Chapter 7
Experimental Study of Craniofacial Superimposition Methodologies, Tools, and Criteria

7.1 Introduction

The objective of this chapter is to analyze the reliability of different CFS methodologies and the corresponding technical approaches to the CFS identification technique. Moreover, we aim to examine the subjectivity and discriminative power of the different criteria (detailed in Sect. 7.2) for assessing the skull-face correspondence either proposed in the literature or by any of the MEPROCS partners.

This novel study is expected to provide important insights to better understand: (1) which are the most convenient characteristics of every method included in this study, (2) which are the most and less discriminative criteria; and (3) which criteria are dependent more on the expert and which criteria are more independent, that is, less subjective. The two latter points could give an idea of how many and which criteria are needed to reach a reliable conclusion. Those criteria that are determined to be more discriminatory could be later included as a recommended standard for CFS.

7.2 Study on the Performance of Different Craniofacial Superimposition Approaches

7.2.1 Experimental Study

Each participant was requested to tackle each of the provided cases using the typical protocol that they would follow at their institutions. The participants were requested to fill an identification form with description of the protocol/methodology (i.e., software, equipment, orientation process, landmarks, assessment criteria) employed. For each case, a final identification decision (either positive or negative) should be

Table 7.1 Summary of the characteristics of the datasets employed for the study

Case	# Skulls	# Photographs	# Candidates	# SFO'S	# Positive cases Male set	Female set
N1	1	3	3	3	1	0
N2	1	3	3	3	1	1
N3	1	3	3	3	0	1
N4	1	3	3	3	0	0
N5	4	4	4	16	2	2
N6[a]	1	2	1	2	0	0
N7[a]	1	2	1	2	2(1)[a]	2(1)[a]
TOTAL	7	20	18	32(30)[a]	6(5)[a]	6(5)[a]

[a]Although cases 6 and 7 are composed of two photographs, that is, they involve two SFOs, these two cases result in just one identification result, considering the matching of the skull over the two photographs

reported along with the rationale supporting the decision and at least one image illustrating the overlay/superimposition outcome.

The dataset used in this reliability test consisted of two sets, divided by sex, of seven CFS case studies each of them. These 14 CFS cases involve a total number of 60 SFO problems as given in Table 7.1. The dataset was collected at the University of Tennessee after obtaining informed consent from the responsible party for the deceased, and provided to the MEPROCS project as data share protocol established through the University of Dundee.

The dataset consisted generically of a set of ante-mortem photos, photos of the skull (with scales), and a set of 3D models of the skull acquired by laser scanning technology (Fastscan Polhemus Scorpion scanner). Physical 1:1 replicas of the skull 3D models were provided to those participants performing video-superimposition. Each set of case studies had the following structure: cases 1–4 mimic a scenario with one skull and three possible candidates, where only one ante-mortem photo of each candidate is available. In case 5, a more complex scenario is simulated, including four skulls and four possible candidates, with only one available ante-mortem photo of each candidate. In cases 6 and 7, the scenario simulated includes one skull and only one possible candidate, with several photos of the candidate available for analysis (see Table 7.1 for the case studies detailed explanation).

The performance of each participant was measured by computing true-positive, false-positive, true- and false-negative rates, and overall accuracy. All indicators were calculated for each sex and all case studies pooled together. Experience and familiarity with craniofacial identification techniques was also taken into account and level of experience of the participants was classified according to the following scheme:

- No previous experience and no CFS-related training.
- No previous experience but CFS-related training.
- Short previous research experience and CFS-related training.
- Moderate previous experience with CFS real cases and CFS-related training.
- Broad experience with CFS real cases.

The study was carried out by 26 participants from the following institutions: University of Granada (Spain), University of Dundee (Scotland), Legal Medicine and Forensic Sciences Institute (Peru), North Carolina State University (USA), Complutense University of Madrid (Spain), University of Melbourne (Australia), Azienda Ospadaliera-Universitaria di Trieste (Italy), Russian Academy of Sciences (Russia), Portuguese Judiciary Police (Portugal), Moscow Region State Bureau of Forensic Examination (Russia), Spanish Civil Guard (Spain), Turkish Council of Forensic Medicine (Turkey), National Research Institute of Police Science (Japan), University of Milan (Italy), South African Police Service (South Africa), University of Vilnius (Lithuania), and University Sains Malaysia (Malaysia). In Table 7.2 all the participants (numbered from 1 to 26) are listed in the study with the corresponding level of experience. Since not all the participants completed the whole study, information of the dataset(s) tackled by each of them is provided as well.

Tables 7.3, 7.4, 7.5, and 7.6 summarize the methodologies employed by the participants grouped by the technological approach followed. They were classified following the taxonomy given in The Scientific Working Group for Forensic Anthropology (2012b), that is, computer-aided semi-automatic 3D-2D superimposition (Table 7.3), computer-aided manual 3D-2D superimposition (Table 7.4), computer-aided manual video superimposition (Table 7.5), and computer-aided manual photo superimposition (Table 7.6). The first column also indicates both the type of dataset used and the global performance. The datasets are either male, female, or both. The global performance of the participant methodology refers to the percentage of correct decisions. Significant details of each of them are briefly explained according to software and equipment employed, how the SFO process is tackled, and the kind of skull-face relationship assessment made (decision making).

7.2.2   Results

A total number of 1152 CFS problems have been tackled within this study. While previous Tables 7.3, 7.4, 7.5, and 7.6 reported on the global performance (correct decisions) of each participant-methodology, the following three tables report the results obtained by each participant, considering separately the two different datasets independently (Tables 7.7 and 7.8, male and female respectively) and both together (Table 7.9). Detailed performance indicators such as true positive (TP), false positive (FP), true negative (TN), and false negative (FN) are given in any case.

Considering only the global performance, participants P2, P3, and P4 achieved the higher rates surpassing 90.00% up to 94.29% achieved by P4 (Tables 7.3 and 7.4). They share a similar SFO approach (computer-aided 3D-2D), although P2 made use of a semiautomatic software. During the decision-making stage, they all employed the morphological criteria in Wilkinson (2004), and two of them also analyzed the criteria introduced in Austin-Smith and Maples (1994) and Yoshino (2012). However, only P2 tackled both female and male datasets. Similar

Table 7.2 Participants of the study, their experience related to CFS, and datasets tackled

Participant number		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23	P24	P25	P26
CFS experience		4	3	3	3	5	4	2	2	2	2	2	2	2	3	2	4	4	1	4	1	4	5	3	5	5	5
Submitted results	Males	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Females	✓	✓	✓	✗	✓	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✗	✓	✓

Table 7.3 Summarization of computer-aided semiautomatic 3D-2D superimposition (CAs3DS) approaches that participated in the study

Participant number and global performance	Software and equipment	Skull-face overlay	Decision making
P2 Both datasets 91.43%	Face2Skull™	Automatic SFO by matching cranial and facial landmarks (manually located). Manual refinement of the overlay when needed.	Followed guideline and criteria proposed in Austin-Smith and Maples (1994), Yoshino (2012), Wilkinson (2004).
P19 Both datasets 83.01%	3D-TADD	Used 3D-TADDsoftware to scale the photograph and position the 3D model.	Morphological and landmark assessment. Anatomical points and contours of the head, the middle and lower part of the face, eyebrows, ears lobes, the line between the lips, the wings of the nose and supramental furrow are automatically detected. Manual refinement of points and contours when needed.

performances were obtained by P26 (88.30%) and P18 (87.69%). While the first followed video superimposition approach, the second employed computer 3D-2D software. Again, a morphological approach was the key aspect, leading their skull-face relationship assessment. From these rates, participants' performance decreases almost linearly until the worst results by P23, who based both the SFO and decision making in a landmark comparison.

Looking deeply at the individual performance, it is quite obvious that higher rates of true negatives were achieved in comparison to true positives. Just focusing on those participants who carried out the study over both male and females datasets (Table 7.9), we observe that four of them achieved true negative rates equal or higher than 90.00% (P2, P14, P18, P26). However, the same four participants achieved 80.00%, 50.00%, 62.50%, and 66.67% of true positive, respectively. According to the average behavior considering the two datasets, the mean true positive rate is 52.63%, while the mean true negative rate is 84.20%. Consequently, the false-positive rate is significantly lower than the false-negative rate. It is important to remark that the number of negative cases (50) is five times the number of positive cases (10).

Table 7.10 reports performance indicator of the different participants – methodologies grouped by the level of experience of the participant. There are not significant differences related to the level of experience of the participants. There is not a correlation between the performance and the level of experience of the practitioners.

