Beliakov, A. Pradera, T. Calvo, Aggregation functions: A guide for practitioners, Vol. 221, Springer, 2007.

[32] J. E. Buikstra, D. H. Ubelaker, Standards for data collection from human skeletal remains.

[33] M. F. Anderson, D. T. Anderson, D. J. Wescott, Estimation of adult skeletal age-at-death using the sugeno fuzzy integral, Am J Phys Anthropol 142 (1) (2010) 30–41.

[34] H. Imai, V. Torra, On a modeling of decision making with a twofold integral., in: EUSFLAT Conf., 2003, pp. 714–717.

[35] H. Tahani, J. M. Keller, Information fusion in computer vision using the fuzzy integral, IEEE T Syst Man Cyb 20 (3) (1990) 733–741.

[36] Artec 3d scanners, www.artec3d.com/3d-scanner/artec-spider/.

[37] 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 (3) (2015) 569–81.

[38] M. Friedman, A comparison of alternative tests of significance for the problem of m rankings, Ann Math Stat 11 (1) (1940) 86–92.

[39] O. J. Dunn, Multiple comparisons among means, J Am Stat Assoc 56 (293) (1961) 52–64.

[40] A. K. Jain, S. Z. Li, Handbook of face recognition, Vol. 1, Springer, 2005.

[41] D. Austin-Smith, W. R. Maples, The reliability of skull/photograph superimposition in individual identification, Journal of Forensic Science 39 (2) (1994) 446–455.

[42] D. Chai, Y. Lan, C. Tao, R. Gui, Y. Mu, J. Feng, W. Wang, J. Zhu, A study on the standard for forensic anthropologic identification of skull-image superimposition, J. Forensic Sci 34 (6) (1989) 1343–1356.

[43] 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.e1–6.

[44] M. Yoshino, K. Imaizumi, S. Miyasaka, S. Seta, Evaluation of anatomical consistency in craniofacial superimposition images, Forensic science international 74 (1) (1995) 125–134.

[45] O. Ibáñez, R. Vicente, D. S. Navega, C. Wilkinson, P. T. Jayaprakash, M. I. Huete, T. M. Briers, R. Hardiman, F. Navarro, E. Ruiz, F. Cavalli, K. Imaizumi, R. Jankauskas, E. Veselovskaya, A. Abramov, P. Lestón, F. Molinero, J. Cardoso, J. Cagdir, D. Humpire, Y. Nakanishi, A. Zeuner, A. H. Ross, D. Gaudio, S. Damas, Study on the performance of different craniofacial superimposition approaches (i), Forensic Sci Int 257 (2015) 496–503.

[46] O. Ibáñez, R. Vicente, D. Navega, C. Campomanes-Álvarez, C. Cattaneo, R. Jankauskas, M. Huete, F. Navarro, R. Hardi-

man, E. Ruiz, et al., Meprocs framework for craniofacial superimposition: Validation study, Legal Medicine 23 (2016) 99–108.

[47] C. N. Stephan, B. Amidan, H. Trease, P. Guyomarc'h, T. Pulsipher, J. E. Byrd, Morphometric comparison of clavicle outlines from 3d bone scans and 2d chest radiographs: a shortlisting tool to assist radiographic identification of human skeletons, Journal of forensic sciences 59 (2) (2014) 306–313.

[48] G. Beliakov, How to build aggregation operators from data, International Journal of Intelligent Systems 18 (8) (2003) 903–923.

Appendix A. Definitions

We introduce some basic definitions related with aggregation functions based on [31].

Definition 2. *An aggregation function is a function of $n > 1$ arguments that maps the (n-dimensarional) unit cube onto the unit interval $f \colon [0,1]^n \to [0,1]$, with the properties*

(i) $f \underbrace{(0, 0, \ldots, 0)}_{n-times} = 0 \; and \; f \underbrace{(1, 1, \ldots, 1)}_{n-times} = 1.$

(ii) $\mathbf{x} \leq \mathbf{y}$ *implies* $f(\mathbf{x}) \leq f(\mathbf{y})$ *for all* $\mathbf{x}, \mathbf{y} \in [0,1]^n$

Definition 3. *An extended aggregation function is a mapping*

$$F : \bigcup_{n \in 1, 2, \ldots} [0,1]^n \to [0,1],$$

such that the restriction of this mapping to the domain $[0,1]^n$ for a fixed n is $n-ary$ aggregation function f, with the convention $F(x) = x$ for $n = 1$.

This allows us to define and work with such families of functions of any number of arguments.

There are several semantics of aggregation, and the main classes are determined according to these semantics. In some cases we require that the high and low inputs average each other, in other cases aggregation functions model logical connectives (disjunction and conjunction), so that the inputs reinforce each other, and sometimes the behavior of aggregation functions depends on the inputs. The four main classes of aggregation functions are [31]:

- Averaging,
- Conjunctive,
- Disjunctive,
- Mixed.

