Available online at www.sciencedirect.com

SCIENCE DIRECT®

Forensic Science International 159S (2006) S147–S158

Forensic Science International

www.elsevier.com/locate/forsciint

Craniofacial reconstruction using a combined statistical model of face shape and soft tissue depths: Methodology and validation

Peter Claes [a], Dirk Vandermeulen [a,*], Sven De Greef [b], Guy Willems [b], Paul Suetens [a]

[a] Katholieke Universiteit Leuven, Faculties of Engineering and Medicine, Medical Image Computing, ESAT & Radiology, Herestraat 49, B-3000 Leuven, Belgium

[b] Katholieke Universiteit Leuven, Faculty of Medicine, School of Dentistry, Oral Pathology and Maxillo-Facial Surgery, Department of Forensic Odontology, Kapucijnenvoer 7, B-3000 Leuven, Belgium

Available online 15 March 2006

Abstract

Forensic facial reconstruction aims at estimating the facial outlook associated with an unidentified skull specimen. Estimation is generally based on tabulated average values of soft tissue thicknesses measured at a sparse set of landmarks on the skull. Traditional 'plastic' methods apply modeling clay or plasticine on a cast of the skull, approximating the estimated tissue depths at the landmarks and interpolating in between. Current computerized techniques mimic this landmark interpolation procedure using a single static facial surface template. However, the resulting reconstruction is biased by the specific choice of the template and no face-specific regularization is used during the interpolation process. We reduce the template bias by using a flexible statistical model of a dense set of facial surface points, combined with an associated sparse set of skull-based landmarks. This statistical model is constructed from a facial database of ($N = 118$) individuals and limits the reconstructions to statistically plausible outlooks. The actual reconstruction is obtained by fitting the skull-based landmarks of the template model to the corresponding landmarks indicated on a digital copy of the skull to be reconstructed. The fitting process changes the face-specific statistical model parameters in a regularized way and interpolates the remaining landmark fit error using a minimal bending thin-plate spline (TPS)-based deformation. Furthermore, estimated properties of the skull specimen (BMI, age and gender, e.g.) can be incorporated as conditions on the reconstruction by removing property-related shape variation from the statistical model description before the fitting process. The proposed statistical method is validated, both in terms of accuracy and identification success rate, based on leave-one-out cross-validation tests applied on the facial database. Accuracy results are obtained by statistically analyzing the local 3D facial surface differences of the reconstructions and their corresponding ground truth. Identification success rate is obtained by comparing, based on correlation, Euclidean distance matrix (EDM) signatures of the reconstructed and the original 3D facial surfaces in the database. A subjective identification success rate is quantified based on face-pool tests. Finally a qualitative comparison is made between facial reconstructions of a real-case skull, based on two typical static face models and our statistical model, showing the shortcomings of current face models and the improved performance of the statistical model.

Keywords: Forensic facial reconstruction; Computer-aided; Statistical modeling and reconstruction; Validation

1. Introduction

When confronted with a corpse that is unrecognizable due to its state of decomposition, skeletonisation, mutilation or incineration, and if no other identification evidence is available, craniofacial reconstruction can be considered. The goal of craniofacial reconstruction is to recreate an estimate of the face of an individual at the time of death. Hopefully, this will trigger recognition by relatives such that further identification evidence can be gathered on a restricted list of candidates. Although craniofacial reconstruction is a valuable tool in the initiation of the process of identification, positive identification has to be obtained eventually by classic techniques such as radiographic and dental comparisons or DNA-analysis.

Several 3D manual methods for facial reconstruction have been developed and are currently used in practice. These reconstructions consist of physically modeling a face on a skull replica (the target skull) with clay or plasticine. However, manual reconstruction methods require a lot of anatomical and artistic modeling expertise and are as a result highly subjective.