Table 7.4 Summarization of computer-aided manual 3D-2D superimposition (CAm3DS) approaches that participated in the study

Participant number and global performance	Software and equipment	Skull-face overlay	Anatomical consistency assessment criteria	Observation
P4 Male dataset 94.29%	Freeform Modelling Plus Phantom® Desktop™ Haptic Device Adobe Photoshop™ Wacom tablet	The cranium was orientated first in a trial and error manner, by means of moving and rotating the stylus of the haptic device. Starting points for the alignment were eyes/orbits in conjunction with either the maxillary teeth, if visible, or the nasal bridge. Follow "fuzzy landmarks" and morphology for orientation. The mandible was orientated in compliance with the facial expression and the limits of a physiologically natural position.	The resulting images were then imported into Adobe Photoshop and outline tracings of the skull and face were produced using Wacom tablet. Morphological assessment based on Austin-Smith and Maples (1994), Yoshino (2012), Wilkinson (2004).	The software has two variables to "play" with the camera projection.
P3 Female dataset 91.43%	Freeform Modelling Plus Phantom® Desktop™ Haptic Device Adobe Photoshop™	Manual orientation with software and haptic device. Anatomical criteria of morphological assessment taken into account during this process.	Morphological assessment based on (Wilkinson 2006).	Haptic device used for moving the mandible.
P18 Both datasets 87.69%	Blender Adobe Photoshop™	Manual orientation using Blender and Photoshop.	Morphological assessment.	
P14 Both datasets 85.71%	Geomagic Studio10 (3D software) Adobe Photoshop™ CS6	Geomagic Studio used to articulate cranium and mandible. Orientation of the skull by trial and	Landmark and morphological consistency analysis.	Consider first biological profile, asymmetries, and pathologies. Give special

Table 7.4   (continued)

Participant number and global performance	Software and equipment	Skull-face overlay	Anatomical consistency assessment criteria	Observation
		error manipulation of the 3D skull model using Photoshop		consideration to teeth comparison.
P20 Male dataset 81.66%	Autodesk 3D Studio MAX Photoshop™ CS6	Orientation (yaw, pitch and roll) setting in the same plane porion, tragus, Whitnall's tubercle, and ectocanthion (Fenton et al. 2008).	Morphological (outlines) and landmark matching assessment.	Asymmetries are key signs for individualization.
P21 Male dataset 81.82%	Adobe Photoshop™	Both 3D model and AM photo oriented and resized in Photoshop. Separate use of cranium and mandible.	Morphological and landmark assessment.	
P5 Both datasets 80.88%	Adobe Photoshop ™ CS3. 3D Max Studio 2011	Orientation and overlay achieved using software tools. Process guided by recommendations proposed in Lan and Cai (1993). Use of morphometric planes, and landmarks from Martin and Saller (1966).	Recommendations by Lan and Cai (1993). Morphological assessment.	
P1 Both datasets 79.10%	RapidForm 2006 Adobe Photoshop™	3D skull model oriented by simultaneous view of the model and the AM photo. Resize of the image according to cranial and facial landmarks.	Followed criteria described in Navarro-Merino (2011).	
P17 Both	Craniofacial correspondence analysis program	The program represents the algorithm of	The program of craniofacial correspondence based on the data on soft facial tissue thickness and on the analysis of	

(continued)

Table 7.4 (continued)

Participant number and global performance	Software and equipment	Skull-face overlay	Anatomical consistency assessment criteria	Observation
datasets 77.19%	(in-house software)	transformations from dimensional and descriptive traits of a skull into the corresponding parameters of the face.	relationships between facial and skull features, correlative and regression analysis of the facial and skull features, as well as a descriptive analysis of indexes and variations of categorical and quantitative characteristics.	
P15 Both datasets 76.81%	Geomagic Studio software 3D Rugle	The 3Dscan is rotated manually (using 3D Rugle) through all possible axes (X, YandZ) to Frankfort horizontal plane. The orientation of the skull is then electronically realigned so that the anatomical position becomes the starting position.	Morphological and landmark assessment.	Geomagic is also used to complete the mandible articulation. Each photograph is cropped and resized to maximize the dimensions of the face in the frame.
P22 Both datasets 71.43%	Claytools Geomagic, Inc. 3D Rugle Adobe Photoshop™ Elements	The obj 3D data both of "cranium" and "mandible" were imported to a 3D modeling tool, Claytools (Geomagic, Inc.). A craniofacial superimposition was performed with "3D Rugle" software. These are imported to an Adobe Photoshop Elements software and superimposed again.	Morphological and landmark assessment (measure distances between landmarks).	A perspective effect can be virtually applied to the 3D skull data using this software (3D Rugle).
P16 Both datasets 67.24%	Rhinoceros© or Autodesk 3D Studio© software Adobe Photoshop™ software	Right size is established from inter-Whitnall's malar tubercular distance and/or glabella–gnathion	Morphological and landmark assessment.	Generally do not use 3D laser scan, because participants have access to a CT-scan with 0.5 mm isotropic

(continued)

Table 7.4   (continued)

Participant number and global performance	Software and equipment	Skull-face overlay	Anatomical consistency assessment criteria	Observation
		distance. Orientation is established matching Whitnall's malar tubercles–porion, where possible. Superimposition based on shapes. Landmarks help guiding the process		voxel that helps in obtaining a very anatomically detailed 3D image.
P23 Male dataset 55.88%	Geomagic Studio software 3D Studio Max software VAM software	Landmark-based orientation following Gordon and Steyn (2012).	Landmark-based consistency analysis following Gordon and Steyn (2012).	0, 1, or 2 disagreements of landmarks are considered a positive match. 3 or more a considered a negative match.

While the participants who achieve higher rates of correct decisions are in category 3 (84.68%), those generating lower results are, surprisingly, category four, together with category two grouped participants with the lowest performance (75.00% and 74.32%, respectively).

Finally, Table 7.11 depicts the overall accuracy according to the technological approach used by each participant. In overall, the approach followed by Participants 2, 3 (CAm3DS approach) and 24, 25, and 26 (CAmVS approach) is the most accurate (88.49% and 84.56%, respectively). These technological approaches represent the past and the future of the CFS technological development.

7.2.2.1   Set of Criteria for Assessing the Skull-Face Overlay Relationship

With all the data generated, some of the most representative experts in craniofacial identification joined in a discussion intended to identify and agree on the most important issues that have to be considered to properly employ the CFS technique. Tables 7.12, 7.13, 7.14, and 7.15 depict the identification of a set of common criteria for assessing the skull-face correspondence.

Table 7.5 Summarization of computer-aided manual video superimposition (CAmVS) approaches that participated in the study

Participant number and global performance	Software and equipment	Skull-face overlay	Decision making
P24 Mala dataset 80.00%	Digital video mixer, digital video camera, and screen monitor	The orientation was done by aligning: Whitnall's tubercle with ectocanthions; external auditory meatus with the tragus; vertical midline of the face using the glabella, nasion, anterior nasal spine, prosthion, menton, and pogonion. The size was adjusted taking soft tissue depth into consideration.	First, the correlation of the shape characteristics between facial skeleton and face in the photograph was compared. Then, morphological assessment was performed.
P25 Both datasets 83.33%	Digital video mixer, digital video camera, and screen monitor AVerMedia TV Series Product	Skull: (1) orientated on an own-made support enabling rotation in three planes; (2) zoomed in/out manually by video camera to obtain corresponding size; (3) rotated manually to obtain best position according to vertical (glabella-nasion-prosthion), horizontal (ectocanthion-ectocanthion and porion-porion, when available) and anteroposterior (prosthion-porion) planes.	Morphological assessment of landmarks and outlines.
P26 Both datasets 88.13%	Adobe Photoshop™ CS3 Video system (Wilkinson 2004)	Follow the procedure described in Jayaprakash et al. (2001)	Asses traits indicated in Jayaprakash et al. (2001).

Table 7.6 Summarization of computer-aided manual photo superimposition (CAm3PS) approaches that participated in the study

Participant number and global performance	Software and equipment	Skull-face overlay	Decision making	Observation
P6–P13 68.57– 82.35% Male dataset	Adobe Photoshop™ or GNU GIMP. Dennis Slice's PhotoMatch	Photoshop to resize images. Then, PhotoMatch to manipulate transparency, scaling, and rotation to match as many points as possible.	Special signs: crowding, broken nasal bones, asymmetry, etc.	Do not use the 3D data but just the PM skull photos. Limited orientation and alignment capabilities.