Definition 4. *An aggregation function f has averaging behavior (or is averaging) if for every $\mathbf{x}$ it is bounded by $min(\mathbf{x}) \leq f(\mathbf{x}) \leq max(\mathbf{x})$.*

Definition 5. *An aggregation function f has conjunctive behavior (or is conjunctive) if for every $\mathbf{x}$ it is bounded by $f(\mathbf{x}) \leq min(\mathbf{x}) = min(x_1, x_2, \ldots, x_n)$.*

Definition 6. *An aggregation function f has disjunctive behavior (or is disjunctive) if for every $\mathbf{x}$ it is bounded by $f(\mathbf{x}) \geq max(\mathbf{x}) = max(x_1, x_2, \ldots, x_n)$.*

Definition 7. *An aggregation function f is mixed if it does not belong to any of the above classes, i.e., it exhibits different types of behavior on different parts of the domain.*

Last, we define an important property of aggregation functions for this application: symmetry.

Definition 8. *An aggregation function f is called symmetric, if its values does not depend on the permutation of the aggregation of the elements, i.e.,*
$f(x_1, x_2, \ldots, x_n) = f(x_{P(1)}, x_{P(2)}, \ldots, x_{P(n)})$,
for every x and every permutation $P = (P(1), P(2), \ldots, P(n))$ of $(1, 2, \ldots, n)$.

ACCEPTANCE AND RESIGNATION OF THE PUBLICATION'S CO-AUTHORS

Publication/article: ***HIERARCHICAL INFORMATION FUSION FOR DECISION MAKING IN CRANIOFACIAL SUPERIMPOSITION***

The co-authors:

Mr./Ms.	Oscar Ibáñez
Mr./Ms.	Oscar Cordón
Mr./Ms.	Caroline Wilkinson
Mr./Ms.	
Mr./Ms.	
Mr./Ms.	

Declare that they:

☒ Accept and authorize the use of the above mentioned publication/article as part of the documentation for the deposit and defence of the doctoral thesis of Mr./Ms. **Carmen Campomanes Álvarez** titled *AUTOMATION OF THE ASSESSMENT OF CRANIOFACIAL SUPERIMPOSITION USING SOFT COMPUTING AND COMPUTER VISION*,

☒ Have not used the above mentioned publication/article as part of the documentation for the deposit and defence of another doctoral thesis and/or refuse to use it for a future doctoral thesis

Granada, 2 , mayo 2017

Signed: Oscar Ibáñez   Signed: Oscar Cordón   Signed: Caroline Wilkinson

Bibliography

[AAMG⁺06]   S. Al-Amad, M. McCullough, J. Graham, J. Clement, and A. Hill. Craniofacial identification by computer-mediated superimposition. *Journal of Forensic Odonto-Stomatology*, 24:47–52, 2006.

[AAW10]   M. F. Anderson, D. T. Anderson, and D. J. Wescott. Estimation of adult skeletal age-at-death using the sugeno fuzzy integral. *American Journal of Physical Anthropology*, 142(1):30–41, 2010.

[AFLGMFAL04]   S. Aja-Fernández, R. Luis-García, M. Martín-Fernández, and C. Alberola-López. A computational tw3 classifier for skeletal maturity assessment. A computing with words approach. *Journal of Biomedical Informatics*, 37:99–107, 2004.

[AISB95]   W. A. Aulsebrook, M. Y. Iscan, J. M. Slabbert, and P. Beckert. Superimposition and reconstruction in forensic facial identification: a survey. *Forensic Science International*, 75:101–120, 1995.

[ARP07]   A. M. Albert, K. Ricanek, and E. Patterson. A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. *Forensic Science International*, 172(1):1–9, 2007.

[ASC⁺09]   D. De Angelis, R. Sala, A. Cantatore, M. Grandi, and C. Cattaneo. A new computer-assisted technique to aid personal identification. *International Journal of Legal Medicine*, 123:351–356, 2009.

[ASM94]   D. Austin-Smith and W. R. Maples. The reliability of skull/photograph superimposition in individual identification. *Journal of Forensic Science*, 39(2):446–455, 1994.

[BCC06]   I. Bloch, O. Colliot, and R. M. Cesar. On the ternary spatial relation "between". *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 36(2):312–327, 2006.

[BCDS09]   L. Ballerini, O. Cordón, S. Damas, and J. Santamaría. Automatic 3d modeling of skulls by scatter search and heuristic features. In E. Avineria, M. Koepen, K. Dahal, Y. Sunitiyoso, and R. Roy, editors, *Applications of Soft Computing, Updating the State of the Art*, page 149–158. Springer, Berlin, Germany, 2009.

[BDW86]   R. J. Bastiaan, G. D. Dalitz, and C. Woodward. Video superimposition of skulls and photographic portraits-a new aid to identification. *Journal of Forensic Science*, 31:1373–1379, 1986.