* Corresponding author at: Katholieke Universiteit Leuven, Faculty of Engineering and Medicine, Medical Image Computing, ESAT & Radiology, Herestraat 49, B-3000 Leuven, Belgium. Tel.: +32 16 349022; fax: +32 16 349001.
E-mail address: dirk.vandermeulen@esatkuleuven.be (D. Vandermeulen).

0379-0738/$ – see front matter © 2006 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.forsciint.2006.02.035

Furthermore, these reconstructions take a lot of time (several days), and, hence, are often limited to a single reconstruction.

Computer-based methods, on the other hand, are consistent (given the same input data and modeling assumptions, the same output results) and objective (knowing all the modeling assumptions). Moreover, since these methods can be executed in a short time, multiple reconstructions from the same skull using different modeling assumptions (older, thicker. . .) can be obtained.

Current computerized reconstruction techniques are limited, though, in the model used for reconstructing the complete facial outlook. First, either a generic face template or a specific best look-alike template, based on skull similarities or estimated subject properties (BMI, gender and age, e.g.), is chosen. Subsequently, the skin surface associated to the target skull is estimated by warping the model template, based on a generic, ''smooth'' deformation that maps landmark points on the model skull onto the corresponding landmark points on the target skull. Multiple reconstructions based on different values of BMI, age and gender are obtained by choosing a different initial face template. Two major shortcomings are apparent using such a single template model in combination with a face-unspecific deformation. First, the reconstruction can be incorrectly biased by the choice of the template. Indeed, when using a single subject-specific best look-alike template, based on whatever similarity properties, unwanted facial features of the template remain visible in the final reconstruction. Using a generic face template, on the other hand, results in too smooth and unspecific a reconstruction. In [1,2] this problem is solved by deforming multiple face templates towards the given skull and combining the results. This, however, results in a longer computation time required to generate such a reconstruction. Secondly, the generic deformations applied to the templates are not face-specific, but only ''smooth'' in a mathematical sense. No problem arises when the differences between the model and target skull-based landmarks are small. However, if these differences are relatively large, the required deformation will be more pronounced, resulting in a possibly unrealistic, caricature-like or implausible facial reconstruction.

Current computer-based facial reconstruction methods differ mainly by the selection of landmark points or skull features used to deform the model towards a given target skull. Some techniques [3–8] fit a facial template to the endpoints of a set of virtual dowels positioned on a 3D digitized model of the target skull. The dowel lengths represent averages of ancestry-, gender- and age-matched tissue depths at a limited number of predefined cephalometric landmarks. There is no direct correlation, however, between the reported tissue depths and the associated skin surface shape of an individual. Other computer-based techniques deform a 3D reference skull to a target skull based on crest lines (lines of maximal local curvature) [9], anatomical landmarks [10], feature points [11] or distance map representations [1]. The calculated skull deformation is then spatially extrapolated and applied to the skin surface associated to the reference skull. A reference skull is selected based on similarity in ancestry, gender and age. Reference skulls and corresponding facial surfaces are obtained using CT scanning, which limits the selection of the reference database to patient data because of the involved irradiation dose. Furthermore, CT images are acquired of subjects in a horizontal, supine position. As a result, due to gravitational forces, facial shapes extracted from CT images will differ from the typical facial shape as viewed in standard upright position.

In order to eliminate the template-related bias and to minimize the unrealistic character of the reconstructions caused by large model deformations, we propose a new flexible facial model for craniofacial reconstruction, modeling the combined population-dependent variation and correlation of skin surface shape and tissue depth (represented by skull-landmarks), calculated from a facial database. This requires three subtasks: (1) acquisition of 3D skin surfaces and tissue-depths in an upright position, measured over a sufficiently large and diverse population which are subsequently stored in a database together with the subject's age, body mass index (BMI), gender and ancestry; (2) statistical modeling of the population-dependent variation and correlation of skin surface shape and skull-landmarks; (3) fitting of this statistical model to the individual craniofacial skeleton.