Eight different participants but all of them following the same methodology

Table 7.7 Performance of CFS methodologies on male dataset

Participant N.	N	N of P	N of P given	TP (%)	N of N	N of N given	TN (%)	N of N	N of P given	FP (%)	N of P	N of N given	FN (%)
Baseline	30	5	5	100.00	25	25	100.00	25	0	0.00	5	0	0.00
P1	30	5	3	60.00	25	21	84.00	25	4	16.00	5	2	40.00
P2	30	5	4	80.00	25	24	96.00	25	1	4.00	5	1	20.00
P4	30	5	4	80.00	25	24	96.00	25	1	4.00	5	1	20.00
P5	30	5	4	80.00	25	23	92.00	25	2	8.00	5	1	20.00
P6	4	2	1	50.00	2	2	100.00	2	0	0.00	2	1	50.00
P7	30	5	2	40.00	25	21	84.00	25	4	16.00	5	3	60.00
P8	30	5	1	20.00	25	21	84.00	25	4	16.00	5	4	80.00
P9	29	4	1	25.00	25	23	92.00	25	2	8.00	4	3	75.00
P10	30	5	1	20.00	25	20	80.00	25	5	20.00	5	4	80.00
P11	29	5	3	60.00	24	17	70.83	24	7	29.17	5	2	40.00
P12	30	5	1	20.00	25	21	84.00	25	4	16.00	5	4	80.00
P13	30	5	2	40.00	25	20	80.00	25	5	20.00	5	3	60.00
P14	30	5	4	80.00	25	22	88.00	25	3	12.00	5	1	20.00
P15	30	5	3	60.00	25	21	84.00	25	4	16.00	5	2	40.00
P16	20	3	0	0.00	17	11	64.71	17	6	35.29	3	3	100.00
P17	20	5	4	80.00	15	11	73.33	15	4	26.67	5	1	20.00
P18	29	4	3	75.00	25	23	92.00	25	2	8.00	4	1	25.00
P19	26	3	2	66.67	23	18	78.26	23	5	21.74	3	1	33.33
P20	30	5	3	60.00	25	21	84.00	25	4	16.00	5	2	40.00
P21	28	5	4	80.00	23	18	78.26	23	5	21.74	5	1	20.00
P22	30	5	5	100.00	25	20	80.00	25	5	20.00	5	0	0.00
P23	29	4	2	50.00	25	14	56.00	25	11	44.00	4	2	50.00
P24	30	5	0	0.00	25	24	96.00	25	1	4.00	5	5	100.00
P26	29	4	3	75.00	25	22	88.00	25	3	12.00	4	1	25.00
P25	30	5	3	60.00	25	20	80.00	25	5	20.00	5	2	40.00

Table 7.8 Performance of CFS methodologies on female dataset

Participant N.	N	N of P	N of P given	TP (%)	N of N	N of N given	TN (%)	N of N	N of P given	FP (%)	N of P	N of N given	FN (%)
Baseline	30	5	5	100.00	25	25	100.00	25	0	0.00	5	0	0.00
P1	29	5	2	40.00	24	20	83.33	24	4	16.67	5	3	60.00
P2	30	5	4	80.00	25	24	96.00	25	1	4.00	5	1	20.00
P3	30	5	3	60.00	25	24	96.00	25	1	4.00	5	2	40.00
P5	28	4	1	25.00	24	18	75.00	24	6	25.00	4	3	75.00
P14	30	5	1	20.00	25	23	92.00	25	2	8.00	5	4	80.00
P15	29	5	1	20.00	24	20	83.33	24	4	16.67	5	4	80.00
P16	28	5	3	60.00	23	19	82.61	23	4	17.39	5	2	40.00
P17	30	5	1	20.00	25	22	88.00	25	3	12.00	5	4	80.00
P18	29	4	2	50.00	25	22	88.00	25	3	12.00	4	2	50.00
P19	27	4	3	75.00	23	17	73.91	23	6	26.09	4	1	25.00
P20	30	5	3	60.00	25	22	88.00	25	3	12.00	5	2	40.00
P22	30	5	3	60.00	25	18	72.00	25	7	28.00	5	2	40.00
P26	30	5	3	60.00	25	24	96.00	25	1	4.00	5	2	40.00
P25	30	5	4	80.00	25	23	92.00	25	2	8.00	5	1	20.00

Table 7.9 Performance of CFS methodologies on male and female dataset

Participant N.	N	N of P	N of P given	TP (%)	N of N	N of N given	TN (%)	N of N	N of P given	FP (%)	N of P	N of N given	FN (%)
Baseline	30	5	5	100.00	25	25	100.00	25	0	0.00	5	0	0.00
P1	59	10	5	50.00	49	41	83.67	49	8	16.33	10	5	50.00
P2	60	10	8	80.00	50	48	96.00	50	2	4.00	10	2	20.00
P5	58	9	5	55.56	49	41	83.67	49	8	16.33	9	4	44.44
P14	60	10	5	50.00	50	45	90.00	50	5	10.00	10	5	50.00
P15	59	10	4	40.00	49	41	83.67	49	8	16.33	10	6	60.00
P16	48	8	3	37.50	40	30	75.00	40	10	25.00	8	5	62.50
P17	50	10	5	50.00	40	33	82.50	40	7	17.50	10	5	50.00
P18	58	8	5	62.50	50	45	90.00	50	5	10.00	8	3	37.50
P19	53	7	4	57.14	46	40	86.96	46	6	13.04	7	3	42.86
P20	60	10	6	60.00	50	43	86.00	50	7	14.00	10	4	40.00
P22	60	10	8	80.00	50	38	76.00	50	12	24.00	10	2	20.00
P25	60	10	7	70.00	50	43	86.00	50	7	14.00	10	3	30.00
P26	59	9	6	66.67	50	46	92.00	50	4	8.00	9	3	33.33

Table 7.10 Overall accuracy of CFS grouped by level of experience of the participant

Participants	Sample	Correct decisions	Ground truth	Decision Positive	Decision Negative	Decision (%) Positive	Decision (%) Negative
All	Global	78.99%	Positive	100	90	52.63	47.37
			Negative	152	810	15.80	84.20
1	Global	81.88%	Positive	14	9	60.87	39.13
			Negative	16	99	13.91	86.09
2	Global	74.32%	Positive	15	33	31.25	68.75
			Negative	43	205	17.34	82.66
3	Global	84.68%	Positive	22	12	64.71	35.29
			Negative	20	155	11.43	88.57
4	Global	75.00%	Positive	18	14	56.25	43.75
			Negative	34	126	21.25	78.75
5	Global	80.75%	Positive	31	22	58.49	41.51
			Negative	39	225	14.77	85.23

Table 7.11 Overall accuracy of CFS grouped by technological approach employed

Participants N.	Tech. approach	Correct decision	Ground truth	Decision Positive	Decision Negative	Decision (%) Positive	Decision (%) Negative
P2, P19	CAs3DS	88.49%	Positive	12	5	70.59	29.41
			Negative	8	88	8.33	91.67
P1, P3, P4, P5, P14, P15, P16, P17, P18, P20, P21, P22, P23	CAm3DS	78.90%	Positive	62	47	56.88	43.12
			Negative	92	458	16.73	83.27
P24, P25, P26	CAmVS	84.56%	Positive	13	11	54.17	45.83
			Negative	12	133	8.28	91.72
P6, P7, P8, P9, P10, P11, P12, P13	CAmPS	74.06%	Positive	12	24	33.33	66.67
			Negative	31	145	17.61	82.39

7.3   Study on the Criteria Assessing Skull-Face Correspondence in Craniofacial Superimposition

The MEPROCS consortium designed the current study, which aims to analyze the subjectivity and discriminative power of the different criteria (defined in the previous section of this chapter) for assessing the skull-face correspondence either proposed in the literature or by any of the MEPROCS partners.

7.3.1   *Experimental Study*

The dataset used in this study consisted of 18 different CFS problems, some of them composed of more than one image of the same subject, 24 SFOs in total.

Table 7.12 Marking lines used to analyze anatomical consistency

No.	Criteria	View	Do you employ it?
Group 1	Superimposition of the following marking lines (face–skull):	Frontal (F)	–
1.1	excanthion-excanthion (A) – ectocanthion-ectoconchion (A′) Ectocanthion line	F	
1.2	glabella-gnathion (B) – glabella-gnathion (B′) Frontal central line	F	
1.3	superciliary-superciliary (C) – superciliary-superciliary (C′) Supraciliary line	F	
1.4	horizontal line at subnasal (D) – horizontal line at nasospinal (D′) Subnasal line	F	
1.5	cheilion-cheilion (E) – occusal line/horizontal line at stomion (E′) Cheilion line	F	
1.6	horizontal line at gnathion (H) – horizontal line at gnathion (H′) Gnathion line	F	
1.7	endocanthion-cheilion (G) – entocanthion-caninion (G′) [right] Entocanthion vertical line	F	
1.8	endocanthion-cheilion (G) – entocanthion-caninion (G′) [left] Entocanthion vertical line	F	

Skull 3D models were obtained from patients whose head has been scanned with a cone beam computed tomography (CBCT).

The skull 3D models employed suffer from two different problems. Firstly, we only have part of the whole skull, from the jaw to the upper orbits, without including parietal, occipital, and part of the temporal areas. Secondly, the 3D model is, to a greater or lesser extent, noisy and may not be accurately represented. All those problems have the same origin: the use of CBCTs instead of CTs. High-resolution CTs together with photographs of the patient/volunteer were not accessible. However, one benefit of the CBCT data is that the volunteer was upright rather than supine, as commonly recorded for CT scanning.