[Ber96]   A. Bertillon. *The Bertillon System of Identification*. McClaughry RW, Chicago, IL, 1st edition, 1896.

[Ber07]   E. S. Berner. *Clinical Decision Support Systems: Theory and Practice.* Springer, New York, NY, 2007.

[BFM97]   T. Bäck, D. B. Fogel, and Z. Michalewicz, editors. *Handbook of Evolutionary Computation.* IOP Publishing Ltd., Bristol, UK, 1st edition, 1997.

[BH89]   L. M. Brocklebank and C. J. Holmgren. Development of equipment for the standardization of skull photographs in personal identifications by photographic superimposition. *Journal of Forensic Science*, 34(5):1214–1221, 1989.

[BK95]   I. Bajnóczky and L. Királyfalvi. A new approach to computer-aided comparison of skull and photograph. *International Journal of Legal Medicine*, 108:157–161, 1995.

[BKA+03]   Y. Bilge, P. Kedici, Y. Alakoc, K. Ulkuer, and Y. Ilkyaz. The identification of a dismembered human body: a multidisciplinary approach. *Forensic Science International*, 137:141–146, 2003.

[BKR+10]   C. Birngruber, K. Kreutz, F. Ramsthaler, J. Krähahn, and M. A. Verhoff. Superimposition technique for skull identification with AFLOAT® software. *International Journal of Legal Medicine*, 124:471–475, 2010.

[Blo96]   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, pages 1307–1312, 1996.

[Blo99]   I. Bloch. On fuzzy distances and their use in image processing under imprecision. *Pattern Recognition*, 32:1873–1895, 1999.

[BM92]   P. Besl and N. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 14(2):239–256, 1992.

[Bon97]   P. P. Bonissone. Soft computing: the convergence of emerging reasoning technologies. *Soft Computing*, 1:6–18, 1997.

[Boo97]   F. L. Bookstein. *Morphometric tools for landmark data: geometry and biology.* Cambridge University Press, 1997.

[BPC07]   G. Beliakov, A. Pradera, and T. Calvo. *Aggregation functions: A guide for practitioners*, volume 221. Springer, 2007.

[BR02]   F. Bernardini and H. Rushmeier. The 3d model acquisition pipeline. In *Computer Graphics Forum*, volume 21, pages 149–172, 2002.

[BR03]   I. Bloch and A. Ralescu. Directional relative position between objects in image processing: a comparison between fuzzy approaches. *Pattern Recognition*, 36(7):1563–1582, 2003.

[Bro75]   P. Broca. *Instructions craniologiques et craniométriques de la Société d'anthropologie de Paris*, volume 2. G. Masson, 1875.

[BSA05]   H. Biwasaka, K. Saigusa, and Y. Aoki. The applicability of holography in forensic identification: a fusion of the traditional optical technique and digital technique. *Journal of Forensic Science*, 50(2):393–399, 2005.

[BU94]   J. E. Buikstra and D. H. Ubelaker. Standards for data collection from human skeletal remains. 1994.

[BW99]   K. R. Burns and J. Wallington. *Forensic anthropology training manual.* Prentice Hall Saddle River, NJ, 1999.

[CÁCD13]   B. R. Campomanes-Álvarez, O. Cordón, and S. Damas. Evolutionary multi-objective optimization for mesh simplification of 3*d* open models. *Integrated Computer-Aided Engineering*, 20(4):375–390, 2013.

[CAIN⁺14]   B. R. Campomanes-Álvarez, O. Ibáñez, F. Navarro, M. Botella, S. Damas, and O. Cordón. Computer vision and soft computing for automatic skull–face overlay in craniofacial superimposition. *Forensic Science International*, 245:77–86, 2014.

[CÁIN⁺15]   B. R. Campomanes-Álvarez, O. Ibáñez, F. Navarro, I. Alemán, O. Cordón, and S. Damas. Dispersion assessment in the location of facial landmarks on photographs. *International Journal of Legal Medicine*, 129(1):227–236, 2015.

[CAMFI⁺17]   C. Campomanes-Alvarez, R. Martos-Fernández, O. Ibáñez, O. Cordón, and C. Wilkinson. Modeling skull-face anatomical/morphological correspondence for craniofacial superimposition-based identification. 2017. Submitted.

[CGM⁺13]   M. Cummaudo, M. Guerzoni, L. Marasciuolo, D. Gibelli, A. Cigada, Z. Obertovà, Z. Ratnayake, P. Poppa, P. Gabriel, S. Ritz-Timme, et al. Pitfalls at the root of facial assessment on photographs: a quantitative study of accuracy in positioning facial landmarks. *Internatinal Journal of Legal Medicine*, 127(3):699–706, 2013.