2. Materials and methods

2.1. Database acquisition

A 3D facial entry (sample) in the database consists of a 3D skin surface representation coupled with soft tissue depths measured at 52 (10 midline plus 2 times 21 bilateral) manually indicated anatomical landmarks on the skin surface plus an indication of the nose tip. These landmarks are defined in and are measured according to [12]. The skin surface shape of an individual is measured with a mobile 3D photographic device (ShapeCam, Eyetronics (www.eyetronics.com)). Before capturing the 3D images, the 52 landmarks are marked on the face with a blue eyeliner pencil by a forensic odontologist. The 3D coordinates of these blue points are extracted from the 3D images by simple image processing and the nose tip is indicated manually. The soft tissue depths at the 52 cephalometric landmarks are measured with a compact and lightweight mobile digital ultrasound ''A-mode'' scanner (Epoch 4B with a 10 MHz 0.6 mm∅ transducer, Panametrics Inc., Waltham, USA). Further details and validation of the soft tissue depth measuring system can be found in [12].

The database used in this paper consists of 118 facial entries and associated tissue depths. All individuals are Caucasoid and subdivided in 48 female and 70 male subjects. The age distribution ranges between 18 and 55 years with an average of 25. The BMI distribution has an average of 22 kg/m^2 and minimum and maximum of 18 and 38, respectively. An example facial entry of the database is depicted in Fig. 1.

2.2. A statistical face and soft-tissue depth model

We construct a combined statistical facial surface and soft tissue depth tissue model from the acquired database. This model consists of a vector of 3D coordinates of a set of densely

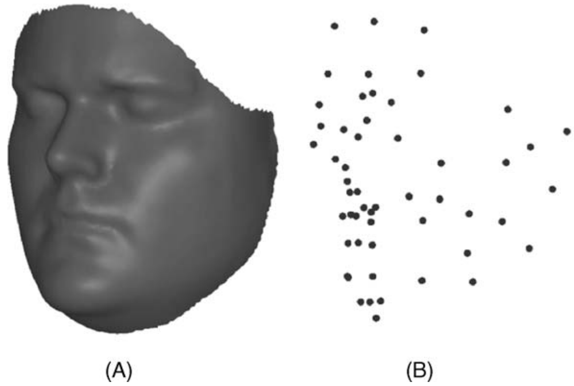

Fig. 1. An example facial entry from the database. (a) Shaded rendering of a dense point representation of the facial surface; (b) 53 skull-based landmarks; (c) combined facial surface and skull-based landmark view.

sampled points on the 3D skin surfaces and the 52 skull-landmarks with the nose tip as a final additional landmark. Skull-landmark points are positioned at a distance equal to the measured soft tissue depth along the inward pointing normal to the skin surface at the blue skin-landmark points. Inter-subject correspondences between skull-landmarks are automatically determined using a non-linear robust point matching procedure [13,14], while inter-subject correspondences between the dense set of points on the skin surfaces are obtained by geodesic surface matching [15].

The database is statistically modelled using a principal component analysis (PCA) of the thus obtained inter-subject point correspondences. The result of this PCA is a geometrically averaged facial template, which is calculated together with a correlation-ranked set of statistically independent modes of principal variations. These principal modes or components are in turn vectors of 3D coordinates defined as linear combinations of point position changes and capture the likely variation of these points from the average as observed over all subjects in the database. Every face in the database can now be represented as a weighted linear combination of these principal components and the average face shape. Furthermore, since the principal components are mutually uncorrelated and ranked according to the amount of variation in the data they explain, the weight associated to each principal component represents a statistical likelihood of occurrence.