Frontal and lateral photographs were taken of the same patients to create a set of positive cases, while other people with similar facial geometry were photographed in order to compose a set of negative cases. Nine of eighteen cases were positives and the other nine were negatives. Twelve of the photographs were lateral and twelve were frontal, half of them belonging to positive cases and the other half to negative cases.

The participants were provided the same 24 SFOs as a single image with four different layers: facial photograph with and without landmarks and skull projection with and without landmarks.

Table 7.13 Landmarks used to evaluate soft tissue thickness

No.	Criteria	View	Do you employ it?
Group 2	Overall consistency of the facial outline and facial soft tissue thickness at the following pair of homologous points (skull-face):	Frontal (F)/ Lateral-Oblique (L-O)	–
2.1	gnathion – menton	F	
2.2	gnathion – menton	L-O	
2.3	nasion – nasion	L-O	
2.4	glabella – glabella	L-O	
2.5	subespinale – subnasale	L-O	
2.6	pogonion – pogonion	L-O	
2.7	rhinion – rhinion	L-O	
2.8	gonion – gonion	F	
2.9	zygion – zygion	F	
2.10	trichion – trichion	L-O	
2.11	The minimal tissue thickness all along the contour is considered from the point of view of its symmetry by side for evaluating the match as acceptable.		
2.12	Ant lacrimal crest – medial canthus	F/O	
2.13	Prosthion – supra-labiale	F/L-O	
2.14	Alare – Alare	F/L-O	
2.15	Gonion – gonion	L	
2.16	Zygomaxilare – malare	L	
2.17	Whitnall's tubercle – lateral canthus	F/L-O	
2.18	occlusion mid-incisors – stomion	F/L-O	
2.19	porion – tragion	L-O	
2.20	crista conchalis – supra-alare	F/L-O	
2.21	intercanine distance (75%) – chelion	F	
2.22	eyeball position – pupilare	F	
2.23	supraorbitale – sag eyebrow	F	
2.24	two tangents nasal – pronasale	F/L-O	
2.25	1st premolar/canine radiating line – chelion	F/L-O	
2.26	mastoidale – subaurale	L-O	
2.27	infraorbital foramen – chelion	F/L-O	

For the sake of an objective analysis, it was important to focus the attention of the participants on the criteria for analyzing the skull and the face relationship only. This study should include both positive and negative SFOs. The procedure to obtain each type of SFO was different.

For positive cases, optimal SFOs were achieved using the following procedure. The DICOM images resulting from the CBCT machine were automatically processed to obtain the corresponding 3D face and 3D skull models. After positioning homologous points in both the 3D face model and the photograph, the former

Table 7.14   Consistency of the bony and facial outlines/morphological curves

No.	Criteria	View	Do you employ it?
Group 3	Overall consistency of the bony and facial outlines/ morphological curves:	Frontal(F)/ Lateral-Oblique (L-O)	–
3.1	The outline of the frontal bone follows the forehead outline.	L-O	
3.2	The skull and head height is similar (account for variation in soft tissue and distortion in the perception created by presence of hair).	L-O	
3.3	The width of the cranium fills forehead area of the face.	F	
3.4	The length of the skull from menton to bregma fits within the face.	F	
3.5	The lateral line of the zygomatic bone matches the outline of cheek.	F	
3.6	The chin outline is consistent with the mental outline.	L-O	
3.7	The gonial outline follows the outline of jaw angle.	L-O	
3.8	The outline of frontal process of the zygomatic bone can be aligned with the process seen in the face. The outline of the zygomatic arch can be fitted between the skull and the face. (This criterion is more easily appreciated in individuals with minimal soft tissue thickness).	L-O	
3.9	The arcus supraciliariaris follows supraorbital margin.	L-O	
3.10	The temporal line is consistent with the outline of the forehead (Sometimes the temporal line cannot be distinguished).	F	
3.11	The outline of the face and the outline of the skull all along the contour follow each other maintaining symmetrical flow by side.	F	
3.12	The outline of the nose in the face represented by shade distribution follows the outline of the nasal bone in the skull maintaining symmetrical flow by side.	F	
3.13	The asymmetries in the facial organs especially the nose reveal consistency with the asymmetries in the organs of the skull including the nasal structures. These include 1. Asymmetries in the nasal area including the nasal bone, piriform aperture, and nasal spine 2. Asymmetries in the zygomatic area, especially the extent of protrusion of the arch 3. Asymmetries in the occlusal line caused by protruding or overriding anterior dentition reflected as asymmetries in the corresponding part of the lip closure line 4. Asymmetries in the gonia	F	
3.14	The arcus supraciliariaris follows supraorbital margin.	F	

(continued)

Table 7.14 (continued)

No.	Criteria	View	Do you employ it?
Group 3	Overall consistency of the bony and facial outlines/ morphological curves:	Frontal(F)/ Lateral-Oblique (L-O)	–
3.15	The outline of the of the nasal bones follows the outline of the nose in the skull with minimal tissue thickness allowance	L	
3.16	The sagittal outline of the nasal cartilage is the mirror image of the contour of the pyriform aperture, relative to Line №1 passing through the rhinion point (1) and parallel to Line №2 joining the nasion(2) and the prostion (3) anthropometric points. Line№1 splits the entire nasal cartilage into two symmetric mirrored halves: the protruding part of the nose cartilage is the mirror image of the cartilage filling the pyriform aperture of the cranium.		
3.17	Lateral nasal bulges.	F	
3.18	Oblique line of the mandible.	F	
3.19	Dental information (bony to bony consistency).	F/L-O	

was automatically projected onto the latter so as to obtain an ideal match. Then, the parameters originating from that match between the 3D face model and the photograph were applied to the 3D skull model, resulting in an objective and accurate SFO. The latter superimposition is considered a ground-truth SFO.

Figure 7.1 shows an overview of the whole ground-truth data creation process.

For negative cases, the SFOs were performed using Face2Skull[TM] software. An expert was asked to obtain the best possible SFO and to judge the skull-face relationship without being informed of the actual negative relationship to avoid biasing the SFO process.

For the criteria assessment study, 37 forensic experts were asked to indicate which specific criteria they are going to use for evaluating the 24 skull-face relationships. The criteria are organized in four groups according to the family criteria: lines (group 1), landmarks-soft tissue (group 2), outlines (group 3), and positional relationship (group 4). Group 1 is composed of 28 criteria, group 2 has 27 criteria, group 3 is a set of 19 criteria, and group 4 is made up of 21 criteria.

Forensic experts were asked to evaluate the skull-face correspondence following a systematic approach. For each SFO, the degree of consistency of all the criteria previously selected was indicated using the following values: 0: not evaluable, 1: not match, 2: poor match, 3: doubtful match, 4: good match and 5: perfect match.

In order to avoid personal interpretations, MEPROCS partners assigned in advance (before giving the instructions to the participants) the value 0 to those criteria they considered unable to be visually checked due to the noisy nature of the image, the absence of the bony part, or the pose of the photograph. That was carried out for each single SFO case.

Table 7.15 Positional relationship analyzed to assess anatomical consistency

No.	Criteria	View	Do you employ it?
Group 4	Overall consistency positional relationships between the skull and face:	Frontal (F)/ Lateral-Oblique (L-O)	–
4.1	The prosthion lies posterior to the anterior edge of the upper lip. The occlusal and the lip closure line are consistent.	F	
4.2	The lateral angle of the eye lies within the lateral wall of the orbit.	L-O	
4.3	The lateral orbital margin at the Whinall's turbercle matches or approximates the of the ectocathionposition	L-O	
4.4	The width and height of piriform aperture lie within the borders of the nose.	F	
4.5	The stomion lies at the central incisors (at the occusal line).	L-O	
4.6	The lateral margin of the piriform aperture matches or approximates the alare.	L-O	
4.7	The stomion lies at the central incisors (incisal margin of the upper incisors).	F	
4.8	The nasion is higher than the nasal root.	L-O	
4.9	The prosthion lies posterior to the anterior edge of the upper lip. The occlusal and the lip closure line are consistent.	L-O	
4.10	The medial margin of orbit aligns and superimposes with the endocanthion.	F	
4.11	The nasion is higher than the nasal root.	F	
4.12	The Whitnall's tubercle aligns with the ectochantus on the horizontal plane and ,vertically, the ectochantus lies medial to the tubercle. The orbital width is consistent with the eye-slit width.	F	
4.13	The chelion lies between the canine and the first pre-molar (at the occusal line).	F	
4.14	The porion aligns just posterior to the tragus, slightly inferior to the crus of the helix.	L-O	
4.15	The lower margin of piriform aperture matches the subnasale.	L-O	
4.16	The eyebrow generally follows the upper edge of the orbit over the medial two-thirds. At lateral superior one-third of the orbit, the eyebrow continues horizontally as the orbital rim begins to curve inferiorly.	F	
4.17	The external auditory meatus opening lies medial to the tragus of the ear. (Place a projecting marker at the ear canal to assess this criterion more easily).	F	
4.18	The chelion lies between the canine and the first pre-molar (at the occusal line).	L-O	

