HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

Egyptian Journal of Forensic Sciences

journal homepage: http://www.journals.elsevier.com/egyptian-journal-of-forensic-sciences

ORIGINAL ARTICLE

Performance of distance-based matching algorithms in 3D facial identification

CrossMark

Petra Urbanová *

Laboratory of Morphogy and Forensic Anthropology, Department of Anthropology, Faculty of Science,
Masaryk University, Kotlarska 2, 611 37 Brno, Czech Republic

Abstract Facial image identification is an area of forensic sciences, where an expert provides an opinion on whether or not two or more images depict the same individual. The primary concern for facial image identification is that it must be based on sound scientific principles. The recent extensive development in 3D recording technology, which is presumed to enhance performances of identification tasks, has made essential to question conditions, under which 3D images can yield accurate and reliable results. The present paper explores the effect of mesh resolution, adequacy of selected measures of dissimilarity and number of variables employed to encode identity-specific facial features on a dataset of 528 3D face models sampled from the Fidentis 3D Face Database ($N \sim 2100$). In order to match 3D images two quantitative approaches were tested, the first based on closest point-to-point distances computed from registered surface models and the second grounded on Procrustes distances derived from discrete 3D facial points collected manually on textured 3D facial models. The results expressed in terms of rank-1 identification rates, ROC curves and likelihood ratios show that under optimized conditions the tested algorithms have the capacity to provide very accurate and reliable results. The performance of the tested algorithms is, however, highly dependent on mesh resolution and the number of variables employed in the task. The results also show that in addition to numerical measures of dissimilarity, various 3D visualization tools can be of assistance in the decision-making.

© 2016 The International Association of Law and Forensic Sciences (IALFS). Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Facial image identification involves many areas of forensic or commercial security applications,[1] where tasks of one-to-one (verification, authentication) or one-to-many matching (identification) are performed. The most frequent facial identification in the forensic settings is made by eyewitnesses in real time. Alternatively, surveillance, monitoring or screening by police officers may also require that an identification task be conducted in real time. Nevertheless, face identification executed as a part of forensic casework is almost exclusively grounded on off-line pre-recorded evidence. If executed manually by professionals trained to recognize and systematically examine

* Tel.: +420 5 49496206.
  E-mail address: urbanova@sci.muni.cz.

identity-coding facial features, these tasks, known as facial image comparison,[2] are only effective when conducted as a part of small-scale image processing. For large-scale comparisons, automated or semi-automated approaches customarily referred to as facial recognition and performed by computer-aided algorithms, are more practical choices.

It is generally acknowledged that matching image-based forensic evidence from crime scenes (surveillance videos) with suspects (mug shots, ID photographs) is extremely dependent on image medium,[3] image quality,[2] perpetrator's pose or disguise.[4] Bruegge[5] noted that although various law-enforcement agencies had been successfully using automated face recognition systems for screening, other tasks, such as negative identifications (e.g., ability to eliminate a person) are more difficult to achieve. Recently, new developments in 3D technologies have shown that at least some of the key shortcomings of forensic facial image identification can be overcome and performances can be improved.[6] Specifically, 3D images preserve realistic depth information, which minimizes the effect of head pose, framing and lighting – image properties that largely interfere with manual as well as auto-mated facial identification.[7]

There are a vast number of strategies capable of highlighting or extracting individually or group-specific facial representations depicted in both 2D and 3D images. In addition to manual comparative approaches employing side-by-side comparison or image superimpositions,[8] morphological analysis specifying color, texture, shape and size similarities in facial appearance has been routinely utilized in the framework of forensic anthropology. Generally, anthropologists are accustomed to approach tasks of similarity in terms of either visual trait assessment or somatometrics (anthropometrics, photo-anthropometry).[9,10] In the visual (morphological) approach the face is subdivided into gross anatomical components, e.g., forehead, cheeks, ears etc., and for each component a set of component characteristics is described in detail.[11–15] Recently, guidelines to forensic facial comparison together with a standardized list of facial features have been provided by The Facial Identification Scientific Working Group (FISWG).[16] In addition to the verbal description, tracing of individual facial features on images in order to outline similarities or differences have been suggested.[17]

In its traditional form, photo-anthropometrics relies on measuring linear distances and angles or derived indices[18,19] and/or inter-landmark distances on digital or printed photographs.[17,20] The method is most effective when the proportional relationships are aimed at comparing one image to another than at matching a single image against multiple targets.[11] However, if applied to regular two-dimensional images the approach has been labeled as very susceptible to head poses and additional technical noise present in images.[17] For 3D images, it has been demonstrated that the employment of the anthropometric approach is far less problematic.[6,21–24]

In addition to the traditional methods, multiple advanced quantitative techniques have been proposed in both 2D and 3D image comparison.[25] For instance, landmark-based or outline-based approaches subscribed under the rubric of geometric morphometrics have been utilized.[26,27] Unlike registration-free measurements or visual assessment, geometric data require a prior treatment in order to minimize the variance in position and rotation. For that purpose, Bookstein's shape coordinates or Procrustes fit in its numerous variants, e.g., Generalized Procrustes superimposition, Ordinary Procrustes fit, Generalized resistant fit are employable.[28]

Recently, owing to the progress in computing power of personal computers capable of processing dense point clouds and high-resolution meshes 3D surface comparisons have been gradually used.[6,29] Alike discrete facial points or outlines, 3D surface processing of human faces requires registration in order to minimize the variance due to location and rotation. This can be achieved by trivial approaches such as the 3-point or multiple-point registration or in a more sophisticated manner by various surface registration techniques.[30] The most common procedures utilize either Iterative Closest Point (ICP) algorithm[31] or one of its variants, e.g., EM-ICP algorithm.[32] This further establishes connectivity between points (vertices) of two and more meshes. The point-to-point connectivity is subsequently utilized to compute surface-to-surface deviations, of which various measures of dissimilarity can be retrieved (e.g., Hausdorff distance, average deviations, root mean square).[33–35] Alternatively, in order to compute a univariate measure of dissimilarity the reduction of dimensionality via Principal Components Analysis in conjunction with Mahalanobis or Euclidean distances have been widely used.[36,37] A review to available techniques has been summed up by Abate et al.[23] and more recently by Abdelwahab et al.[24]

In the present paper, pitfalls of examining 3D images are explored by testing algorithms incorporated into FIDENTIS Analyst, software developed as a potential tool for semi-automated and automated image comparisons in the framework of forensic anthropology. Using a dataset sampled from a large database of high-resolution 3D scans the performance of dissimilarity measures to identify 3D facial scans belonging to the same individual is tested.

2. Material

The studied sample consisted of 528 adult volunteers, 232 males and 296 females of Czech and Slovak nationality aged between 18.14 and 83.74 years (average 28.02 years) at the time of data acquisition. The average age for females counted 27.62 years while for males it averaged at 28.54 years. If tested no statistically significant differences in age between sexes were revealed (t-test, $t = -1.009$ p-value = 0.313). The participants were mostly recruited from the pool of students and employees of the author's home institution. None of them were diagnosed with any major medical conditions affecting facial morphology.

Of the studied sample, 28 volunteers – 7 males, 21 females aged from 19 and 36 years (average: 23.36 years, females: 23.07 years, males: 26.12 years) embodied the tested sample, whereas the remaining 500 individuals comprised the control sample.

3. Methods

For all participants, face geometry and texture was recorded with Vectra M1 3D facial scanner (Canfield Scientific, Inc, Fairfield, NJ). Vectra M1 is a double camera system with a limited narrow range (100°) which does not allow capturing the entire face at once. In result, lateral parts of the face and ears are not measured properly and distorted geometry and

blurry texture are frequently generated. To avoid the disturbances each face was captured in three head positions: frontal view, 30° left and right semi-profiles. Data acquisition was conducted indoors under controlled lighting conditions at the facilities of the Laboratory of Morphology and Forensic Anthropology, Department of Anthropology, Faculty of Science, Masaryk University (Brno, Czech Republic). Prior to scanning, all participants were asked to take off their glasses, accessories, earrings etc. Locks of hair covering the face or falling over ears were pulled back with a plastic headband and hair clips or brush back behind ears. Male participants were advised to shave the morning of the scanning session. In the course of data acquisition, each participant was instructed to maintain a natural head position and neutral facial expression with eyes open.

In total, 556 three-dimensional facial models were recorded – 500 single scan for the control sample and 2 scans per individuals for the tested subset of 28 individuals. Since then, all models have been included into the Fidentis 3D Face Database (http://www.fidentis.cz/database.php) available at the home institution for research purposes.

The acquired 3D images were subsequently processed in Mirror® Medical Imaging Software. The scans were manually trimmed of unwanted background, raw scanning noise and technically unsuitable parts, such as peripheral and eye regions in the semi-profiles or lateral parts of the nose in the frontal scans. After having been cleaned of unsuitable parts, the three partial 3D scans were aligned manually in program MeshLab v1.3.2 open source application for mesh processing (http://mesh-lab.sourceforge.net). Firstly, the left and right semi-profile scans were aligned approximately with the frontal model to complement the face using 3-point alignment algorithm. The automatic "register mesh" function was then run to match corresponding surfaces. Ultimately, the three models were merged into a single model using the Poisson filter. The filter parameters were set to preserve the highest level of details of the models (Octree Depth set to 12, Solver Divide to 10). In order to facilitate subsequent processes, all 3D meshes were pre-registered by manually marking 6 facial landmarks (right and left tragion, left lower palpebra, right and left entokanthion and pronasale) and by running an in–house MeshLab script which superimposed the model to a template face.

Final adjustments were carried out using GOM Inspect software – the meshes were checked for errors ("Eliminate Mesh Errors" function) and holes ("Close Holes – Automatically" function). They were trimmed manually of peripheral parts containing ears and reduced to the final resolution of 10,000 vertices. Additional mesh reduction was conducted in order to explore the effect of mesh resolution. For that purpose, the original number of vertices was gradually decreased to meshes of 5,000 (5k); 1,000 (1k); 500 (5h); 100 (1h) and 50 (5d) points.

One scan per individual from the tested sample was taken for a probe and in the pair-wise fashion (1:N) compared against targets composed of the control sample and the individual's corresponding scan ($N = 501$). The same was conducted for the control subset where each scan was compared against the remaining control scans.

Two types of facial features were extracted in order to identify 3D faces – closest point-to-point distances and discrete facial points (landmarks).