[CLT⁺92]   D. S. Chai, Y. W. Lan, C. Tao, R. J. Gui, Y. C. Mu, J. H. Feng, W. D. Wang, and J. Zhu. A study on the standard for forensic anthropologic identification of skull-image superimposition. In *Bulletin du Service de Documentation Generale*, volume 79, pages 269–77. Organization Internationale de Police Criminelle (INTERPOL), 1992.

[DCI⁺11]   S. Damas, O. Cordón, O. Ibáñez, J. Santamaría, I. Alemán, M. Botella, and F. Navarro. Forensic identification by computer-aided craniofacial superimposition: a survey. *ACM Computing Surveys*, 43(4):27, 2011.

[DCS11]   S. Damas, O. Cordón, and J. Santamaría. Medical image registration using evolutionary computation: An experimental study. *IEEE Computational Intelligence Magazine*, 6:26–42, 2011.

[DCV⁺86]   V. P. Delfino, M. Colonna, E. Vacca, F. Potente, and F. Introna Jr. Computer-aided skull/face superimposition. *The American Journal of Forensic Medicine and Pathology*, 7:201–212, 1986.

[DIC17]   S. Damas, O. Ibáñez, and O. Cordón. *Handbook on craniofacial superimposition.* Springer, 2017. In press.

[DK00]   P. Diamond and P. Kloeden. Metric topology of fuzzy numbers and fuzzy analysis. In D. Dubois and H. Prade, editors, *Fundamentals of fuzzy sets. The handbooks of fuzzy sets*, volume 7 of *The Handbooks of Fuzzy Sets Series*, page 583–637. Springer US, 2000.

[DMB+11]   Q. H. Dinh, T. C. Ma, T. D. Bui, T. T. Nguyen, and D. T. Nguyen. Facial
            soft tissue thickness prediction using anthropometric distances. In N. T. Nguyen,
            B. Trawinski, and J. J. Jung, editors, *New Challenges for Intelligent Information
            and Database Systems in Computational Intelligence*, volume 351 of *Studies in
            Computational Intelligence*, pages 117–126. Springer Berlin Heidelberg, 2011.

[Dor83]   R. Dorion. Photographic superimposition. *Journal of Forensic Science*,
            28(3):724–734, 1983.

[DP83]   D. Dubois and H. Prade. On distance between fuzzy points and their use
            for pausible reasoning. In *International Conference on Systems, Man and
            Cybernetics*, pages 300–303, 1983.

[Dun61]   O. J. Dunn. Multiple comparisons among means. *The Annals of Mathematical
            Statistics*, 56:52–64, 1961.

[Dut91]   S. Dutta. Approximate spatial reasoning: integrating qualitative and quantitative
            constraints. *International Journal of Approximate Reasoning*, 5(3):307–330, 1991.

[DWK+15]   S. Damas, C. Wilkinson, T. Kahana, E. Veselovskaya, A. Abramov,
            R. Jankauskas, P.T. Jayaprakash, E. Ruiz, F. Navarro, M.I. Huete, E. Cunha,
            F. Cavalli, J. Clement, P. Leston, F. Molinero, T. Briers, F. Viegas, K. Imaizumi,
            D. Humpire, and O. Ibáñez. Study on the performance of different craniofa-
            cial superimposition approaches (ii): best practices proposal. *Forensic Science
            International*, 257:504–508, 2015.

[ES03]   A. E. Eiben and J. E. Smith. *Introduction to Evolutionary Computing*. Natural
            Computing Series. Springer-Verlag, Berlin Heidelberg, 1st edition, 2003.

[Fau93]   O. Faugeras. *Three-Dimensional Computer Vision. A Geometric Viewpoint*. The
            MIT Press, Cambridge, MA, 1st edition, 1993.

[FCPB08]   M. Fantini, F. De Crescenzio, F. Persiani, and S. Benazzi. 3d restitution,
            restoration and prototyping of a medieval damaged skull. *Rapid Prototyping
            Journal*, 14(5):318–324, 2008.

[FHS08]   T. W. Fenton, A. N. Heard, and N. J. Sauer. Skull-photo superimposition and
            border deaths: Identification through exclusion and the failure to exclude. *Journal
            of Forensic Science*, 53(1):34–40, 2008.

[Fre61]   H. Freeman. On the encoding of arbitrary geometric configurations. *Electronic
            Computers, IRE Transactions on*, (2):260–268, 1961.

[Fre75]   J. Freeman. The modelling of spatial relations. *Computer Vision Graph*, 4(2):156–
            171, 1975.

[Fri40]   M. Friedman. A comparison of alternative tests of significance for the problem of
            m rankings. *The Annals of Mathematical Statistics*, 11:86–92, 1940.

[GB37]   J. Glaister and J. C. Brash. *Medico-legal aspects of the Ruxton case*. E. and S.
            Livingstone, Edinburgh, UK, 1st edition, 1937.