(continued)

Table 7.15 (continued)

No.	Criteria	View	Do you employ it?
Group 4	Overall consistency positional relationships between the skull and face:	Frontal (F)/ Lateral-Oblique (L-O)	–
4.19	The anterior nasal spine lies posterior to the base of the nose near the most posterior portion of the lateral septal cartilage.	L-O	
4.20	Gonial flare in the skull and the postero-lateral jaw angle outline in the face.	L-O	
4.21	Gonial flare in the skull and the postero-lateral jaw prominence in the face.	F	
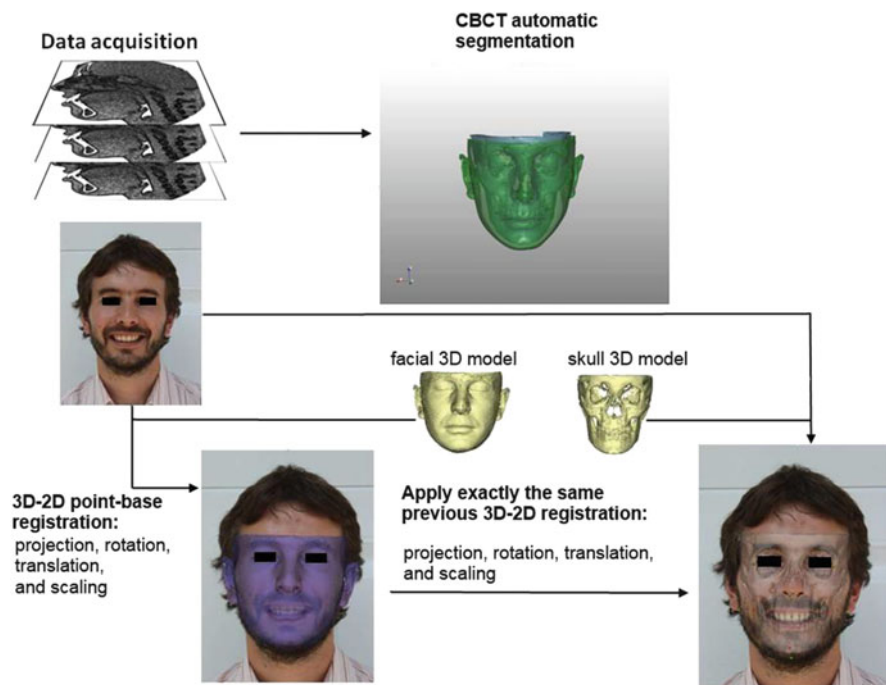
Fig. 7.1 Overview of the ground-truth data creation process

Finally, for each SFO case (and also for each CFS case when it implies more than one SFO), participants were asked to indicate the final identification decision according to the following scale: −3: strong support of not being the same person,−2: moderate support of not being the same person,−1: limited support of not being the same person, 0: not determined, +1: limited support of being the same

person, +2: moderate support of being the same person, and +3: strong support of being the same person. Therefore, the dataset is composed by:

The forensic expert, the specific SFO case and its state (positive or negative), the photograph of the SFO case (frontal or lateral pose), the criteria used by the expert in order to evaluate the corresponding SFO case, the family of the criteria, the degree of consistency of these latter criteria given by the expert (0: not evaluable, 1: not match, 2: poor match, 3: doubtful match, 4: good match and 5: perfect match), and the decision of the expert for the corresponding case (−3: strong support of not being the same person, −2: moderate support of not being the same person, −1: limited support of not being the same person, 0: not determined, +1: limited support of being the same person, +2: moderate support of being the same person, and +3: strong support of being the same person).

7.3.1.1 Data Analysis

We have developed three studies with the following characteristics:

1. According to the data employed:

   (a) With all the data.
   (b) Filtering (removing) the experts with a proficiency lower or equal than 0.5.
   (c) Filtering (removing) the scenarios (SFO cases) with higher standard deviation (fourth quartile).

2. According to the view of the photographs: frontal *versus* lateral poses
3. According to the family of criteria: lines, landmarks-soft tissue, outlines, and positional relationship

The statistical analysis developed relied on several concepts that are introduced below together with an example:

– Cases with decision (CD): the cases in which the expert's decision is different from 0 (not undetermined)
– Expert proficiency: the proportion of cases with decision in which the expert evaluated the status of the case correctly

$$EP = \frac{TP + TN}{CD}, \tag{7.1}$$

where TP is the number of positive cases with a positive decision and TN is the number of negative cases with a negative decision, and CD is the number of cases with decision

7.3.1.2 Correlation Between Two Variables

Before computing which are the most relevant criteria, we have calculated the correlation between the status of the identification case and the decision of the

forensic expert (correlation-based expert proficiency). Furthermore, we have also estimated the correlation between the value of a criterion and the status of a case (criterion correlation with ground truth). That correlation assesses the tendency of a criterion to have higher values on positive cases and lower values on negative ones.

Correlation is a statistical technique that can show whether and how strongly pairs of variables are related. The Pearson correlation coefficient is the most widely used. It measures the strength of the linear relationship between normally distributed variables. When the variables are not normally distributed or the relationship between the variables is not linear, it may be more appropriate to use the Spearman rank correlation method. Spearman's coefficient, like any correlation calculation, is appropriate for both continuous and discrete variables, including ordinal variables (Wilkinson 2006). Due to the nature of our dataset, we have applied the Spearman rank correlation method.

The Spearman correlation coefficient is defined as the Pearson correlation coefficient ρ between the ranked variables. For a sample of size n, the n raw scores X_i, Y_i are converted to ranks x_i, y_i, and ρ is computed from

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}, \tag{7.2}$$

where $d_i = x_i - y_i$ is the difference between ranks. Identical values (rank ties or value duplicates) are assigned a rank equal to the average of their positions in the ascending order the values.

The sign of the Spearman correlation indicates the direction of association between X (the independent variable) and Y (the dependent variable). If Y tends to increase when X increases, the Spearman correlation coefficient is positive. If Y tends to decrease when X increases, the Spearman correlation coefficient is negative. A Spearman correlation of zero indicates that there is no tendency for Y to either increase or decrease when X increases. The Spearman correlation increases in magnitude as X and Y become closer to being perfect monotone functions of each other. When X and Y are perfectly monotonically related, the Spearman correlation coefficient becomes 1.

We have performed the Spearman correlation with a statistical test in order to estimate the significance of the results. The considered level of statistical significance was 0.05.

For the correlation-based expert proficiency, the specific aim was to test the null hypothesis (H_0) stating that the correlation between the decision of the expert and the status of the case are not correlated. In the case of the criterion correlation with the ground truth, the goal was to test the null hypothesis (H_0) stating that the correlation between the value of the criterion and the status of the case are not correlated.

As complementary studies, we have added the following analyses:

– Criterion weighted correlation with ground truth: same as the correlation with ground truth, except that the correlation coefficient associated with each expert is weighted according to his proficiency.

– Criterion variability: it is computed as the mean standard deviation of the criterion evaluation over the same case. It aims to assess the subjectivity of a criterion.

7.3.1.3   Linear Regression

Correlation makes no a priori assumption as to whether one variable is dependent on the other(s) and is not concerned with the relationship between variables; instead, it gives an estimate as to the degree of association between the variables. In fact, correlation analysis tests for interdependence of the variables.

As regression attempts to describe the dependence of a variable on one (or more) explanatory variables, it implicitly assumes that there is a one-way causal effect from the explanatory variable(s) to the response variable, regardless of whether the path of effect is direct or indirect.

Therefore, we have complemented the latter correlation analysis with a linear regression test.

Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable. This functional relationship may then be formally stated as an equation, with associated statistical values that describe how well this equation fits the data.

A linear regression line has an equation of the form $Y = a + bX$, where X is the explanatory variable and Y is the dependent variable. The slope of the line is b, and a is the intercept (the value of y when $x = 0$).

The most common method for fitting a regression line is the method of least-squares. This method calculates the best-fitting line for the observed data by minimizing the sum of the squares of the vertical deviations from each data point to the line (if a point lies on the fitted line exactly, then its vertical deviation is 0). Because the deviations are first squared, then summed, there are no cancellations between positive and negative values.

We have performed the linear regression test in order to estimate if the value of a criterion depends on the status of the case. The goal is to test the null hypothesis (H_0) stating that the value of a criterion has not an influence in the status of the case.

7.3.2   Results

With the aim of providing a feasible forum of discussion we have focused the analysis on only one scenario, the 1.a, that is, all the data (participants and CFS cases) are considered at the same time. Additionally, in some parts of the document, we also refer to the second scenario where the data was divided into two different sets according to the view of photograph: frontal and lateral.