### 3.1. Closest point to point distances

In the closest point-to-point distances, the minimum distance from each point of a scan from the tested subset (a probe) to the nearest vertex of control scans (targets) were calculated while the direction of such vectors was omitted (only absolute values were taken into account). Prior to computation a probe and its corresponding array of candidates were registered by fitting all target meshes on to the probe. Registration algorithms were run on the edited and hole-filled meshes (saved in obj format) using the point-to-point variant of ICP algorithms featured in FIDENTIS Analyst v. 1.27beta application.[38] The incorporated variant of the algorithm establishes the correspondence between points of two meshes following the nearest neighbor criteria. For the purpose of the study the algorithm was set to an error rate of 0.05 and maximum iterations of 10. For each comparison, the global measure of dissimilarity was based on statistical descriptors – root mean square (RMS) and 75th percentile (PERC75) extracted from sets of closest point-to-point distances.

The entire procedure was then re-computed with size-adjusted meshes by employing the Scaling Iterative Closest Point (SICP) algorithm as implemented in FIDENTIS Analyst program. The algorithm applies a scaling parameter as proposed by Du et al.[39,40] in order to normalize size-related variance between two compared meshes.

3.2. Landmarks

Each facial scan was described by a set of 42 landmarks (Fig. 1, Table 1). 3D Cartesian coordinates for the entire set of landmarks were collected manually on texture-present 3D models with MeshLab 1.3.2 program using an established lab protocol.[41] Procrustes distance defined as a sum of squared distances between corresponding landmarks at its minimum was selected as the measure of dissimilarity. Procrustes distances correspond to shape differences between two objects in the morphospace. Prior to the computation it requires that the configuration of landmarks be registered into a common coordinate system and fitted to minimize least-square variations. The registration was conducted by the Ordinary Procrustes analysis where the set of landmarks corresponding to a target was fitted onto the configuration of a probe. This produced a set of Procrustes coordinates invariant in rotation, translation and size.

Similar to the mesh-based analysis the number of landmarks was gradually reduced to explore its effect on the overall performance. In the first step, 6 bilateral points of the auricular region were removed (only tragion was included). This was done as a precaution to relatively common situations in 3D image acquisition wherein lateral facial regions, ears includes, are obscured by hair or recorded in an insufficient quality. The subsequent reduction of landmarks was executed in regard to the inter-individual discriminative power accounted for each point. In each consecutive step 3 points with the lowest level of inter-individual variance (expressed in terms of mean squares) were removed. After the removal the variance was recomputed for the new set of points, the pair-wise Procrustes fit (OPA) included, and another 3-point set was removed. The bilateral points were taken as independent in the computation but eventually removed at the same time. In addition to 10
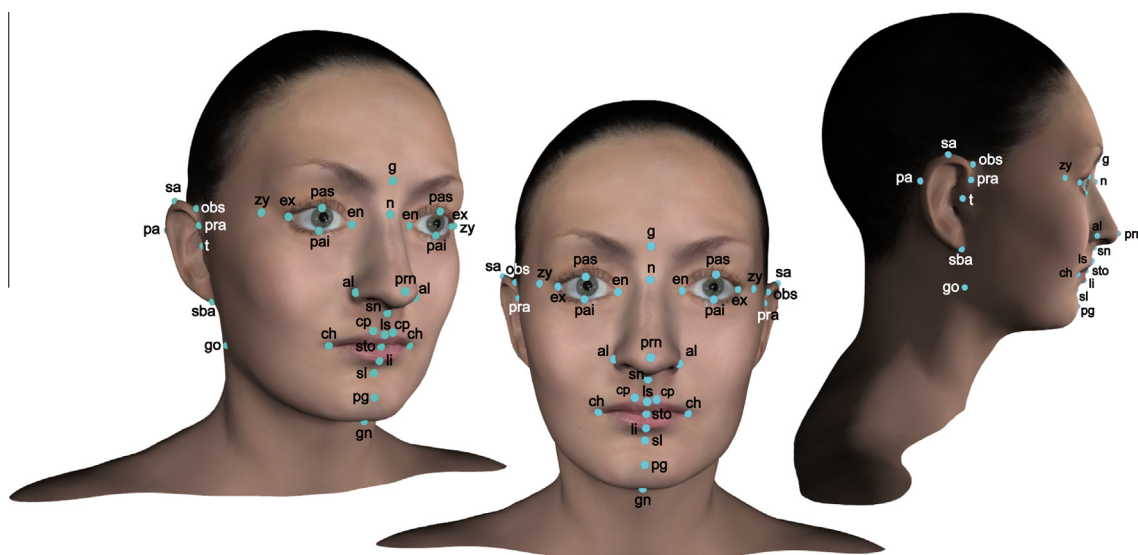
Figure 1   The set of landmarks acquired for each 3D image.

configurations constructed on the basis of the variance criterion, an alternative 6-point set was tested. The landmark configuration featuring right and left exocanthion, right and left endocanthion, pronasale and stomion was based on the theoretical assumptions that these points represent the primary source of individually specific variations.[42]

For each probe, the acquired results were sorted in descending order by values of the extracted measures of dissimilarity. For the 28 tested individuals match and non-match decision-making was in concordance with the provided ranking. If the target assigned with the rank of 1 matched the reference scan (probe) (i.e., it was derived from the same individual) then a true positive result was recorded. If any other scan was ranked at number one then a false positive result was recorded. Accordingly, rank-10 identification rates were computed. Matches up to the rank of 10 were recorded as true positive results; otherwise false positive results were registered.

The performance of the dissimilarity measures was evaluated using univariate ROC curves analysis. The optimal thresholds were determined based on a criterion which finds the closest point to the top-left corner of the ROC curve (i.e., it minimizes the sum of squared 1-sensitivity and 1-specificity values). In addition, results were re-computed with a threshold maximizing specificity over sensitivity and vice versa. The performance was expressed in terms of area under curves (AUC), for which 95% confidence interval was calculated using 500 bootstrappings.

Sensitivity (true positive rate, TPR, recall) was computed as true positive count divided by number of specimens in the tested sample ($N = 28$). Specificity (true negative rate, TNR) was computed as true negative counts divided by number of targets ($N = 501$). In addition, positive likelihood ratio (true positive count/false positive count) and negative likelihood ratio (false negative count/true negative count) were computed.

The acquired results were further tested for differences between males and females and for age-related variation by Mantel test, Mann–Whitney U test and Wilcoxon signed-rank test.

Differences between meshes were visualized by means of color maps using the rainbow spectrum encoding system. Additional advanced visualization techniques employing transparency, contour rendering, sections and fog simulation were used. All functionalities are incorporated in FIDENTIS Analyst software.

4. Results

4.1. Mesh-based identification

Descriptive statistics for rank-1 identification together with classification rates are displayed in Table 2. M–W U test showed that for both measures of dissimilarity the average values were significantly lower for the true positive results than when the true negatives were called except for the comparisons based on meshes of 50 vertices (Table 3).

Values of closest point-to-point distances and their relevant univariate measures of dissimilarity were shown to be vastly dependent on mesh resolution (Figs. 2 and 3). Fig. 3 shows that for meshes of less than 500 vertices the distances associated with low-numbered ranks roughly correspond or fall into the range of the high-numbered ranks corresponding to the high-resolution meshes. Similarly, the discriminative power decreases as the mesh resolution deteriorates. However, Fig. 4 shows that the decrement in the accuracy rate is not as steep as the results could suggest and the high identification rate holds up to the mesh reduction of 500 vertices.

For the mesh resolution of 10k points, RMS-based algorithm provided 96.4% of correctly identified 3D faces. With the PERC75-based measure employed the algorithm succeeded to identify all individuals correctly. This identification rate was valid even after the mesh resolution was reduced to 5k and 1k points. For the PERC75-based algorithm the rate fell significantly when meshes were reduced to 100 points (78.6% of correctly identified cases) and continued to fall for further reduced meshes (42.9% of correctly identified cases). It also shows that RMS provided less consistent results than when PERC75 was used as the measure of differences.

Table 1 List of landmarks, their definitions and presence in configurations of descending number of landmarks as employed in landmark-based identification.

Configurations	Landmark	Definition
6L–42L	Gnathion (gn)[1]	The most anterior inferior point located in the midline at the lower margin of the mandible
6L–42L	Gonion (go)[1*]	The lateral inferior point located at the mandibular angle
6L–42L	Pogonion (pg)[1]	The most anterior midline point located at the chin with the head positioned in the Frankfurt horizontal plane
6L–42L	Zygion (zy)[1*]	The most lateral point of the face located on line running through the right and left eye centers
9L–42L	Glabella (g)[1]	The outermost midline point between the eyebrows
9L–42L	Tragion (t)[1*]	The point located at the upper margin of tragus in the little notch where the cartilage is attached while the head is positioned in the Frankfurt horizontal place
12L–42L	Sublabiale (sl)[1]	The midpoint of the mentolabial sulcus
12L–42L	Exocanthion (ex)[1*]	The point at the outer commissure of the eye where the outer margin of the upper eyelid meets the lower eyelid
15L–42L	Nasion (n)[1]	The deepest midline point of the nasal root
15L–42L	Palpebra superior (pas)[1*]	The point located at the intersection of a line going through the eye center (parallel to the mid-sagittal plane) and the caudal (lower) margin of the upper eyelid
18L–42L	Labrale inferius (li)	The most anterior midline point at the lower margin of lower vermilion (by Fetter [13], modified)
18L–42L	Palpebra inferior (pai)[1*]	The point which is located at the intersection of a line going through the eye center (parallel to the mid-sagittal plane) and the upper margin of the lower eyelid
21L–42L	Stomion (sto)[1]	The point located at the intersection of the closed mouth fissure and the midline
21L–42L	Cheilion (ch)[1*]	The point located at the labial commissure
24L–42L	Pronasale (prn)	The most anterior midline point of the nasal tip with the head positioned in the Frankfurt horizontal plane
24L–42L	Endocanthion (en)[1*]	The point at the inner commissure of the eye where the inner margin of the upper eyelid meets the lower eyelid
27L–42L	Labrale superius (ls)	The midpoint of the upper vermilion line
27L–42L	Crista philtri (cp)[1*]	The point located at the intersection of the horizontal line connecting the superior point of the upper vermilion and the lower margin of crista philtra
30L–42L	Subnasale (sn)[1]	The lowest posterior midline point at the angle formed by the outline of nasal septum and upper lip
30L–42L	Alare (al)[1*]	The most lateral anterior point of the wing of the nose
42L	Superaurale (sa)[1*]	The most superior point located at the upper margin of the auricle with the head positioned in the Frankfurt horizontal plane
42L	Subaurale (sba)[*]	The most inferior point located at the lower margin of the earlobe with the head positioned in the Frankfurt horizontal plane
42L	Postaurale (pa)[1*]	The most posterior point of the posterior margin of the auricle with the head is positioned in the Frankfurt horizontal plane
42L	Otobasion superius (obs)[1*]	The point where the upper margin of the auricle attaches to the head
42L	Otobasion inferius (obi)[1*]	The point where the earlobe attaches to the cheek skin (lower attachment)
42L	Praeaurale (pra)[1*]	The point at the intersection of the line between otobasion superius and otobasion inferius points, at the level of the postaurale point

[*] Bilateral landmarks.
[1] Definitions adapted from Fetter.[13]

If a non-match was met at the first rank, the positive identification was mostly achieved by the rank of 10. This was, however, valid only for the meshes of more than 500 vertices. Although the number of matches based on low-resolution meshes at the rank of 10 increased significantly comparing to the rank-1 rates, it could not match that of better-quality meshes. Many positive matches based on meshes of 50 points were even met beyond single digit ranks. In fact, the highest-ranked match for one of the probes was met at rank 208 and 41 for RMS and PERC75, respectively.