In Fig. 2 the average face and the effects of changing the average face by changing the weights associated to the first two modes of variation are shown. The first mode models the facial differences between compressed and elongated faces along an anterior–posterior axis. The second mode of variation seems to be characterized by the same variation as the first mode but

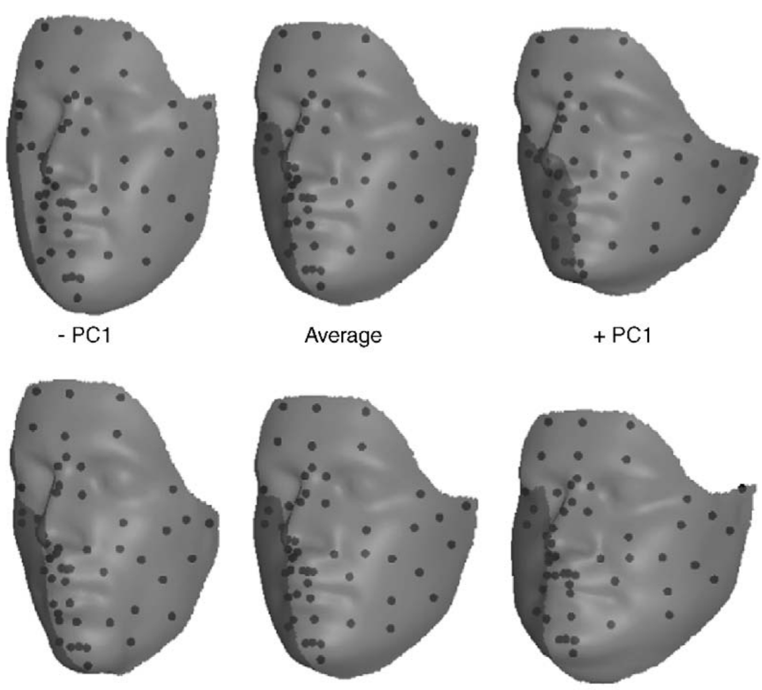

Fig. 2. Visualization of the first two modes of statistical covariation present in the current database. Top row: variation along the first principal component in negative (left) and positive (right) direction. Bottom row: variation along the second principal component in negative (left) and positive (right) direction.