[GD48]   I. Gordon and M. Drennan. Medico-legal aspects of the wolkersdorfer case. *South
            African Medical Journal*, 22:543–549, 1948.

[Geo93]   R. M. George. Anatomical and artistic guidelines for forensic facial reconstruction. In M. Y. Iscan and R. Helmer, editors, *Forensic Analysis of the Skull*, pages 215–227. Wiley Liss, New York, NY, 1993.

[GFLH09]   S. García, A. Fernández, J. Luengo, and F. Herrera. A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability. *Soft Computing*, 13:959–977, 2009.

[GODA+16]   D. Gaudio, L. Olivieri, D. De Angelis, P. Poppa, A. Galassi, and C. Cattaneo. Reliability of craniofacial superimposition using three-dimension skull model. *Journal of Forensic Sciences*, 61(1):5–11, 2016.

[GPA+06]   L. Galantucci, G. Percoco, G. Angelelli, C. López, F. Introna, C. Liuzzi, and A. De Donno. Reverse engineering techniques applied to a human skull, for cad 3d reconstruction and physical replication by rapid prototyping. *Journal of Medical Engineering and Technology*, 30(2):102–111, 2006.

[GS01]   A. K. Ghosh and P. Sinha. An economised craniofacial identification system. *Forensic Science International*, 117(1):109–119, 2001.

[GS12]   G. M. Gordon and M. Steyn. An investigation into the accuracy and reliability of skull-photo superimposition in a south african sample. *Forensic Science International*, 216:198.e1–6, 2012.

[GV86]   G. Gerla and R. Volpe. The definition of distance and diameter in fuzzy set theory. *Stutia Univ Babes–Bolyai Math*, 31:21–26, 1986.

[GW08]   R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley, Upper Saddle River, NJ, 3rd edition, 2008.

[HB97]   D. Hearn and M. P. Baker. *Computer graphics. C version*. Prentice-Hall, Upper Saddle River, NJ, 2nd edition, 1997.

[Hel86]   R. Helmer. Identifizierung der leichenuberreste das Josef Mengele [in German]. *Archives Kriminology*, 177:130–144, 1986.

[HG77]   R. Helmer and O. Grüner. Vereinfachte schädelidentifizierung nach dem superprojektionsverfahren mit hilfe einer video-anlage. *International Journal of Legal Medicine*, 80(3):183–187, 1977.

[HG09]   H. He and E. A. Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284, 2009.

[HIWK15]   M. I. Huete, O. Ibáñez, C. Wilkinson, and T. Kahana. Past, present, and future of craniofacial superimposition: Literature and international surveys. *Legal Medicine*, 17:267–278, 2015.

[HO01]   N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9:159–195, 2001.

[ICCÁ+15]   O. Ibáñez, F. Cavalli, B. R. Campomanes-Álvarez, C. Campomanes-Álvarez, A. Valsecchi, and M. I. Huete. Ground truth data generation for skull–face overlay. *International Journal of Legal Medicine*, 129(3):569–81, 2015.

[ICD12]   O. Ibáñez, O. Cordón, and S. Damas. A cooperative coevolutionary approach dealing with the skull-face overlay uncertainty in forensic identification by craniofacial superimposition. *Soft Computing*, 18:797–808, 2012.

[ICDS09]   O. Ibáñez, O. Cordón, S. Damas, and J. Santamaría. An experimental study on the applicability of evolutionary algorithms to craniofacial superimposition in forensic identification. *Information Science*, 79:3998–4028, 2009.

[ICDS11]   O. Ibáñez, O. Cordón, S. Damas, and J. Santamaría. Modeling the skull-face overlay uncertainty using fuzzy sets. *IEEE Transactions Fuzzy Systems*, 16:946–959, 2011.

[Ind09]   E. Indriati. Historical perspectives on forensic anthropology in indonesia. In S. Blau and D. H. Ubelaker, editors, *Handbook of Forensic Anthropology and Archaeology*, pages 115–125. Left Coast Press, Walnut Creek, CA, 2009.

[IS01]   K. Ikeuchi and Y. Sato. *Modeling from Reality*. The Springer International Series in Engineering and Computer Science. Springer-Verlag, Berlin, 1st edition, 2001.

[Isc81]   M. Iscan. Integral forensic anthropology. *Practicing Anthropology*, 3(4):13–30, 1981.

[Isc93]   M. Y. Iscan. Introduction to techniques for photographic comparison. In M. Y. Iscan and R. Helmer, editors, *Forensic Analysis of the Skull: Craniofacial Analysis, Reconstruction, and Identification*, pages 57–90. Wiley, New York, NY, 1993.

[IT03]   H. Imai and V. Torra. On a modeling of decision making with a twofold integral. In *EUSFLAT Conference*, pages 714–717, 2003.