As expected, size-invariant meshes provided lower accuracy rates for both measures of dissimilarity than those by the original dataset. This applies particularly to the lower-resolution meshes. However, in agreement with the size-present meshes, using the high-quality meshes the algorithm was shown to perform excellently with both measures of dissimilarity capable of matching 100% of probes. By the resolution of 500 vertices, the PERC75-based metric was able to identify all probes correctly, which is consistent with the results on the original meshes. When tested by Wilcoxon signed-rank test raw point-to-point distances were significantly smaller in size-invariant comparisons than in the original comparisons (data not shown). Alike in the original meshes, raw point-to-point distances increased as the mesh resolution decreased (Table 2).

Mantel tests showed that both measures of dissimilarity yielded statistically significant positive correlations with sex and age (Table 4). While for the sex-related dependency R-values ranged from 0.32 to 0.41 and the differences were slightly more pronounced for RMS-based distance matrices, the age dependency provided R coefficients between 0.16 and 0.19 with higher values in PERC75-based metrics when computed on the high-resolution meshes. These results were also

Table 2 Rank-1 identification rate and descriptive statistics as recorded for the mesh-based testing.

	Resolution	RMS						PERC75					
		10k	5k	1k	5h	1h	5d	10k	5k	1k	5h	1h	5d
	Rank-1 identification (%)	96.43	96.43	100.00	96.43	96.43	57.14	100	100.00	100.00	100.00	78.57	42.86
	Rank-10 identification (%)	100.00	96.43	100.00	96.43	100.00	85.71	100.00	100.00	100.00	100.00	100.00	78.57
Original meshes	Means (rank 1)	1.230	1.611	2.925	3.939	7.193	9.926	1.289	1.712	3.233	4.384	7.913	10.907
Probes (tested sample) (in mm)	Max (rank 1)	2.613	3.252	4.362	5.478	9.118	13.967	1.745	2.346	4.005	5.082	9.976	14.083
	Min (rank 1)	0.951	1.269	2.573	3.645	6.211	7.456	1.085	1.446	2.845	4.064	6.363	8.112
	Means (matches only)	1.179	1.611	2.925	3.882	7.169	9.287	1.289	1.712	3.233	4.384	7.653	9.730
	Max (matches only)	1.599	3.252	4.362	4.229	9.118	11.806	1.745	2.346	4.005	5.082	8.551	12.503
	Min (matches only)	0.951	1.269	2.573	3.645	6.211	7.456	1.085	1.446	2.845	4.064	6.363	8.112
	Means (non-matches)	2.613	–	–	5.478	7.841	10.779	–	–	–	–	8.868	10.764
Control subset (in mm)	Means (rank 1)	4.172	4.724	6.329	6.428	10.889	16.319	4.408	5.099	7.125	7.115	12.510	13.968
	Max (rank 1)	12.921	18.137	17.018	13.195	19.232	29.628	15.699	22.949	21.498	15.629	24.539	24.010
	Min (rank 1)	1.577	1.868	3.435	4.115	7.152	8.813	1.689	2.118	3.833	4.684	7.838	9.087
	Rank-1 identification (%)	100.00	100.00	100.00	96.43	92.86	42.86	100.00	100.00	100.00	100.00	75.00	39.29
	Rank-10 identification (%)	100.00	100.00	100.00	100.00	92.86	75.00	100.00	100.00	100.00	100.00	92.86	64.29
Size-invariant meshes	Mean (rank 1)	1.141	1.617	2.916	3.919	7.195	9.432	1.243	1.726	3.252	4.391	7.819	10.194
Probes (tested sample) (in mm)	Max (rank 1)	1.665	2.859	4.097	5.162	8.861	11.573	2.354	2.337	3.866	5.278	9.748	12.458
	Min (rank 1)	0.922	1.273	2.573	3.646	6.221	7.522	1.017	1.448	2.931	4.097	6.133	8.006
	Means (matches only)	1.141	1.617	2.916	3.930	7.193	9.989	1.243	1.726	3.252	4.391	7.937	10.896
	Max (matches only)	1.665	2.859	4.097	5.283	9.318	13.839	2.354	2.337	3.866	5.278	10.837	13.896
	Min (matches only)	0.922	1.273	2.573	3.646	6.221	7.522	1.017	1.448	2.931	4.097	6.133	8.006
	Means (non-matches)	–	–	–	5.283	8.304	10.704	–	–	–	–	8.939	11.928
Control subset (in mm)	Means (rank 1)	3.131	3.543	4.916	5.204	9.109	11.782	3.251	3.762	5.474	5.788	10.380	13.598
	Max (rank 1)	8.200	8.838	9.892	9.337	15.997	17.857	8.733	9.679	12.217	10.449	17.215	22.028
	Min (rank 1)	1.492	1.805	3.101	3.825	6.635	7.936	1.552	1.976	3.479	4.348	7.422	7.999

Table 3 Results of M–W U test comparing values of dissimilarity measures for the tested and control subsets at rank of 1.

	Measure of dissim.	N° of vertices	Rank Sum (control)	Rank Sum (tested)	U	Z	p-value
Original meshes	RMS	5d	132,798	6858	6452	0.697	0.486
		1h	137,518	2138	1732	6.705	0.000
		5h	138,832	824	418	8.378	0.000
		1k	139,206	450	44	8.854	0.000
		5k	138,584	1072	666	8.062	0.000
		10k	139,246	410	4	8.905	0.000
	PERC75	5d	132,414	7242	6836	0.208	0.835
		1h	137,580	2076	1670	6.784	0.000
		5h	138,940	716	310	8.515	0.000
		1k	139,250	406	0	8.910	0.000
		5k	139,167	489	83	8.804	0.000
		10k	139,250	406	0	8.910	0.000
Size-invariant meshes	RMS	5d	132,162	7494	6912	−0.111	0.911
		1h	138,150	1506	1100	7.510	0.000
		5h	139,071	585	179	8.682	0.000
		1k	139,244	412	6	8.902	0.000
		5k	139,192	464	58	8.836	0.000
		10k	139,055	601	195	8.662	0.000
	PERC75	5d	130,636	9020	5386	−2.054	0.040
		1h	138,176	1480	1074	7.543	0.000
		5h	138,986	670	264	8.574	0.000
		1k	139,250	406	0	8.910	0.000
		5k	139,222	434	28	8.874	0.000
		10k	139,250	406	0	8.910	0.000
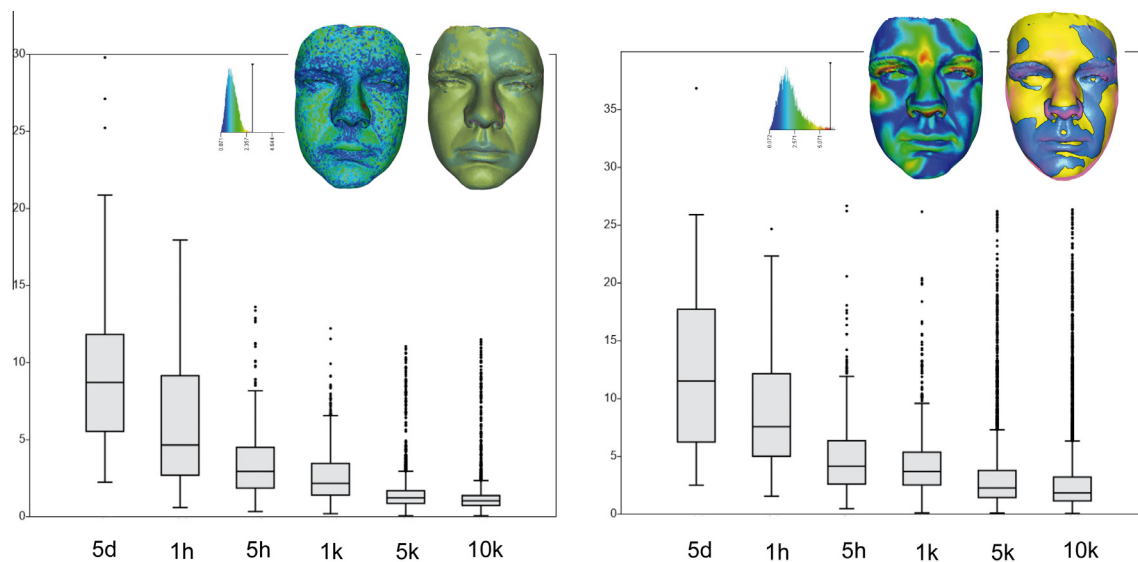
Figure 2 Values of closest point-to-point distances for a true positive and a true negative result at various mesh resolutions accompanied by 3D visualizations depicting local dissimilarities (results for meshes of 10k displayed).

projected onto the demographic profile of the mismatches. If a mismatch was met at rank-1 in meshes of 100 points and higher, then either all (RMS-based metrics) or the majority (PERC75-based metrics) of the false positive results fell into the same sex and age category. For meshes of the lowest resolution (meshes of 50 vertices), the mismatches were less respectful of the same sex and age group. For RMS-based metrics, 2 male probes were falsely identified as female targets at the rank of 1, with one of them also being misclassified into the older neighboring age category, whereas for PERC75-based data two female probes were misclassified as males, albeit of the adequate age category. The results obtained for scaled models concurred with those by the original meshes.