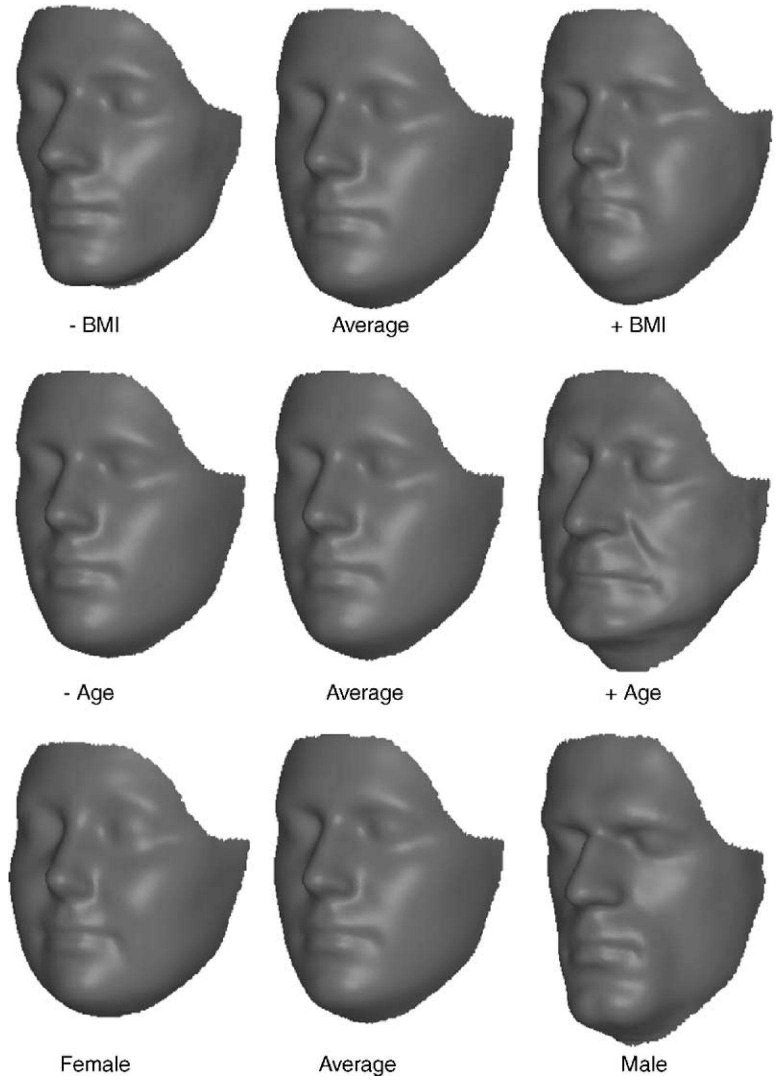

Fig. 3. Illustration of property changes in BMI (top row), age (middle row) and gender (bottom row), applied to the average face.

along the superior–inferior axis. A total of 117 modes of variation can be calculated from a database consisting of 118 entries.

Every facial entry in the database can now be represented using the parametric representation of the statistical model. Indeed, instead of using a vector description of the spatial coordinates of the skin surface points and the 53 skull-based landmarks, the facial entry can now be modelled as the sum of the averaged face and a weighted linear combination of the principal modes of variation. The parameters describing the facial entry are the weights in this linear combination. A probability value can be associated to each set of parameters, with maximal probability for the average face. By altering the parameters, in between statistically determined boundaries as learned from the database, faces spanning a face-space similar to the database can be generated. This is the key point of using a statistical model for craniofacial reconstructions.

Besides modelling pure inter-point correlations, property-related changes, such as changes in facial shape and soft tissue depth as a function of age, BMI, gender and ancestry can be modeled as well using the same PCA machinery [16]. Fig. 3 shows these attribute related deformations applied to the average face.

2.3. Statistical model fitting procedure

A schematic overview of the fitting procedure is given in Fig. 4. The first step in the facial reconstruction process is digitization of the skull, which can be done by CT-scanning of the skull and extracting the skull surface or by direct reconstruction of the skull surface using a 3D scanning device. A forensic anthropologist estimates the gender, age, BMI and ancestry of the skull as well as possible and indicates the 52 anatomical landmarks on the skull. An estimate of the nose tip is then made according to [17].

In a second step, the database is normalized according to the given skull properties by applying the property-specific deformations to every face in the database such that the properties of every face are now equal to the given properties. Based on this property normalized database, a new statistical model is determined which, as a result, incorporates the