[IVC+16]   O. Ibáñez, A. Valsecchi, F. Cavalli, M. I. Huete, B. R. Campomanes-Alvarez, C. Campomanes-Alvarez, R. Vicente, D. Navega, A. Ross, C. Wilkinson, et al. Study on the criteria for assessing skull-face correspondence in craniofacial superimposition. *Legal Medicine*, 23:59–70, 2016.

[IVN+15]   O. Ibáñez, R. Vicente, D. S. Navega, C. Wilkinson, P. T. Jayaprakash, M I. Huete, T. M. Briers, R. Hardiman, F. Navarro, E. Ruiz, F. Cavalli, K. Imaizumi, R. Jankauskas, E. Veselovskaya, A. Abramov, P. Lestón, F. Molinero, J. Cardoso, J. Cagdir, D. Humpire, Y. Nakanishi, A. Zeuner, A. H. Ross, D. Gaudio, and S. Damas. Study on the performance of different craniofacial superimposition approaches (i). *Forensic Science International*, 257:496–503, 2015.

[IVN+16]   O. Ibáñez, R. Vicente, D. Navega, C. Campomanes-Álvarez, C. Cattaneo, R. Jankauskas, M. I. Huete, F. Navarro, R. Hardiman, E. Ruiz, et al. MEPROCS framework for craniofacial superimposition: Validation study. *Legal Medicine*, 23:99–108, 2016.

[JL05]   A. K. Jain and S. Z. Li. *Handbook of face recognition*, volume 1. Springer, 2005.

[JSA01]   P. T. Jayaprakash, G. J. Srinivasan, and M. G. Amravaneswaran. Cranio-facial morphanalysis: a new method for enhancing reliability while identifying skulls by photo superimposition. *Forensic Science International*, 117(1):121–143, 2001.

[Kee78]   P. G. W. Keen. *Decision Support Systems: An Organizational Perspective*. Addison Wesley Pub Co, Reading, MA, 1978.

[KS10]   D. M. Kahn and R. B. Shaw. Overview of current thoughts on facial volume and aging. *Facial Plastic Surgery*, 26(05):350–355, 2010.

[KW95]   J. M. Keller and X. Wang. Comparison of spatial relation definitions in computer vision. In *Uncertainty Modeling and Analysis, 1995, and Annual Conference of the North American Fuzzy Information Processing Society. Proceedings of ISUMA-NAFIPS'95., Third International Symposium on*, pages 679–684. IEEE, 1995.

[KY95]   G. J. Klir and B. Yuan. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice-Hall, Upper Saddle River, NJ, 1st edition, 1995.

[Lan90]   Y. Lan. Research report on model tlga-213 image superimposition identification system. In *Special Issue on Criminal Technology Supplement*, page 13. The Fifth Bureau of the National Public Security Department, Beijing, China, 1990.

[Lan92]   Y. Lan. Development and current status of skull-image superimposition-methodology and instrumentation. *Forensic Science Review*, 4(2):125–136, 1992.

[LC85]   Y. Lan and D. Cai. Study on model TLGA-1 skull identification apparatus. In *Special Issue on Criminal Technology Supplement*, page 23. The Fifth Bureau of the National Public Security Department, Beijing, China, 1985.

[LC88]   Y. Lan and D. Cai. A new technology in skull identification. In R. Helmet, editor, *Advances in Skull Identification Via Video Superimposition*, page 3. Kiel, Germany, 1988.

[LC93]   Y. Lan and D. Cai. *Technical advances in skull-to-photo superimposition*. Wiley, New York, 1993.

[LK99]   W. V. Leekwijck and E. E. Kerre. Defuzzification: criteria and classification. *Fuzzy Sets and Systems*, 108(2):159–178, 1999.

[LM03]   M. Laguna and R. Martín. *Scatter Search: Methodology and Implementations*. C Kluwer Academic Publishers, Boston, MA, 2003.

[Maa89]   G. JR. Maat. The positioning and magnification of faces and skulls for photographic superimposition. *Forensic Science International*, 41(3):225–235, 1989.

[Mit97]   T. Mitchell. *Machine Learning*. McGraw Hill, New York, NY, 1st edition, 1997.

[MS56]   R. Martin and K. Saller. *Lehrbuch der Anthropologie in Systematischer Darstellung [in German]*. Gustav Fischer Verlag, Stuttgart, Germany, 1st edition, 1956.

[NFKC91]   B. A. Nickerson, P. A. Fitzhorn, S. K. Koch, and M. Charney. A methodology for near-optimal computational superimposition of two-dimensional digital facial photographs and three-dimensional cranial surface meshes. *Journal of Forensic Science*, 36:480–500, 1991.

[NM11]   F. J. Navarro-Merino. *Superposición Craneofacial para Identificación Humana. Un Estudio en Población Mediterránea [in Spanish]*. PhD thesis, University of Granada, Spain, 2011.