The ROC curve analysis provided results relatively consistent with those by rank-1 approach (Table 5, Figs. 5 and 6).
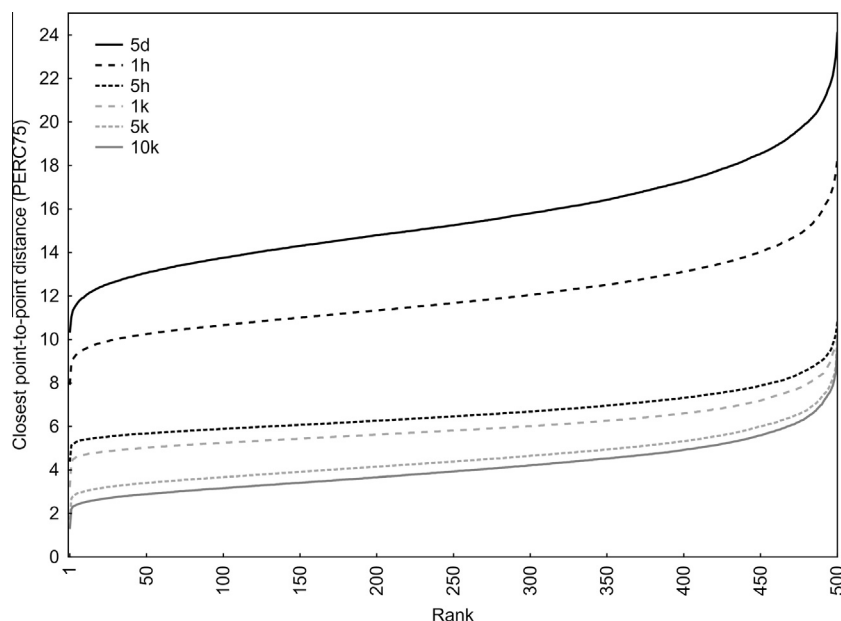
Figure 3 Relationships between ranks and PERC75-based metrics averaged for the tested sample and displayed for various mesh resolutions.
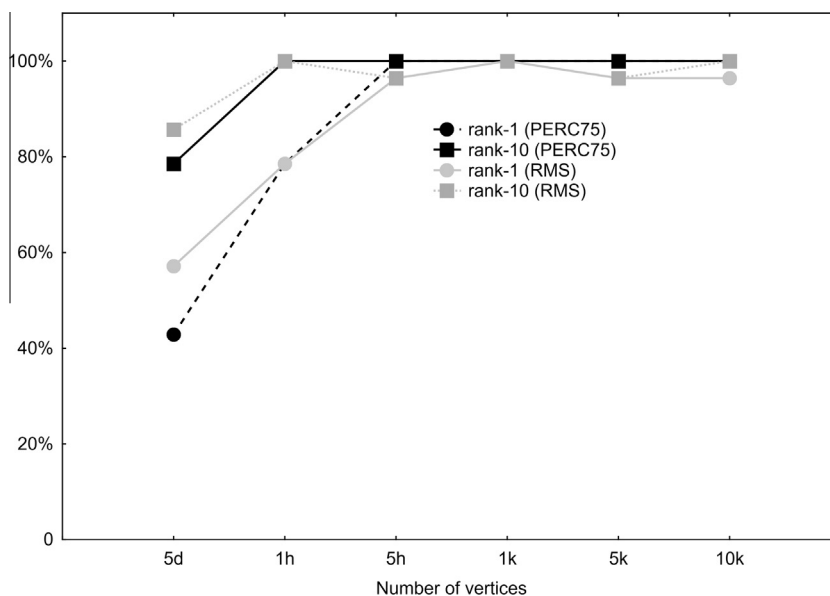
Figure 4 Rank-1 and rank-10 identification rates for RMS and PERC75-based metrics.

The low-resolution meshes provided poor identification rates while metrics derived from the meshes of 10k vertices both identified all target faces correctly. If sorted by performance the best result was shared between the PERC75-based metrics for 10k and 1k mesh resolution. For the size-present meshes, the lowest identification rate was recorded for the PERC75-derived metric in meshes of 50 points. If the point closest to the top left corner was selected as the cut-off value the discriminant power translated into sensitivity of 0.464, specificity of 0.688, positive likelihood ratio of 2.165 and negative likelihood ratio of 0.9177. If, however, either sensitivity or specificity was maximized, it can be seen that even though for the PERC75-based metric the results concurred with those by the original balancing criterion, for the remaining scenarios the values of specificity or sensitivity were maximized at the expense of the opposite. For size-invariant meshes, PERC75-derived metric computed on low-quality meshes yielded the lowest identification rates with a sensitivity of 0.278 and a specificity of 0.464. This translated into positive likelihood ratio of 0.860 and negative likelihood of 1.680.

Of the tested probes, one male 3D facial scan was repeatedly misidentified independent of the mesh resolutions. The graphic visualization between corresponding faces and the face demonstrated that the variation was concentrated in the region under the left mandibular line and can be assigned to technical rather than morphological inconsistencies (Fig. 7).

Table 4 *R*-values provided by the Mantel test. All *R*-values were statistically significant at *p*-value = 0.0001.

N° of vertices	RMS		PERC75	
	Sex	Age	Sex	Age
5d	0.3889	0.1709	0.3702	0.1685
1h	0.4063	0.1822	0.3964	0.1681
5h	0.4066	0.1698	0.399	0.1784
1k	0.3352	0.1569	0.315	0.1615
5k	0.4086	0.1756	0.3914	0.1855
10k	0.4108	0.1799	0.3885	0.1905

4.2. Landmark-based identification

Descriptive statistics for rank-1 identification together with classification rates are displayed in Table 6. No trend pointing out to an overall relationship between the number of landmarks and Procrustes distances at the rank of 1 was observed (Fig. 8). Although larger variance was revealed at the lower ranks suggesting that Procrustes distances increased as the number of landmarks decreased, at higher ranks this trend was absent and the values of the Procrustes distance were relatively independent of the number of landmarks.

M–W *U* test showed that there were statistically significant differences between average Procrustes distances of the tested and the control subsets at rank-1 in all landmark configurations except for the minimum tested number of 6 landmarks (Table 7).

Similar to the mesh-based metrics, the discriminative power of the extracted measure of dissimilarity decreased with the number of quantitative traits, i.e., landmarks (Fig. 9). The rank-1 identification rate of 100% was acquired for the configurations counting 12 and more landmarks. Starting with the set of 12 landmarks the true positive rate decreased to 86% and then to 57% for the 6-point configuration. For the rank-10 identification, the rates remained high, decreasing only for the configurations of 9 and 6 landmarks.

Similar to the mesh-based data, the Procrustes distances were shown to correlate positively with sex and age differences (Table 8). Although statistically significant at 5% level of significance for all tested landmark sets the correlation coefficient did not exceed the value of 0.06 and 0.21 for sex and age respectively. This suggests mild to weak relationships. The correlations increased as the number of landmarks decreased.

The ROC curve analysis, again, provided results relatively consistent with those by rank-1 approach (Table 9). The AUC values fell gradually relative to the decreasing number of landmarks. Surprisingly, the complete set of landmarks yielded a sensitivity of 0.998 at a specificity of 0.96, which was the fifth best performance out of the tested configurations. Once the auricular landmarks were removed the specificity increased to 0.996 at a sensitivity of 1.00. If the point closest to the top left corner was selected as the cut-off value the discriminative power of the 12-landmark configuration which held the rank-1 rate of 100% revealed a specificity of 0.857 and a sensitivity of 0.871. These values translated into positive likelihood ratio of 7.9 and negative likelihood ratio of 0.28. Ultimately, the minimum 6-point configurations provided the

Table 5 ROC curve statistics for mesh-based identification.