Table 7.16 depicts the proportion of cases with decision in which the expert evaluates correctly status of the case (proficiency) and the number of cases given by

Table 7.16 Cases with decision and simple expert proficiency

Forensic expert	Proficiency	Cases with decision
F1	0.62	0.96
F2	0.58	1.00
F3	0.58	0.83
F4	0.50	1.00
F5	0.54	0.79
F6	0.38	0.96
F7	0.62	1.00
F8	0.46	0.96
F9	0.42	0.96
F10	0.42	0.83
F11	0.48	0.96
F12	0.62	1.00
F13	0.54	0.83
F14	0.50	0.92
F15	0.58	1.00
F16	0.67	1.00
F17	0.67	1.00
F18	0.46	0.88
F19	0.54	1.00
F20	0.58	0.92
F21	0.67	0.96
F22	0.67	0.88
F23	0.75	1.00
F24	0.50	1.00
F25	0.50	0.88
F26	0.42	0.83
F27	0.67	1.00
F28	0.42	0.96
F29	0.50	1.00
F30	0.62	1.00
F31	0.75	0.96
F32	0.46	1.00
F33	0.62	1.00
F34	0.67	0.92
F35	0.54	0.92
F36	0.54	0.75
F37	0.58	1.00

the expert with a decision different of the zero value. The average performance was poor. In the best case, the rate is just 75% (only two experts). A percentage of 37 (14 experts) of the participants did not overcome 50% of correct answers. The performance is thus worse than previous studies that also involved the SFO stage (Yoshino et al. 1995; Gordon and Steyn 2012; Jayaprakash et al. 2001; Martin and Saller 1957). Possible explanations for this low performance rates are

– The absence of a complete cranium.
– The quality of some 3D models, which in some cases present noisy parts and artifacts.
– The materials given to the participants do not include the 3D skull models but just a projection on a 2D plane.
– The isolation of the decision-making stage given an SFO.

While the negative influence of the first three is quite evident, the fourth is not clear at all. It is the first time that such a study where SFOs are given at hand is developed. The process of overlaying the skull over the face also involves a continuous comparison of the skull-face relationship that is not performed within this study.

Table 7.17 shows the results of the Spearman test in order to calculate the correlation between the status of the identification case and the decision of the forensic expert. The proficiency in Table 7.16 corresponds to the Spearman correlation coefficient. We observe that seven experts achieve a significant correlation between their decision and the status of the case. F20, F21, F22, F23, F31, F34, and F36 have a Spearman positive correlation >0 with a p-value<0.05. In the rest of cases, we cannot reject the null hypothesis of correlation between the decision of the expert and the status of the case; that is, we cannot assert that the decision of the expert is correlated with the status of the case with a confidence level of 95%.

Table 7.18 shows the number of times a criterion has been evaluated over the total number of evaluations (each participant for each SFO case), "Usage C." It also shows the percentage of participants that employed a criterion at least once, "Usage P." Both statistics are depicted for all the cases ("All"), only frontal cases ("Frontal"), and only lateral/oblique cases ("Lateral"). Those criteria that were employed in less than the 10% of the cases were removed from the corresponding study. In particular, criteria G2.10, G2.26, G3.2, and G3.4 do not reach the required 10% of usage irrespective of the dataset considered (all, only frontal, or only lateral). Criteria that were employed by less than 30% of the participants were also not considered in the corresponding study. The motivation for avoiding these criteria is related to the lack of significance of reduced samples of data.

Table 7.19 presents the results of the correlation between the value of a criterion and the status of the case with a p-value ≤ 0.05, that is, statistically significant results. That correlation assesses the tendency of a criterion to have higher values on positive cases and lower values on negative ones.

Eight experts obtained a correlation between one or more criteria and the status of the case with a confidence level of 95%. Hence, we can affirm that the use of some criteria is significantly correlated with the status of the case; that is, some criteria have higher values on positive cases and lower values in negative ones. In most cases, we obtain a positive correlation: when the degree of a criterion increases, the status of the case tends to a positive identification.

The performed regression analysis computes the independence between the value of a criterion and the status of the case.

Table 7.17 Spearman tests, correlation-based expert proficiency, cases with p-values <0.05 in bold

Forensic expert	Correlation coefficient	p-Value
F1	0.29	0.17
F2	0.21	0.33
F3	0.32	0.13
F4	0.21	0.33
F5	0.33	0.11
F6	−0.07	0.75
F7	0.03	0.91
F8	−0.08	0.71
F9	−0.10	0.64
F10	0.13	0.53
F11	0.24	0.27
F12	0.26	0.21
F13	0.26	0.21
F14	0.10	0.63
F15	0.30	0.15
F16	0.38	0.07
F17	0.34	0.11
F18	0.25	0.23
F19	0.22	0.31
F20	0.61	**0.00**
F21	0.45	**0.03**
F22	0.52	**0.01**
F23	0.49	**0.01**
F24	0.10	0.63
F25	0.23	0.29
F26	−0.03	0.88
F27	0.21	0.33
F28	−0.22	0.30
F29	−0.05	0.81
F30	0.27	0.21
F31	0.54	**0.01**
F32	−0.30	0.15
F33	0.29	0.17
F34	0.50	**0.01**
F35	0.30	0.16
F36	0.40	0.05
F37	0.13	0.54

Table 7.20 depicts the results that have a p-value ≤ 0.05, that is, those cases that reject the null hypothesis. Thus, shows the criteria that have significant influence on the status of the case.

Nine experts obtain a dependency between one or more criteria and the status of the case with a confidence level of 95%. Hence, we can affirm that the use of some

Table 7.18 Criterion usage

Criterion	All UsageC	All UsageP	Frontal UsageC	Frontal UsageP	Lateral UsageC	Lateral UsageP
G2.14	0.72	0.83	0.78	0.83	0.67	0.78
G3.6	0.67	0.97	0.46	0.64	0.88	0.97
G2.13	0.62	0.75	0.64	0.70	0.60	0.75
G2.3	0.59	0.94	0.36	0.48	0.88	0.94
G2.5	0.58	0.86	0.37	0.45	0.81	0.86
G4.16	0.57	0.86	0.84	0.86	0.32	0.54
G3.15	0.56	0.91	0.27	0.43	0.88	0.91
G4.2	0.56	0.89	0.43	0.51	0.70	0.86
G2.1	0.55	0.83	0.80	0.83	0.33	0.43
G4.1	0.54	0.94	0.71	0.78	0.40	0.54
G3.7	0.54	0.89	0.36	0.48	0.72	0.89
G3.12	0.53	0.91	0.83	0.86	0.25	0.43
G2.9	0.52	0.89	0.85	0.89	0.20	0.35
G3.9	0.52	0.75	0.41	0.48	0.65	0.75
G3.11	0.52	0.86	0.84	0.86	0.20	0.43
G4.4	0.52	0.78	0.78	0.78	0.26	0.43
G3.5	0.51	0.91	0.86	0.91	0.18	0.45
G4.15	0.51	0.73	0.42	0.54	0.62	0.70
G4.8	0.51	0.81	0.32	0.40	0.71	0.78
G1.1	0.50	0.86	0.86	0.86	0.16	0.40
G3.14	0.50	0.73	0.70	0.73	0.29	0.43
G2.2	0.48	0.75	0.27	0.32	0.69	0.75
G2.8	0.47	0.78	0.69	0.78	0.25	0.43
G2.16	0.41	0.64	0.23	0.32	0.58	0.64
G2.15	0.40	0.67	0.25	0.37	0.56	0.67
G4.20	0.40	0.73	0.25	0.35	0.56	0.70
G3.8	0.39	0.70	0.26	0.43	0.52	0.62
G4.7	0.39	0.62	0.53	0.56	0.25	0.40
G2.11	0.39	0.45	0.45	0.45	0.33	0.37
G1.6	0.38	0.62	0.59	0.62	0.18	0.29
G2.7	0.38	0.67	0.12	0.18	0.64	0.67
G4.11	0.38	0.73	0.49	0.73	0.26	0.35
G2.4	0.38	0.94	0.24	0.37	0.51	0.91
G4.12	0.37	0.59	0.56	0.59	0.19	0.32
G4.13	0.37	0.67	0.56	0.64	0.18	0.29
G2.12	0.36	0.51	0.50	0.51	0.22	0.43
G4.19	0.36	0.70	0.06	0.16	0.66	0.70
G4.21	0.36	0.62	0.58	0.62	0.11	0.35
G4.18	0.35	0.59	0.22	0.32	0.43	0.56
G2.16	0.32	0.62	0.11	0.21	0.50	0.59
G2.23	0.31	0.51	0.48	0.51	0.11	0.27
G1.7	0.30	0.56	0.48	0.56	0.08	0.24
G3.17	0.28	0.48	0.40	0.45	0.15	0.27
G1.8	0.28	0.56	0.48	0.54	0.06	0.18
G1.2	0.27	0.70	0.40	0.70	0.10	0.29
G2.25	0.25	0.32	0.27	0.32	0.21	0.27