[PCK06]   H. K. Park, J. W. Chung, and H. S. Kho. Use of hand-held laser scanning in the assessment of craniometry. *Forensic Science International*, 160:200–206, 2006.

[PK93]   C. P. Pappis and N. I. Karacapilidis. A comparative assessment of measures of similarity of fuzzy values. *Fuzzy Sets and Systems*, 56(2):171–174, 1993.

[PU02]   M. Prokopec and D. H. Ubelaker. Reconstructing the shape of the nose according to the skull. *Forensic Science Communications*, 4(1), 2002.

[PVP⁺93]   D. V. Pesce, E. Vacca, F. Potente, T. Lettini, and M. Colonna. *Shape analytical morphometry in computer-aided skull identification via video superimposition*. Iscan M. Y., Helmer R. P. Forensic analysis of the skull: craniofacial analysis, reconstruction, and identification. New York: Wiley-Liss, 1993.

[RM86]   D. E. Rumelhart and D. McClelland. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. MIT Press, Cambridge, MA, 1st edition, 1986.

[RMA06]   A. Ricci, G. L. Marella, and M. A. Apostol. A new experimental approach to computer-aided face/skull identification in forensic anthropology. *The American Journal of Forensic Medicine and Pathology*, 27(1):46–49, 2006.

[Ros85]   A. Rosenfeld. Distances between fuzzy sets. *Pattern Recognition Letters*, 3(4):229–233, 1985.

[RPE⁺95]   J. Richtsmeier, C. Paik, P. Elfert, T. Cole, and F. Dahlman. Precision, repeatability and validation of the localization of cranial landmarks using computed tomography scans. *Cleft Palate-Craniofacial Journal*, 32:217–227, 1995.

[SCD07a]   J. Santamaría, O. Cordón, and S. Damas. A scatter search-based technique for pairwise 3d range image registration in forensic anthropology. *Soft Computing*, 11(9):819–828, 2007.

[SCD⁺07b] J. Santamaría, O. Cordón, S. Damas, I. Alemán, and M. Botella. Evolutionary approaches for automatic 3d modeling of skulls in forensic identification. In *Applications of Evolutionary Computing*, number 4448 in Lecture Notes in Computer Science, page 415–422. Springer, Berlin, Germany, 2007.

[SCD⁺09]   J. Santamaría, O. Cordón, S. Damas, J. M. García-Torres, and A. Quirin. Performance evaluation of memetic approaches in 3D reconstruction of forensic objects. *Soft Computing*, 13:883–904, 2009.

[SCD10]   J. Santamaría, O. Cordón, and S. Damas. A comparative study of state-of-the-art evolutionary image registration methods for 3D modeling. *Computer Vision and Image Understanding*, 115:1340–1354, 2010.

[Sek71]   P. C. Sekharan. A revised superimposition technique for identification of the individual from the skull and photograph. *Journal of Criminal Law and Criminology*, pages 107–113, 1971.

[Sek93]   P. C. Sekharan. Positioning the skull for superimposition. In M. Y. Iscan and R. Helmer, editors, *Forensic Analysis of the Skull*, pages 105–118. Wiley Liss, New York, NY, 1993.

[Sen62]   N. Sen. Identification by superimposed photographs. *International Criminal Police Review*, 162:284–286, 1962.

[SHB14]   M. Sonka, V. Hlavac, and R. Boyle. *Image processing, analysis, and machine vision*. Cengage Learning, 2014.

[SLW+09]   S. Singare, Q. Lian, W. P. Wang, J. Wang, Y. Liu, D. Li, and B. Lu. Rapid prototyping assisted surgery planning and custom implant design. *Rapid Prototyping Journal*, 15(1):19–23, 2009.

[SM08]   C. N. Stephan and S. J. Murphy. Mouth width prediction in craniofacial identification: cadaver tests of four recent methods, including two techniques for edentulous skulls. *Journal of Forensic Odonto-Stomatology*, 27(1):2–7, 2008.

[SN02]   B. Scully and P. Nambiar. Determining the validity of Furue's method of craniofacial superimposition for identification. *Malaysian Journal of Computer Science*, 9:17–22, 2002.

[SS08a]   C. N. Stephan and E. K. Simpson. Facial soft tissue depths in craniofacial identification (part i): an analytical review of the published adult data. *Journal of Forensic Science*, 53:1257–1272, 2008.

[SS08b]   C. N. Stephan and E. K. Simpson. Facial soft tissue depths in craniofacial identification (part ii): an analytical review of the published sub-adult data. *Journal of Forensic Science*, 53:1273–1279, 2008.

[Ste03]   C. N. Stephan. Facial approximation: An evaluation of mouth-width determination. *American Journal of Physical Anthropology*, 121(1):48–57, 2003.