	Mesh resolution	Metric	Area under the curve			Cut-off value	Sensitivity	Specificity	LR+	LR−	Sensitivity of 1	Specificity of 1
			Area	SE	95% Confidence interval							
1	10k	PERC75	1.0000	0.000	1.000–1.000	1.61	1.000	1.000	INF	0	1.00	1.00
1	10k	PERC75 scaled	1.0000	0.000	1.000–1.000	1.53	1.000	1.000	INF	0	1.00	1.00
1	1k	PERC75	1.0000	0.000	1.000–1.000	3.80	1.000	1.000	INF	0	1.00	1.00
1	1k	PERC75 scaled	1.0000	0.000	1.000–1.000	3.45	1.000	1.000	INF	0	1.00	1.00
5	1k	RMS scaled	1.0000	0.000	1.000–0.999	3.14	1.000	0.992	125	0	0.99	0.93
6	10k	RMS	0.9997	0.000	0.999–1.000	1.60	0.992	0.992	125	0	0.99	0.96
7	1k	RMS	0.9969	0.003	0.990–1.000	3.46	0.893	0.893	241.1	0.036	0.92	0.93
8	5k	PERC75 scaled	0.9980	0.001	0.996–1.000	2.21	0.964	0.980	50	0	0.98	0.86
9	10k	RMS scaled	0.9960	0.003	0.990–1.000	1.63	0.964	0.952	20.09	0.037	0.93	0.93
10	5k	RMS scaled	0.9960	0.002	0.991–1.000	2.01	0.964	0.970	32.14	0.037	0.95	0.79
11	5k	PERC75	0.9941	0.005	0.980–1.000	2.33	0.986	0.986	60.27	0.036	0.85	0.93
12	5h	RMS scaled	0.9870	0.008	0.971–1.000	3.9	0.929	0.996	232.1	0.072	0.82	0.79
13	5h	PERC75 scaled	0.9810	0.013	0.955–1.000	4.53	0.893	0.946	17.2	0.076	0.63	0.82
14	5h	PERC75	0.9779	0.012	0.953–0.998	4.73	0.988	0.988	74.4	0.108	0.79	0.68
15	5h	RMS	0.9701	0.012	0.942–0.991	4.32	0.944	0.944	14.67	0.189	0.80	0.71
16	5k	RMS	0.9524	0.029	0.893–0.998	2.06	0.972	0.972	32.97	0.147	0.31	0.71
17	1h	PERC75 scaled	0.9230	0.030	0.865–0.982	7.92	0.893	0.812	4.94	0.088	0.252	0.50
18	1h	RMS scaled	0.9210	0.036	0.850–0.992	7.00	0.857	0.912	9.52	0.157	0.12	0.46
19	1h	PERC75	0.8807	0.035	0.802–0.940	8.66	0.868	0.868	5.682	0.288	0.20	0.32
20	1h	RMS	0.8763	0.041	0.397–0.682	7.95	0.834	0.834	4.948	0.214	0.18	0.25
21	5d	RMS	0.5391	0.072	0.370–0.654	9.89	0.778	0.778	1.931	0.735	0.00	0.04
22	5d	PERC75	0.5117	0.071	0.370–0.654	10.9	0.464	0.688	2.165	0.918	0.00	0.00
23	5d	RMS scaled	0.4940	0.072	0.353–0.635	9.66	0.268	0.536	1.102	0.812	0.00	0.04
24	5d	PERC75 scaled	0.3850	0.070	0.248–0.522	10.40	0.278	0.464	0.860	1.680	0.00	0.07
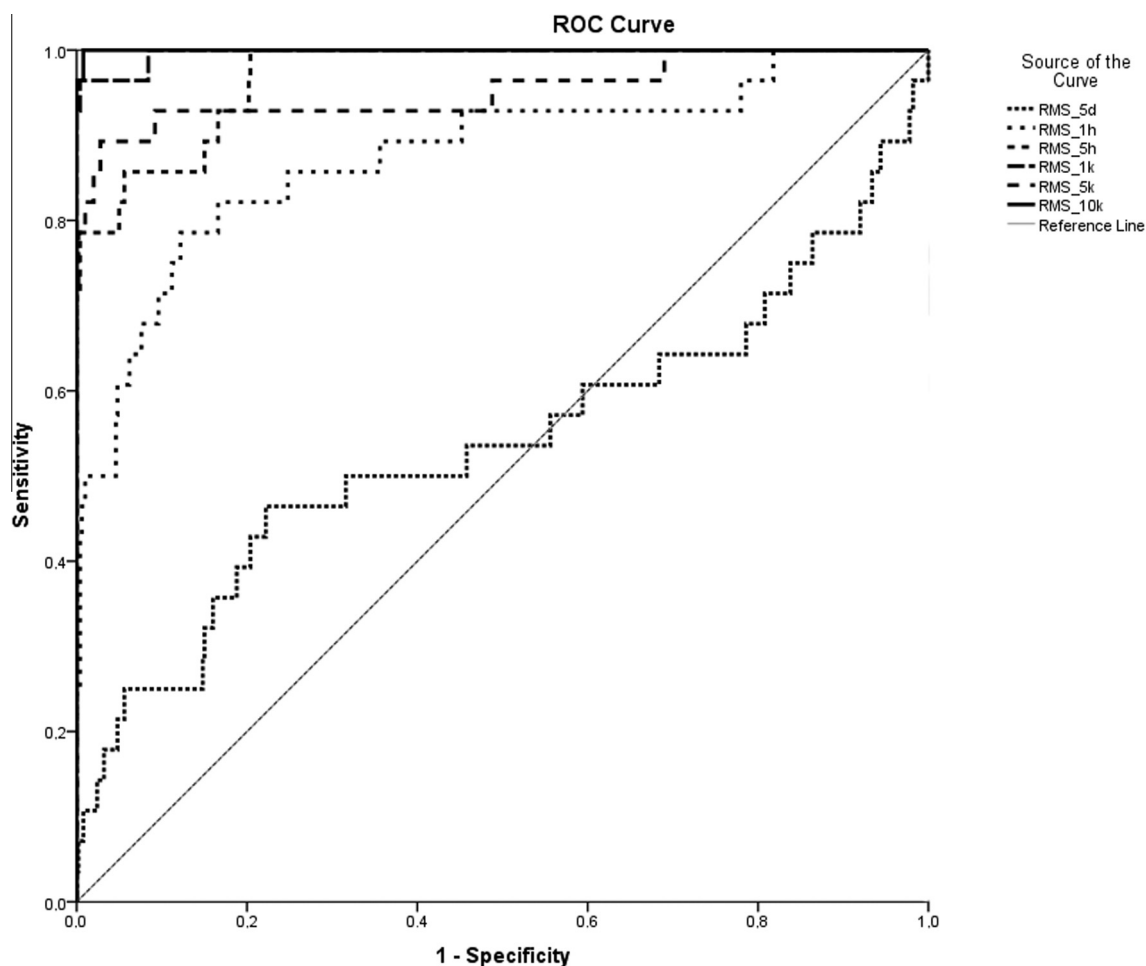
Figure 5   ROC curves illustrating performances of RMS-based measures of dissimilarity on the studied sample.

lowest values of sensitivity while the specificity reached values little below the configuration of 12 landmarks and relatively consistent with the configuration of 9 landmarks.

The results also showed that the alternative 6-point configuration performed better than the original one achieving the sensitivity of 0.632, the specificity of 0.821, which both translated into scores of 3.54 and 0.45 for positive and negative likelihood ratios, respectively. If the preference was given to maximize the values of specificity than none of the configurations provided values of sensitivity higher than 0 except for the 9-point (specificity of 0.43) and 12-point configurations (specificity of 0.21). Values of specificity under the acceptance rate of 0.99 are displayed in Table 9 (see Fig. 10).

5. Discussion

In the framework of forensic sciences, facial images represent important conveyors of information about a character of a crime, perpetrator's appearance or victim's identity. Although forensic facial identification is conducted in majority on traditional photographs or surveillance videos in the recent years a vast number of devices capable of capturing 3D facial morphology have been developed[43] and gradually introduced to the forensic community.[22,26,27] Currently, 3D images are mostly recorded under controlled conditions using in-lab

equipment, but portable cost-efficient 3D recording systems have been also made available for field work.[33] Furthermore, systems designed to accessorize mobile devices such as iPads, cell phones or game consoles, where a stable relationship between a capturing system and a captured subject is not a primary concern, have occurred on markets.[44,45] With the rise of novel 3D technology it has become reasonable to question conditions, under which 3D images can yield accurate and reliable forensic identifications.

In the present paper we explored the effect of mesh resolution given by a number of vertices – the property associated with overall quality of 3D images and the effect of a total count of identity-encoding features derived from these images. Two approaches to quantifying facial features were undertaken. The mesh-based analysis is grounded on the computation of shell-to-shell deviations and is in concordance with current trends in other fields.[34,35] The landmark-based analysis featuring discrete facial points has its origin in the traditional photo-anthropometric approach.

It can be concluded that employing high-resolution meshes and the maximum number of collected facial points both approaches yielded excellent identification rates through their relevant dissimilarity measures. This should not come as a surprise given the fact that the tested 3D meshes were acquired in a controlled in-lab acquisition environment and all 3D models
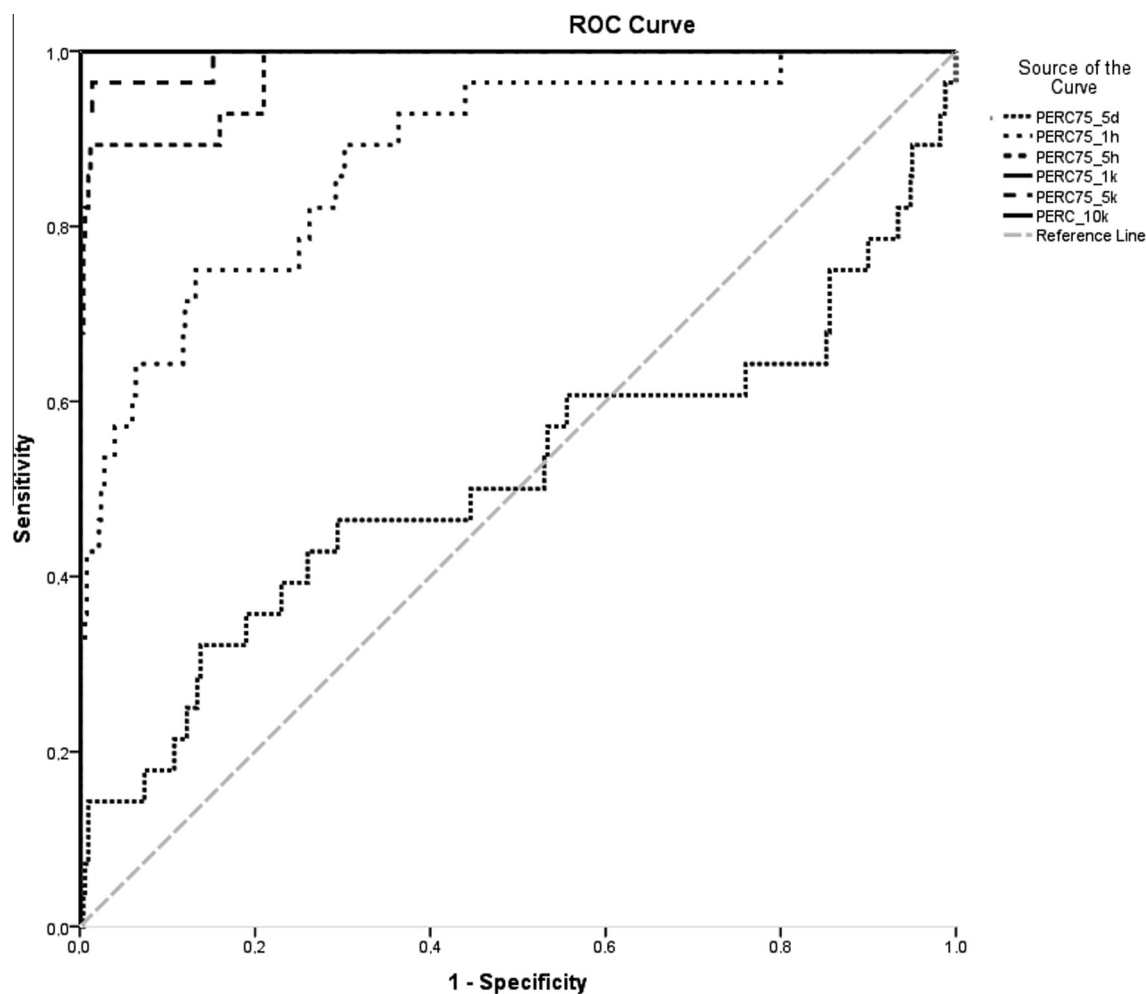
Figure 6 ROC curves illustrating performances of PERC75-based measures of dissimilarity on the studied sample.

were pre-screened in order to exclude faces with facial expressions, occlusion or additional technical irregularities – factors that impose difficulties on automated as well interactive identification tasks.[23] In addition, morphological changes over time have been acknowledged as serious sources of errors. For real world cases a time lapse that passes between two or multiple images submitted for forensic examination can vary significantly. The present results were derived from matching probes and targets recorded only few minutes apart. Therefore, little to none of the intra-individual differences can be contributed to age variations.

More importantly, it was shown that a configuration of 12 points matched successfully all probes at rank-1. This concurs with Spreeuwers[30] who showed that to obtain a recognition rate of 80% for false acceptance rate of 0.1% low-dimensional vectors composed of 12 variables were sufficient. For facial image comparison executed by forensic experts or simple photograph verification carried out by police or security officers, time pressure has been shown to significantly diminish performances.[46] Therefore, knowledge of minimum requirements and a proper selection of facial features capable of achieving robust performance can significantly speed up identification tasks.