(continued)

Table 7.18 (continued)

Criterion	All		Frontal		Lateral		Criterion	All		Frontal		Lateral	
	UsageC	UsageP	UsageC	UsageP	UsageC	UsageP		UsageC	UsageP	UsageC	UsageP	UsageC	UsageP
G2.18	0.46	0.56	0.47	0.54	0.46	0.54	G2.27	0.24	0.27			0.03	0.10
G4.3	0.46	0.70	0.42	0.48	0.50	0.59	G2.22	0.24	0.45	0.44	0.45		
G4.6	0.46	0.73	0.33	0.484	0.59	0.64	G2.24	0.24	0.27			0.05	0.16
G4.10	0.44	0.78	0.75	0.78	0.13	0.40	G2.21	0.21	0.37	0.33	0.37	0.24	0.29
G4.9	0.44	0.78	0.22	0.29	0.65	0.75	G3.16	0.19	0.35	0.12	0.16	0.09	0.32
G4.5	0.44	0.73	0.33	0.48	0.54	0.73	G3.10	0.18	0.54	0.26	0.54	0.17	0.48
G1.4	0.43	0.67	0.64	0.67	0.22	0.35	G3.19	0.18	0.48	0.18	0.48	0.26	0.75
G1.5	0.42	0.73	0.66	0.73	0.18	0.35	G3.1	0.18	0.75	0.09	0.13		
G3.13	0.42	0.73	0.70	0.73	0.13	0.29	G2.20	0.17	0.24			0.24	0.51
G2.17	0.41	0.45	0.45	0.45	0.37	0.40	G4.14	0.16	0.51	0.03	0.08	0.22	0.48
G1.3	0.41	0.75	0.68	0.73	0.14	0.35	G2.19	0.13	0.48	0.02	0.10	0.15	0.37
G3.3	0.41	0.81	0.73	0.81	0.09	0.29	G4.17						
G3.18	0.41	0.67	0.56	0.62	0.26	0.43							

Table 7.19 Spearman test, correlation statistically significant between criterion and the status of the case

Forensic expert	Criterion	Correlation coefficient	p-Value
F3	G2.9	0.48	0.02
F3	G2.14	0.51	0.01
F4	G1.4	0.43	0.03
F4	G3.18	0.49	0.01
F7	G2.18	0.41	0.04
F11	G3.12	0.47	0.02
F11	G3.13	0.44	0.03
F11	G3.15	0.65	0.00
F16	G1.2	0.42	0.03
F16	G1.5	0.41	0.04
F16	G1.8	0.52	0.00
F16	G2.3	0.49	0.01
F16	G2.6	0.52	0.00
F16	G2.14	0.56	0.00
F16	G3.6	0.46	0.02
F16	G4.14	0.50	0.01
F19	G2.8	0.47	0.02
F19	G3.7	0.43	0.03
F19	G4.15	0.50	0.01
F21	G3.19	0.41	0.04
F30	G3.18	0.48	0.01

criteria is significantly dependent on the status of the case; that is, some criteria have higher values on positive cases and lower values in negative ones.

It is important to note that we have achieved similar results in the correlation and regression tests (Tables 7.19 and 7.20). The criteria that present the most influence in the status of the case are G2.14, G2.17, G2.18, and G3.12.

A boxplot with the expert's assessment across the scenarios is depicted in Fig. 7.2. This boxplot shows the significant variability within each of the expert's responses. In general, both negative and positive cases have similar performance rates although a lower variability resulted in the evaluation of the positive ones. While there are only two negative cases (4-2 and 11-1) where most of the participants ($\geq 75\%$) made a correct evaluation, there are four positive cases with a similar successful evaluation (3-1, 7-1, 13-1, and 18-1). Looking at the median values (black horizontal line inside the boxes), there are three negative cases that were incorrectly assessed by most of the participants: SFO cases 4-1, 15-1, and 16-1. The median values of the other three cases (8-1, 10-1, and 17-1) fall in the undetermined category (value 0). Similarly, there are three positive cases that were incorrectly evaluated by most of the participants: SFO cases 5-1, 5-2, and 12-1. For all these cases, 75% of the participants did not make the correct identification. Differences were not observed between the evaluations of lateral versus frontal views.

Table 7.20 Linear regression test, influence statistically significant between criterion and the status of the case

Forensic expert	Criterion	F value	p-Value
F3	G2.9	5.97	0.02
F3	G2.14	7.53	0.01
F4	G1.4	5.00	0.03
F4	G3.18	6.66	0.01
F7	G2.18	5.96	0.02
F11	G3.12	4.62	0.04
F11	G3.15	11.56	0.00
F11	G4.7	4.10	0.05
F16	G1.5	4.66	0.04
F16	G1.8	7.97	0.00
F16	G2.3	6.82	0.01
F16	G2.6	7.97	0.00
F16	G2.14	8.00	0.00
F16	G3.6	4.23	0.05
F16	G3.12	4.45	0.04
F16	G4.14	7.61	0.01
F18	G4.3	4.22	0.05
F19	G2.8	5.03	0.03
F19	G2.17	4.50	0.04
F19	G2.18	4.14	0.05
F19	G4.15	5.76	0.02
F21	G3.19	5.55	0.02
F23	G2.11	4.00	0.05
F30	G3.18	7.48	0.01
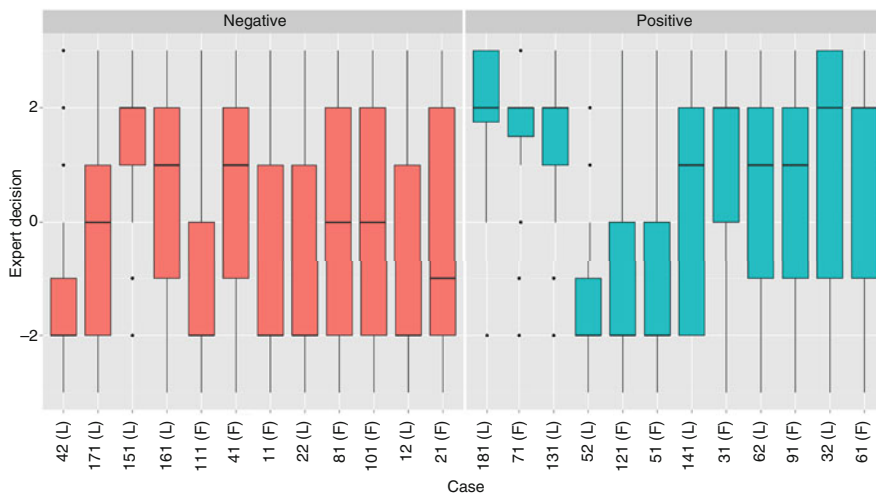
Fig. 7.2 Statistical representation of the expert's assessment for each (negative and positive) SFO case. Expert decisions (between −3 and +3) on the y-axis and SFO cases on the x-axis. F and L, in brackets after the number of the case, indicate frontal and lateral view cases, respectively

The subjectivity was measured as the standard deviation of the evaluations (Table 7.21). The standard deviation was computed on each case, and then the values were averaged. The ranges of values for the criteria are within the interval [1, 5] and thus can conclude that there is a significant distribution in the evaluations by the different participants with standard deviations ranging from 0.85 to 1.31.

A primary goal for the current study is to provide forensic anthropologists with the means (objective data) to select a set of criteria, or to establish a ranked order of