[Ste09]   C. N. Stephan. Craniofacial identification: Techniques of facial approximation and craniofacial superimposition. In S. Blau and D. H. Ubelaker, editors, *Handbook of Forensic Anthropology and Archaeology*, volume 25, pages 304–321. Left Coast Press, Walnut Creek, 2009.

[Sug74]   M. Sugeno. *Theory of fuzzy integrals and its applications*. Tokyo Institute of Technology, 1974.

[Sug77]   Michio Sugeno. Fuzzy measures and fuzzy integrals: a survey. *Fuzzy automata and decision processes*, 78(33):89–102, 1977.

[SY93]   S. Seta and M. Yoshino. *A combined apparatus for photographic and video superimposition*, volume 161. Wiley-Liss, New York, 1993.

[Tao86]   C. Tao. Report on computer programming for model TLGA-1 skull identification. In *Special Issue on Criminal Technology Supplement*, page 41. The Fifth Bureau of the National Public Security Department, Beijing, China, 1986.

[TB98]   J. Taylor and K. Brown. Superimposition techniques. In J. Clement and D. Ranson, editors, *Craniofacial Identification in Forensic Medicine*, pages 151–164. Arnold, London, UK, 1998.

[TJB05]   C. M. Takemura, R. Cesar Jr, and I. Bloch. Fuzzy modeling and evaluation of the spatial relation "Along". In *Progress in Pattern Recognition, Image Analysis and Applications*, pages 837–848. Springer, 2005.

[TK90]   H. Tahani and J. M. Keller. Information fusion in computer vision using the fuzzy integral. *IEEE Transactions on Systems, Man and Cybernetics*, 20(3):733–741, 1990.

[Ube00]   D. H. Ubelaker. A history of smithsonian–fbi collaboration in forensic anthropology, especially in regard to facial imagery. *Forensic Science Communications*, 2(4), 2000.

[UBO92]   D. H. Ubelaker, E. Bubniak, and G. O'Donnell. Computer-assisted photographic superimposition. *Journal of Forensic Science*, 37(3):750–762, 1992.

[VC07]   L. Valencia-Caballero. *Metodología para elaborar reconstrucciones faciales empleando gráficos computarizados tridimensionales [in Spanish]*. PhD thesis, University of Granada, Spain, 2007.

[Vel01]   R. C. Veltkamp. Shape matching: Similarity measures and algorithms. In *Shape Modeling and Applications, SMI 2001 International Conference on.*, pages 188–197. IEEE, 2001.

[WM03]   C. M. Wilkinson and S. A. Mautner. Measurement of eyeball protrusion and its application in facial reconstruction. *Journal of Forensic Science*, 48(1):12–16, 2003.

[WR12]   C. Wilkinson and C. Rynn. *Craniofacial identification*. Cambridge University Press, 2012.

[Wri92]   S. P. Wright. Adjusted p-values for simultaneous inference. *Biometrics*, 48:1005–1013, 1992.

[YIMS95]   M. Yoshino, K. Imaizumi, S. Miyasaka, and S. Seta. Evaluation of anatomical consistency in craniofacial superimposition images. *Forensic Science International*, 74(1):125–134, 1995.

[YKR+08]   M. Yang, K. Kpalma, J. Ronsin, et al. A survey of shape feature extraction techniques. *Pattern Recognition*, pages 43–90, 2008.

[YMK+97]   M. Yoshino, H. Matsuda, S. Kubota, K. Imaizumi, S. Miyasaka, and S. Seta. Computer-assisted skull identification system using video superimposition. *Forensic Science International*, 90(3):231–244, 1997.

[Yos12]   M. Yoshino. Craniofacial superimposition. In C Wilkinson and C Rynn, editors, *Craniofacial Identification*, pages 238–253. University Press, Cambridge, 2012.

[YS00]   M. Yoshino and S. Seta. Skull-photo superimposition. *Encyclopedia of Forensic Sciences*, 2:807–815, 2000.

[Zad65]   L. A. Zadeh. Fuzzy sets. *Information Control*, 8(3):338–353, 1965.

[Zad94]   L. A. Zadeh. Soft computing and fuzzy logic. *IEEE Software*, 11:48–56, 1994.

[Zad96]   L. A. Zadeh. *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lofti A. Zadeh*, volume 6 of *Advances in fuzzy systems : applications and theory*. World Scientific, River Edge, NJ, 1996.

[Zad01]   L. A. Zadeh. Applied soft computing. *Applied Soft Computing*, 1:1–2, 2001.

[ZCB87]   R. Zwick, E. Carlstein, and D. V. Budescu. Measures of similarity among fuzzy concepts: A comparative analysis. *International Journal of Approximate Reasoning*, 1(2):221–242, 1987.

[ZF03]   B. Zitovà and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21:977–1000, 2003.

[ZL04]   D. Zhang and G. Lu. Review of shape representation and description techniques. *Pattern Recognition*, 37(1):1–19, 2004.