In the present study, two minimum landmark configurations were tested. Both of them composed of 6 landmarks. This means that each 3D face was described by 18-dimensional vec-

tors. While the first configuration was derived through the consecutive removal of points with the minimum level of inter-individual variance, the alternative 6-point configuration was selected based on the assumption that the selected points describing positions of eyes, nose and mouth were the foundation of inter-individual differences.[47] Both minimum configurations performed rather poorly with less than 80% of correctly identified probes. Still, the alternative configuration yielded slightly higher rates, which can also be explained by a lower level of acquisition error generally reported for these landmarks. It ought to be emphasized that the obtained results represent an improvement in comparison to those by Kleinberg et al.[9] who reported identification rates as low as 20% for indices derived from a nearly identical set of endpoint (nasion was used instead of pronasale) measured on traditional 2D photographs.

Generally, more consistent results were acquired for the mesh-based approach than for the landmarks. This was especially valid for high-resolution models. It was shown that a single descriptor extracted from 500 and more point-to-point distances was able to identify all probes correctly at the rank of 1 and with the identification rate of 97.8% when the threshold balancing false acceptance and rejection rates (also known as equal error rate) was employed. The values of likelihood ratios suggest that the results could offer "moderate support"
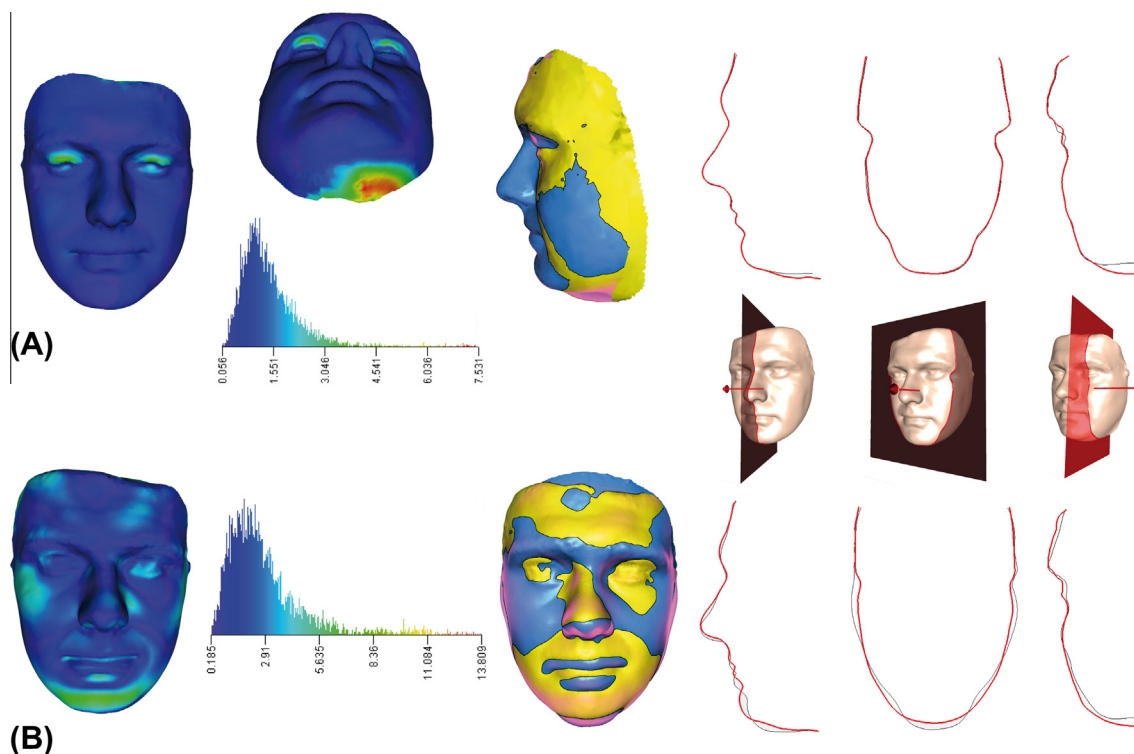
Figure 7 Illustrating pitfalls of 3D face identification based on a single dissimilarity measure. (A) Shows a false negative result at the rank of 1 due to highly localized differences – artifacts of data acquisition. (B) Shows a false positive result at the rank of 1.

Table 6 Rank-1 identification rates and descriptive statistics grounded on landmark-based comparison.

		PD 42L	PD 30L	PD 15L	PD 12L	PD 9L	PD 6L	PD 6L alt
	Rank-1 identification (%)	100.00	100.00	100.00	100.00	89.29	57.14	75.00
	Rank-10 identification	100.00	100.00	100.00	100.00	96.43	96.43	96.43
	Mean (rank 1)	0.0299	0.034	0.033	0.034	0.035	0.036	0.022
Tested subset	Max (rank 1)	0.0503	0.044	0.044	0.046	0.048	0.055	0.038
	Min (rank 1)	0.0207	0.020	0.019	0.018	0.018	0.019	0.011
	Mean (matches only)	0.0299	0.034	0.033	0.034	0.034	0.035	0.020
	Max (matches only)	0.0503	0.044	0.044	0.046	0.048	0.052	0.026
	Min (matches only)	0.0207	0.020	0.019	0.018	0.018	0.019	0.011
	Mean (non-matches)	–	–	–	–	0.041	0.039	0.030
Control subset	Mean (rank 1)	0.0275	0.058	0.049	0.046	0.043	0.034	0.014
	Max (rank 1)	0.0928	0.089	0.074	0.072	0.070	0.065	0.029
	Min (rank 1)	0.0205	0.040	0.034	0.033	0.029	0.013	0.005

to the hypothesis that two scans depict the same individual.[48] Conversely, in the landmark-based analysis the number of 12 points, the lowest-numbered configuration to provide the rank-1 identification rate of 100% yielded likelihood ratios that are far from being considered comparable to the mesh-based results. With the positive likelihood ratio of 7.9 the verbal equivalent to support the origin from the same individual would be weak to limited.

Nevertheless, matching high-resolution 3D models is a computationally expensive task and the processing speed is generally viewed as a trade-off for robust performance. The recording system utilized in the present study produces 3D facial scans of 40,000 and more vertices. Our previous experiences with 3D data processing[41,49] showed that the reduction

to the constant number of 10,000 points provided an acceptable level of data volume together with a welcome decrement of computing power requirements that did not sacrifice the precision in face geometry. Still, using a standard hardware equipment (laptop, 2.4 GHz, 16 GB RAM) it took almost an hour to match a probe against 501 candidates. For the meshes of 1,000 and 500 points the time requirements fell drastically to 3 min and to less than 2 min, respectively. For 3D faces of 100 points it took less than 40 s to acquire results. To ease the computing requirements many proposed algorithms convert textured 3D meshes into depth image, depth maps, XYZ maps or 2.5D representations via orthographic projection.[44,50] However, this may not be the most appropriate solution in all forensic and security applications as the conversion discards
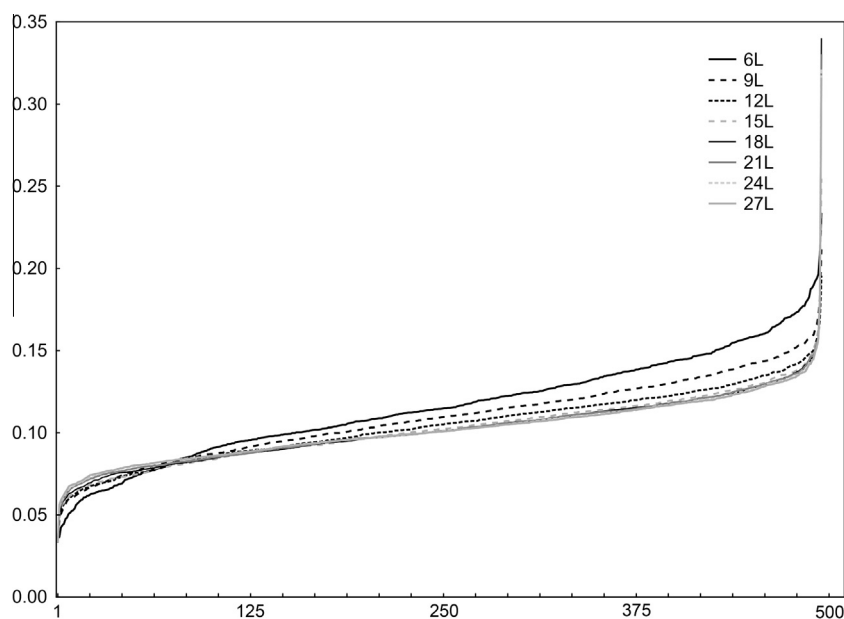
Figure 8 Relationships between ranks and Procrustes distances averaged for the tested sample and displayed for various numbers of landmarks.

Table 7 Results of M-W U test comparing values of dissimilarity measures for tested (matching pairs) and random (rank 1) subsets. Results of M-W U test comparing values of landmark-based dissimilarity measures for tested (matching pairs) and random (rank 1) subsets.

	Rank Sum (control)	Rank Sum (tested)	U	Z	p-value
6L	8865	128,685	5429	1.943	0.052
9L	3227	134,323	2821	−5.289	0.000
12L	1504	136,046	1098	−7.499	0.000
15L	922	136,104	516	−8.244	0.000
18L	656	136,370	250	−8.586	0.000
21L	477	136,549	71	−8.816	0.000
24L	420	136,606	14	−8.889	0.000
27L	418	136,608	12	−8.892	0.000
30L	414	137,136	8	−8.898	0.000
42L	538	138,063	132	−8.740	0.000
6L alt	3967	132,536	3561	−4.320	0.000

a substantial amount of the original 3D information and it may also lead to a significant increase in error rates.[51]

To ease additional computations the present mesh-based approach used univariate global representations of similarity/dissimilarity represented by the root mean square and 75th percentile retrieved from a set of inter-point distances. By doing so, a multidimensional vector was reduced to a single number. The root mean square is a commonly used indicator of geometric deviations in 3D surface data. The 75th percentile computed from raw surface-to-surface distances was previously demonstrated as the most reliable statistical descriptor in 3D facial comparisons.[52] For both parameters, the cut-off value between matching and non-matching 3D faces counted less than 2 mm. Interestingly, it is a comparable or even a smaller distance than that reported in other studies. For instance, for superimposed photographic evidence, Yoshino et al.[53] state that a mean distance of 2.5 mm between corresponding anthropometric points is an adequate threshold

to achieve positive identification. Similarly, Sala et al.[54] assessed identification in 2D to 3D images by a threshold set to 1.8 mm.