Table 7.21 Criterion subjectivity

Criterion	Mean standard deviation	Criterion	Mean standard deviation
G2.20	0.85	G3.16	1.12
G3.10	0.87	G2.12	1.12
G3.1	0.92	G2.17	1.12
G1.5	0.92	G2.14	1.12
G1.8	0.93	G3.11	1.14
G3.18	0.97	G3.9	1.14
G1.3	0.97	G4.12	1.14
G2.19	0.97	G2.3	1.14
G2.24	1.01	G3.14	1.14
G1.2	1.02	G2.15	1.14
G2.16	1.02	G4.20	1.14
G2.23	1.03	G3.6	1.15
G4.19	1.03	G2.22	1.15
G1.7	1.04	G3.8	1.15
G2.25	1.05	G4.4	1.15
G4.10	1.06	G4.6	1.15
G1.6	1.07	G1.1	1.17
G2.21	1.07	G2.11	1.17
G2.9	1.07	G2.5	1.18
G2.27	1.08	G4.15	1.19
G4.13	1.10	G4.8	1.19
G2.7	1.10	G3.5	1.19
G2.18	1.10	G4.3	1.19
G2.8	1.10	G4.16	1.19
G3.7	1.10	G4.2	1.20
G1.4	1.10	G2.1	1.20
G4.5	1.10	G4.18	1.21
G4.14	1.10	G3.19	1.21
G2.4	1.11	G4.11	1.21
G4.1	1.11	G3.13	1.21
G3.3	1.11	G32.2	1.23
G4.21	1.11	G3.12	1.23
G2.13	1.11	G3.15	1.26
G4.9	1.11	G2.6	1.31
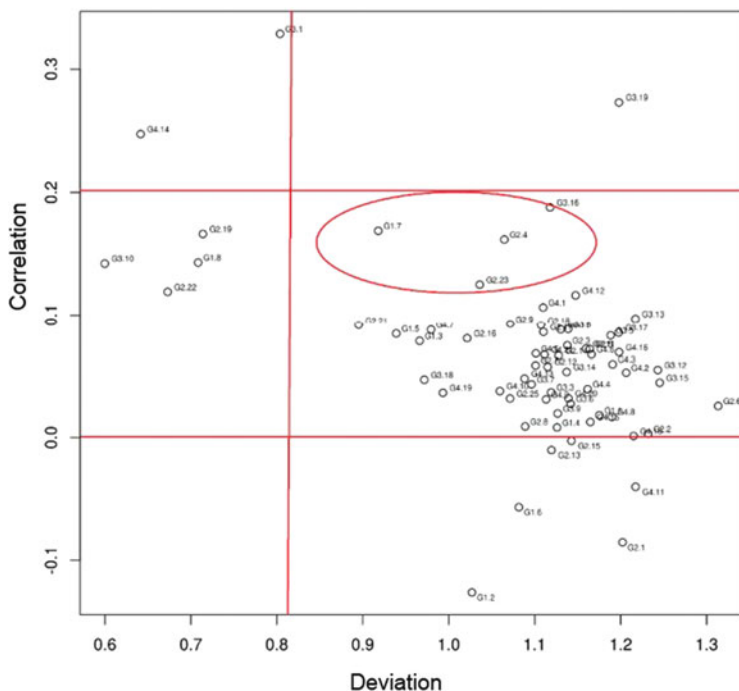
Fig. 7.3 Scatter plot including all the criteria under study spatially distributed according to their subjectivity (x-axis) and discriminative power (y-axis)

preference, with the most discriminative power and easy to evaluate traits. Figure 7.4 visualizes the standard deviation (related to the ease of objective assessment) and the correlation (related to the discriminatory power criterion).

Three red lines split five groups of criteria with one group split into two separate groups with have six groups in total, which represent the best criteria and the highest discriminatory power with the lowest variability in the top left corner (G3.1 and G4.14). Below this region, the criteria that can be considered easy to evaluate, which are important criteria in terms of discriminative power with high correlation values (G3.10, G2.19, G1.8, and G2.22). In the top right corner, the criteria with almost the highest discriminative power with highest variance are grouped (G3.19).The largest area, more or less in the center of the Fig. 7.3, shows the majority of the criteria that in general are not significantly different (G4.12, G4.1, G2.9, G2.16, G4.7, G1.3, G1.5, and G2.21).Within this region, surrounded by a red-colored circle, we have identified a fifth group composed by criteria with a good trade-off between subjectivity and discriminative power (G3.16, G1.7,G2.4,andG2.23).Finally, the right bottom corner groups the least useful criteria with regards to their subjectivity and do not discriminate between face and skull (G1.2, G2.1, G4.11, G1.6, G2.15, and G2.13).

Figure 7.4 depicts the differences between the frontal views that cluster specific criteria according to the pose of the person's face within the photograph.

For the *frontal cases*, five groups can be differentiated. At the top scatter plot, the criteria with highest discriminative power (G3.19) are observed. Below this group,
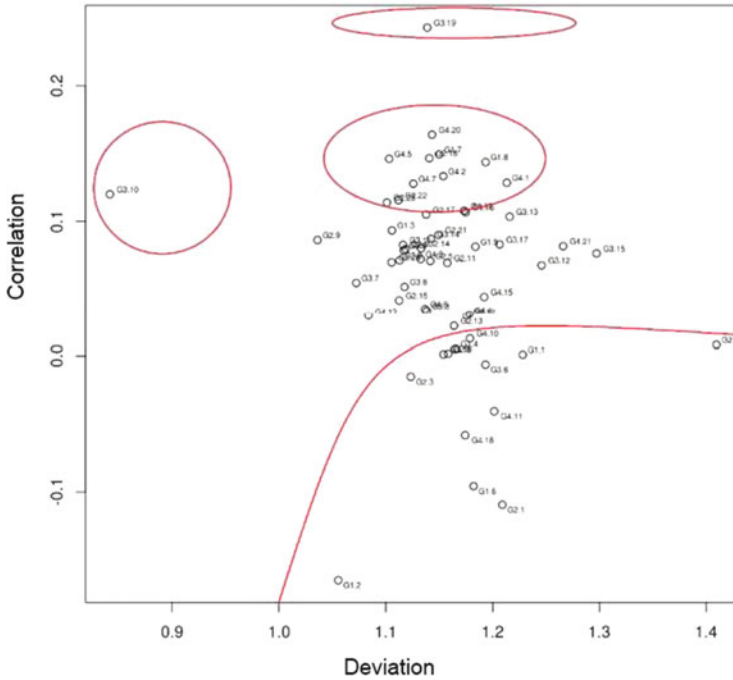
Fig. 7.4 Criterion according to the frontal pose of the person's face in the photograph

the group with the criteria with a good trade-off between subjectivity and discriminative power is observed (G4.20, G4.5, G1.7, G2.18, G1.8, G4.2, G4.7, G4.1, G2.22, and G2.23). On the left side of the scatter plot, easy to evaluate criterion is depicted (G3.10) with the least amount of variability. In addition, it is an important criterion in terms of discriminative power with high correlation values. In the center of the scatter plot, the majority of the criteria are found showing the least amount of difference among them. Finally, the right bottom region groups the least useful criteria with the highest subjectivity and that cannot be used to discriminate between face and skull (G1.2, G2.1, G1.6, G4.18, G4.11, G2.3, G3.6, G1.1, G2.8, G3.18, G1.4, G2.6, and G4.10).

Figure 7.5 depicts the differences between the lateral views that cluster specific criteria according to the pose of the person's face within the photograph.

For the *lateral cases*, although they can be grouped into eight separate groups, the two groups in the center part (between correlation values of 0 and 1.2) are considered as part of the same group of criteria with almost no discriminatory power. On the top left corner, the best criteria are represented (G3.1and G4.17), which also have the greatest discriminatory power and least variability. At the top right corner, encloses the group with the greatest discriminatory power and the greatest variability (G4.14 and G3.19). Below group, a group identified (criteria G2.4, G3.11, G2.9, G1.1, G4.10, G2.12, and G2.19) as still important correlation with the identification decisions and a significant variability is shown. Similarly, criterion G3.18 holds important discriminatory power but has a significantly lower variability. As in the
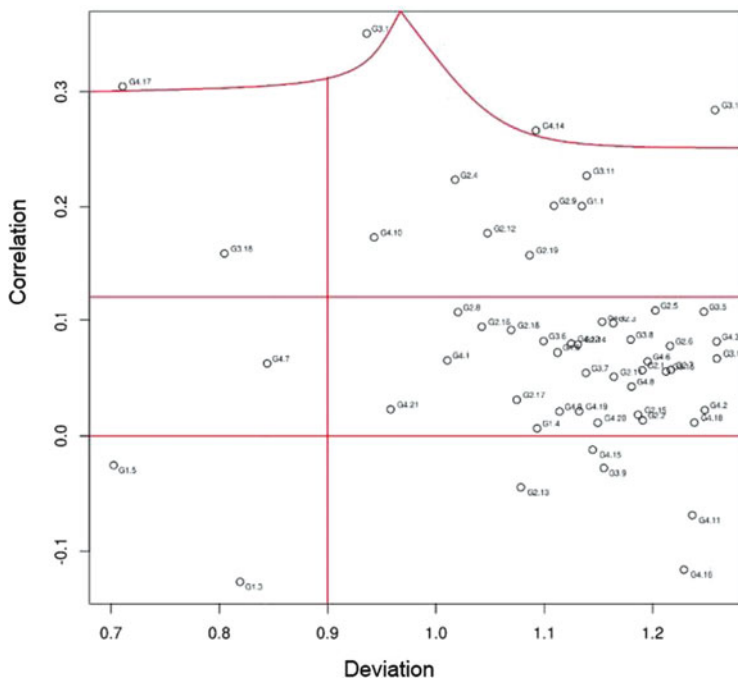
Fig. 7.5 Criteria according to the lateral pose of the person's face in the photograph

other two sections, the central part of the scatter plot contains the majority of the criteria, which do not hold significant correlation values. Finally, the bottom right part of the scatter plot contains the criteria with the greatest subjectivity and that cannot discriminate between face and skull (G4.15, G3.9, G2.13, G4.11, and G4.16). Note that G3.14 refers to the same anatomical correspondence criterion as G3.9 but analyzed on different image views.