Our results also demonstrated that for a given mesh resolution and dissimilarity metric the original meshes outperformed the scaled facial scans. Scaling means that the measures of dissimilarity do not include information about size-related differences. Accordingly, Procrustes distances represented shape distances with no variance accountable for size. Size plays an important role in the formation of sex-related as well as between-person differences encoded in facial features.[26] Interestingly, no differences in sex-related misclassifications were observed between scaled and original meshes. This suggests that extracting size variation does not bias the identification algorithm to the presumed extent.

From the practical point of view the major distinction between the tested approaches was the amount of interactions that was required to process 3D data. While in order to process
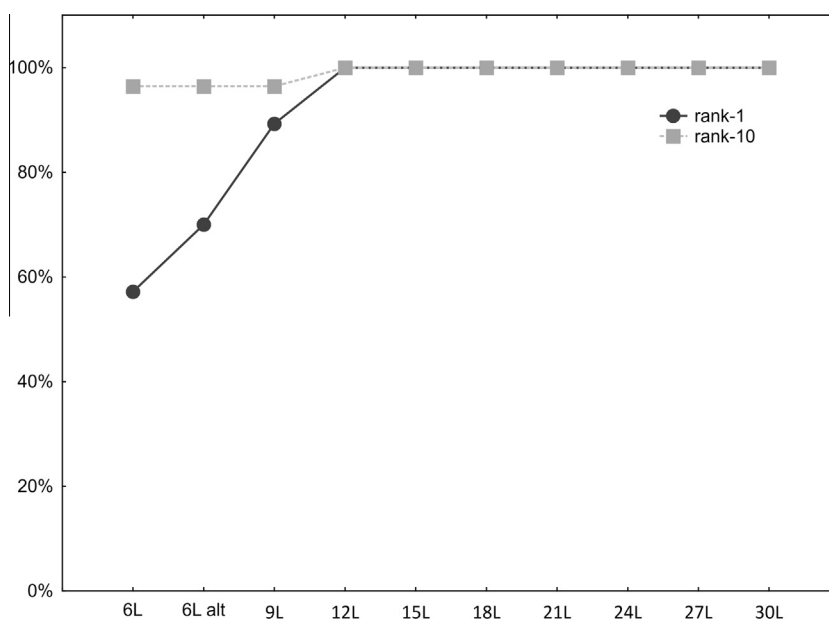
Figure 9 Rank-1 identification rates for landmark-based metrics and varying numbers of landmarks.

Table 8 *R*-values provided by the Mantel test. All *R*-values were statistically significant at *p*-value = 0.0001.

N° of landmarks	Sex	Age
6L	0.0273	0.0686
9L	0.0404	0.1016
12L	0.0503	0.1116
15L	0.0520	0.1361
18L	0.0546	0.1452
21L	0.0528	0.1667
24L	0.0570	0.1757
27L	0.0576	0.1900
30L	0.0581	0.2025
42L	0.0517	0.2279

the complete 3D surface meshes basic manual editing and automatic error checking had to be carried out the process of collecting 3D landmarks was far more laborious. It has to be noted that in contrast to mesh processing this step requires anthropological knowledge and specific training. Otherwise, collected data are subject to major errors. To facilitate the process landmark detection systems algorithms based on geometric features, such as symmetry, curvatures or edges, templates or statistical shape models have been proposed.[55–58] These methods were shown to be of assistance while dealing with a large amount of image evidence, for instance, while retrieving faces from a database,[6] but so far even the most reliable automatic landmark detection systems require some sort of expert's interactions.

Table 9 ROC curve statistics for landmark-based identification.

	Test result variable(s)	Area under the curve			Cut-off value	Sensitivity	Specificity	LR +	LR−	Sensitivity of 0.99	Specificity of 1
		Area	SE	95% Confidence interval							
1	30L	0.9994	0.001	0.998–1	0.0446	1.00	0.996	247.5	0.0	1.00	0.99
2	27L	0.9991	0.001	0.997–1	0.0439	1.00	0.988	82.5	0.0	0.96	0.99
3	24L	0.9990	0.001	0.997–1	0.0429	1.00	0.988	79.55	0.036	0.93	0.99
4	21L	0.9949	0.003	0.998–0.999	0.0432	0.964	0.947	18.36	0.038	0.86	0.95
5	42L	0.9885	0.009	0.968–1	0.0203	0.998	0.964	27.94	0.002	0.96	0.73
6	18L	0.9820	0.009	0.962–0.996	0.0406	0.893	0.952	18.42	0.113	0.79	0.82
7	15L	0.9628	0.015	0.927–0.987	0.0404	0.893	0.913	10.28	0.117	0.57	0.70
8	12L	0.9209	0.026	0.864–0.965	0.0388	0.857	0.871	7.899	0.276	0.50	0.49
9	9L	0.7969	0.042	0.708–0.877	0.0390	0.714	0.697	2.357	0.410	0.29	0.26
10	6L alt	0.7411	0.053	0.639–0.843	0.0124	0.632	0.821	3.541	0.448	0.07	0.07
11	6L	0.3909	0.053	0.292–0.483	0.0359	0.0571	0.681	0.0	1.021	0.00	0.01
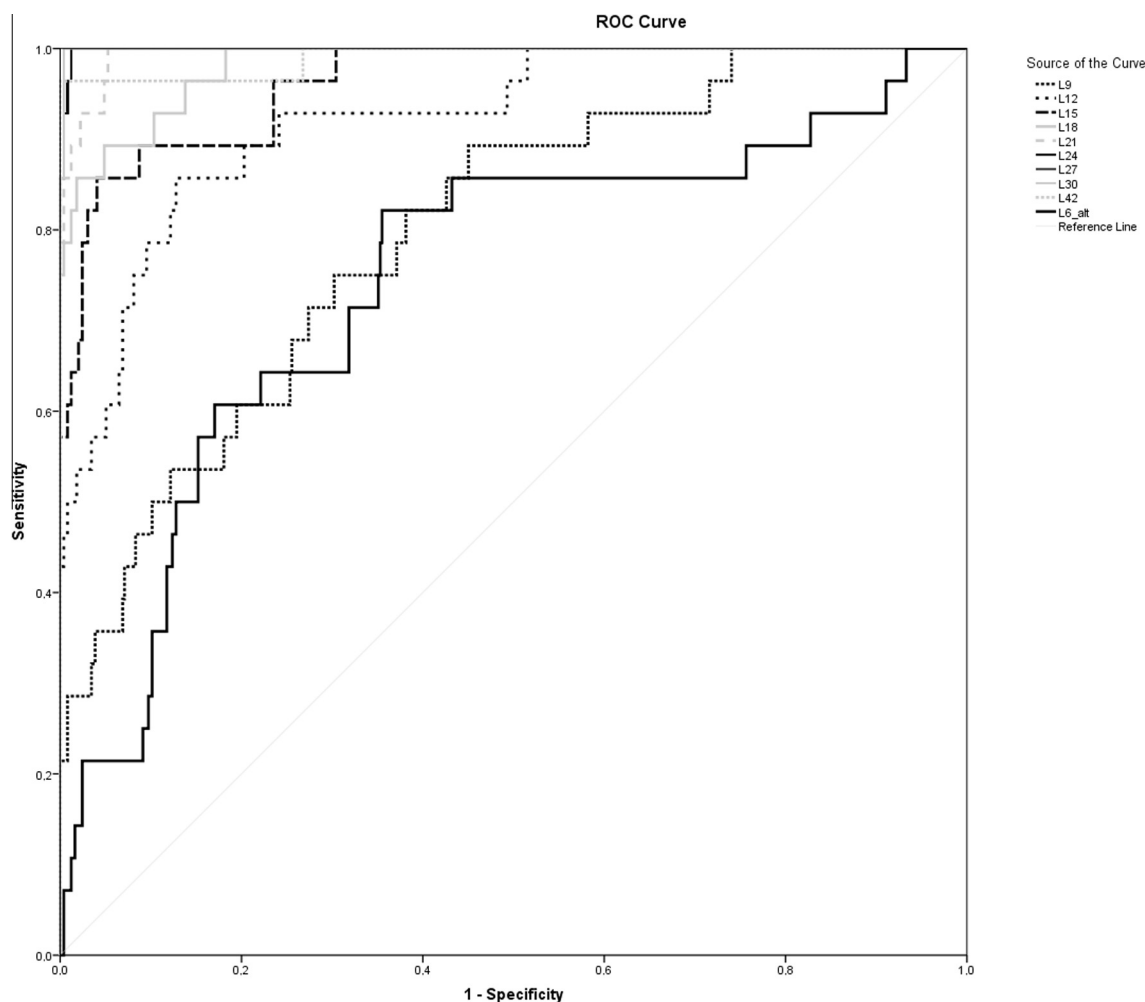
Figure 10 ROC curves illustrating performances of Procrustes distances on the studied sample.

6. Conclusion

Like other techniques employed in the legal context an algorithm utilized for matching 3D digital evidence must be based on quantifiable scientific principles with known error rates. To date, there has not been sufficient scientific evidence supporting the usage of three-dimensional records of human face for forensic identification purposes. The present results showed that under controlled conditions univariate dissimilarity measures retrieved from closest point-to-point distances and landmarks collected manually on 3D meshes were capable of providing very accurate and reliable results. The error rates are, however, highly dependent on mesh resolution and optimal number of landmarks, which may also impose increased computing, time and labor requirements on identification tasks. In following research further effort should be aimed at exploring performances related to a semi-controlled or ad-hoc acquisition environment.

Funding

None.

Ethical approval

The project was approved by the Ethics Committee at Faculty of Science, Masaryk University.

Conflict of interest

None declared.

References

1. Tian T, Kanade J, Cohn J. *Handbook of face recognition*. Springer; 2005.
2. Wilkinson C, Evans R. Are facial image analysis experts any better than the general public at identifying individuals from CCTV images? *Sci Justice* 2009;**49**:191–6.
3. Kemp R, Towell N, Pike G. When seeing should not be believing: photographs, credit cards and fraud. *Appl Cogn Psychol* 1997;**11**(3):211–22.
4. Yoshino M, Noguchi K, Atsuchi M, Kubota S, Imaizumi K, Thomas CD, Clement JG. Individual identification of disguised faces by morphometrical matching. *Forensic Sci Int* 2002;**127**(1–2):97–103.

5. Bruegge RV. Why automated face recognition cannot be used to eliminate suspects. *Proceedings of the american academy of forensic sciences 66th annual scientific meeting, Seattle, February 17–22*. p. 160–1.

6. Gupta S, Markey MK, Bovik AC. Anthropometric 3D face recognition. *Int J Comp Vision* 2010;**90**(3):331–49.

7. Zhang X, Gao Y. Face recognition across pose: a review. *Pattern Recognit* 2009;**42**:2876–96.

8. Busking S, Botha CP, Ferrarini L, Milles J, Post FH. Image-based rendering of intersecting surfaces for dynamic comparative visualization. *Vis Comp* 2011;**27**:347–63.

9. Kleinberg KF, Vanezis P, Burton AM. Failure of anthropometry as a facial identification technique using high-quality photographs. *J Forensic Sci* 2007;**52**(4):779–83.

10. Davis JP, Valentine T, Wilkinson C. Facial image comparison. In: Wilkinson C, Rynn C, editors. *Craniofacial identification*. Cambridge University Press; 2012. p. 136–53.

11. İşcan MY, Loth SR. Photo image identification. In: Knupfer JG, Saukko P, editors. *Siegel*. USA.: Encyclopaedia of Forensic Science, Academic Press; 2000. p. 795–807.

12. Aßmann S, Nohrden D, Schmitt R, Gabriel P, Ritz-Timme S. *Anthropological Atlas of male facial features*. Frankfurt: Verlag für Polizeiwissenschaft; 2008.

13. Fetter V. *Antropologie*. Academia; 1967.

14. Hulanicka B. Anthroposcopic features as a measure of similarity. *Materialy i prace antropologiczne* 1973;**87**:115–53.

15. Urbanová P, Čuta M, Kotulanová Z, Chalás I. *Somatoskopické znaky člověka*, 2010, <http://www.sci.muni.cz/somatoskopie>.

16. Guidelines for Facial Comparison Methods 1.0. Facial Identification Scientific Working Group (FISWG), <https://www.fiswg.org/document/viewDocuments,>; 2012accessed 02.02.12.

17. Porter G, Doran G. An anatomical and photographic technique for forensic facial identification. *Forensic Sci Int* 2000;**114**(2):97–105.

18. Roelofse MM, Steyn M, Becker PJ. Photo identification: facial metrical and morphological features in South African males. *Forensic Sci Int* 2008;**177**(2–3):168–75.

19. Davis JP, Valentine T, Davis RE. Computer assisted photo-anthropometric analyses of full-face and profile facial images. *Forensic Sci Int* 2010;**200**(1–3):165–76.

20. Halberstein RA. The application of anthropometric indices in forensic photography: three case studies. *J Forensic Sci* 2001;**46**(6):1438–41.

21. Bowyer KW, Chang K, Flynn P. A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition. *Comp Vis Image Understanding* 2006;**101**:1–15.

22. Smeets D, Claes P, Vandermeulen D, Clement JG. Objective 3D face recognition: evolution, approaches and challenges. *Forensic Sci Int* 2010;**201**(1–3):125–32.

23. Abate AF, Nappi M, Riccio D, Sabatino G. 2D and 3D face recognition: a survey. *Pattern Recognit Lett* 2007;**28**:1885–906.

24. Abdelwahab M, El-hamid Fawzi A, El-bakry HM, El-dosuky MA. 3D face identification: a survey. *Int J Inf Sci Intell Syst* 2014;**31**(1):121–36.

25. Tangchaitrong K, Messer LB, Thomas CD, Townsend GC. Fourier analysis of facial profiles of young twins. *Am J Phys Anthropol* 2000;**113**(3):369–79.

26. Evison M, Dryden I, Fieller N, Mallett X, Morecroft L, Schofield RV, Bruegge RV. Key parameters of face shape variation in 3D in a large sample. *J Forensic Sci* 2010;**55**(1):159–62.

27. Cattaneo C, Cantatore A, Ciaffi R, Gibelli D, Cigada A, De Angelis D, Sala R. Personal identification by the comparison of facial profiles: testing the reliability of a high-resolution 3D–2D comparison model. *J Forensic Sci* 2012;**57**(1):182–7.

28. Zelditch ML, Swiderski DL, Sheets HD, Fink WL. *Geometric morphometrics for biologists: a primer*. New York: Elsevier Academics Press; 2004.

29. Kau CH, Richmond S, Zhurov AI, Knox J, Chestnutt I, Hartles R, Playle R. Reliability of measuring facial morphology with a 3-dimensional laser scanning system. *Am J Orthod Dentofacial Orthop* 2005;**128**(4):424–30.

30. Spreeuwers L. Fast and accurate 3D face recognition. *Int J Comput Vis* 2011;**93**:389.

31. Besl PJ, McKay ND. A method for registration of 3-D shapes. *IEEE Trans Pattern Anal Mach Intell* 1992;**14**(2):239–56.

32. Granger S, Pennec X. Multi-scale EM-ICP: a fast and robust approach for surface registration. *European conference on computer vision (ECCV 2002), 2353 of LNCS*, p. 418–32.

33. Urbanová P, Hejna P, Jurda M. Testing photogrammetry-based techniques for three-dimensional surface documentation in forensic pathology. *Forensic Sci Int* 2015;**250**(1):77–86.

34. Kau CH, Richmond S, Savio C, Mallorie C. Measuring adult facial morphology in three dimensions. *Angle Orthod* 2006;**76**(5):773–8.

35. Kau CH, Bozic M, English J, Lee R, Bussa H, Ellis RK. Cone-beam computed tomography of the maxillofacial region–an update. *Int J Med Robot* 2009;**5**(4):366–80.

36. Aeria G, Claes P, Vandermeulen D, Clement JG. Targeting specific facial variation for different identification tasks. *Forensic Sci Int* 2010;**201**(1–3):118–24.

37. Delac K, Grgic M, Grgic S. Statistics in face recognition: analyzing probability distributions of PCA, ICA and LDA performance results, Zagreb, Croatia, 15–17 September*Proceedings of the 4th international symposium on image and signal processing and analysis, ISPA 2005*. p. 289–94.

38. Chalás I, Urbanová P, Kotulanová Z, Jandová M, Králík M, Kozlíková B, Sochor J. *Forensic 3D facial identification software (FIDENTIS)*, 2014.

39. Du S, Zheng N, Xiong L, Ying S, Xue J. Scaling iterative closest point algorithm for registration of *m–D* point sets. *J Vis Commun Image Represent* 2010;**21**(5–6):442–52.

40. Du S, Zheng N, Ying S, Wei J. ICP with bounded scale for registration of *m*-D point sets. *Multimedia and Expo, 2007 IEEE International Conference on 2007*, p. 1291–94.

41. Kotulanová Z, Chalás I, Urbanová P. 3D virtual model database of human faces: applications in anthropology and forensic sciences. *Mikulov anthropology meeting*. Brno: Academy of Sciences of the Czech Republic, Institute of Archeology, and Masaryk University, Department of Anthropology; 2014, p. 177–80.

42. Farkas LG. *Anthropometric facial proportions in medicine*. Thomas Books; 1987.

43. Tzou CH, Artner NM, Pona I, Hold A, Placheta E, Kropatsch M, Frey M. Comparison of three-dimensional surface-imaging systems. *J Plast Reconstr Aesthet Surg* 2014;**67**(4):489–97.

44. Min R, Choi J, Medioni G, Dugelay J. Real-time 3D face identification from a depth camera. *Pattern recognition (ICPR), 2012 21st international conference on, Tsukuba 11–15 Nov*, p. 1739–42.

45. Tong J, Zhou J, Liu L, Pan Z, Yan H. Scanning 3D full human bodies using kinects. *IEEE Trans Visualization Comp Graphics* 2012;**18**(4):643–50.

46. Chiller-Glaus SD, Schwaninger A, Hofer F. Identity verification from photographs in travel documents: role of display duration and orientation of performance. *Proceedings of the human factors and ergonomics society 51st annual meeting*, p. 1327–30.

47. Ellis HD, Shepherd JW, Davies GM. Identification of familiar and unfamiliar faces from internal and external features: some implications for theories of face recognition. *Perception* 1979;**8**(4):431–9.

48. Martire KA, Kemp RI, Sayle M, Newell BR. On the interpretation of likelihood ratios in forensic science evidence: presentation formats and the weak evidence effect. *Forensic Sci Int* 2014;**2**:24061–8.

49. Jandová M, Kotulanová Z, Urbanová P. Databáze tro-jrozměrných modelů obličeje dětí a její využití v ortodoncii. *Ortodoncie* 2015;**24**(1):14–21.

50. Gordon GG, Hall P. Face recognition based on depth maps and surface curvature. In: *Geometric methods in computer vision Proceedings of the SPIE*, vol. 1570, 1991. p. 234–247.

51. Urbanová P, Chalás I. Performance of matching algorithms in non-standard expression-variant faces. *Proceedings of the american academy of forensic sciences 68th annual scientific meeting, Las Vegas, February 22–27*, p. 445.

52. Urbanová P, Chalás I, Kozlikova B. How accurate is 3D facial morphology for personal identification? *Proceedings of the american academy of forensic sciences 67th annual scientific meeting, Orlando, February 16–21*, pp. 484.

53. Yoshino M, Matsuda H, Kubota S, Imaizumi K, Miyasaka S. Computer-assisted facial image identification system. *Forensic Sci Commun* 2001;**3**(1):1–4.

54. Sala R, Cantatore A, Cigada A, Zappa E. Personal identification and minimum requirements on image metrological features. *IMEKO, IEEE, SICE. 2nd international symposium on measurement, analysis and modeling of human functions. June 14–16, 2004, Genova, Italy*.

55. Whitmarsh T, Veltkamp RC, Spagnuolo M, Marini S, Ter Haar F. Landmark detection on 3D face scans by facial model registration. *1st international symposium on shapes and semantics*, p. 71–5.

56. Liang S, Wu J, Weinberg SM, Shapiro LG. Improved detection of landmarks on 3D human face data. *Engineering in medicine and biology society (EMBC), 2013, 35th annual international conference of the IEEE*, 6482-85.

57. Galvánek M, Furmanová K, Chalás I, Sochor J. Automated facial landmark detection, comparison and visualization. *Proceedings of the 31st spring conference on computer graphics*. Bratislava, Slovakia: Comenius University; 2015, p. 21–8.

58. Yoshino M, Taniguchi M, Imaizumi K, Miyasaka S, Tanijiri T, Yano H, David C, Thomas L, Clement JG. A new retrieval system for a database of 3D facial images. *Forensic Sci Int* 2005;**148**(2–3):113–20.