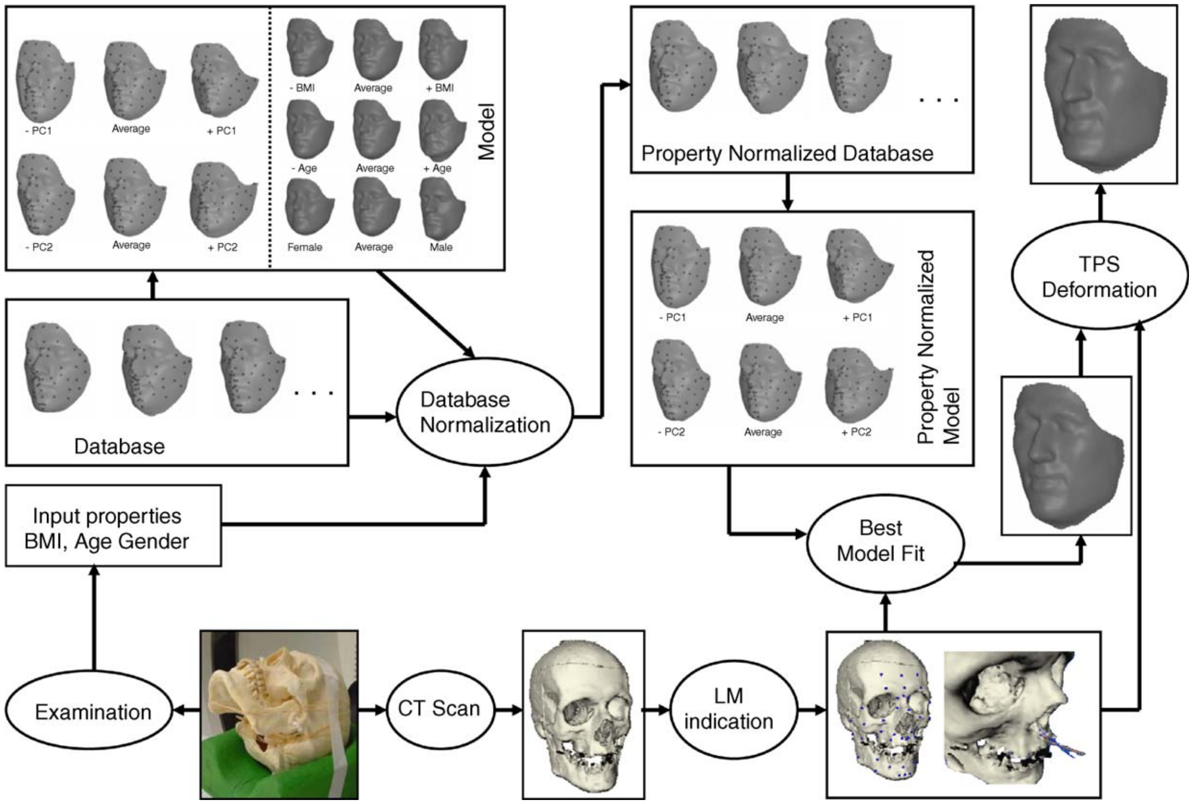

Fig. 4. A schematic overview of the statistical model fitting procedure for craniofacial reconstruction.

remaining inter-subject variations after removal of property-related variations.

In a third step, this model, which can be considered as an elastic mask with elastic dowels on the inside of the mask at the landmark locations, is fitted onto the skull. The modes of possible elastic deformation are the statistical modes of facial variation, with the elasticity or the flexibility of each mode proportional to its statistical likelihood. Fitting of this mask to the external surface of the individual craniofacial skeleton is obtained by finding a set of model parameters (the weights of the principal modes as well as the pose (translation, rotation, scale) parameters of the average model to fit the target skull), such that all the skull-based landmarks of the mask fit the corresponding indicated target skull-based landmarks as well as possible. However, since our database consists of only 118 facial samples, it currently has a limited ''elasticity'' (an insufficient number of degrees of freedom ($N = 117$)) to exactly fit the 53 skull-based landmarks (which requires at least 3×53 degrees of freedom, one for each point coordinate component).

In a final step, this lack of elasticity of the mask is reduced by combining the result of the previous step with a more generic physics-based, minimal bending energy, deformation, as follows. After the statistical (PCA-based) model fit, the remaining fit error at the skull-based landmarks is annihilated by calculating a thin-plate spline (TPS) warping [18] that maps the skull-based landmarks of the model onto the skull-based landmarks of the target. The same spatial warping is then applied to the dense point set of the facial model surface already aligned by the previous statistical model fitting step. Since the skull-based landmarks of the statistical model are already near to the target skull-based landmarks, the amount of generic TPS or non-face-specific deformation required is limited so that more realistic reconstructions are still guaranteed.

2.4. Validation

Validation of the proposed statistical method for craniofacial reconstruction is obtained by a leave-one-out cross-validation procedure. Each facial sample is removed, in turn, from the database and used as a test case. The remaining facial entries are used to create the statistical model. The 53 skull-based landmarks and property values (age, BMI and gender) of the test case are used as input for the statistical model fitting procedure. The resulting reconstructed facial skin surface can then be compared with the skin surface of the test case representing the ground truth.

A quantitative reconstruction error evaluation is performed by calculating the distances between every point on the reconstructed skin surface and its closest point on the measured skin surface of the test case. This is repeated for every facial entry in the database, such that we obtain an average absolute reconstruction error and standard deviation, which can be visualized for every point on the reconstructed skin surface. Comparing the two surfaces this way is interesting for evaluating the reconstruction performance in terms of accuracy

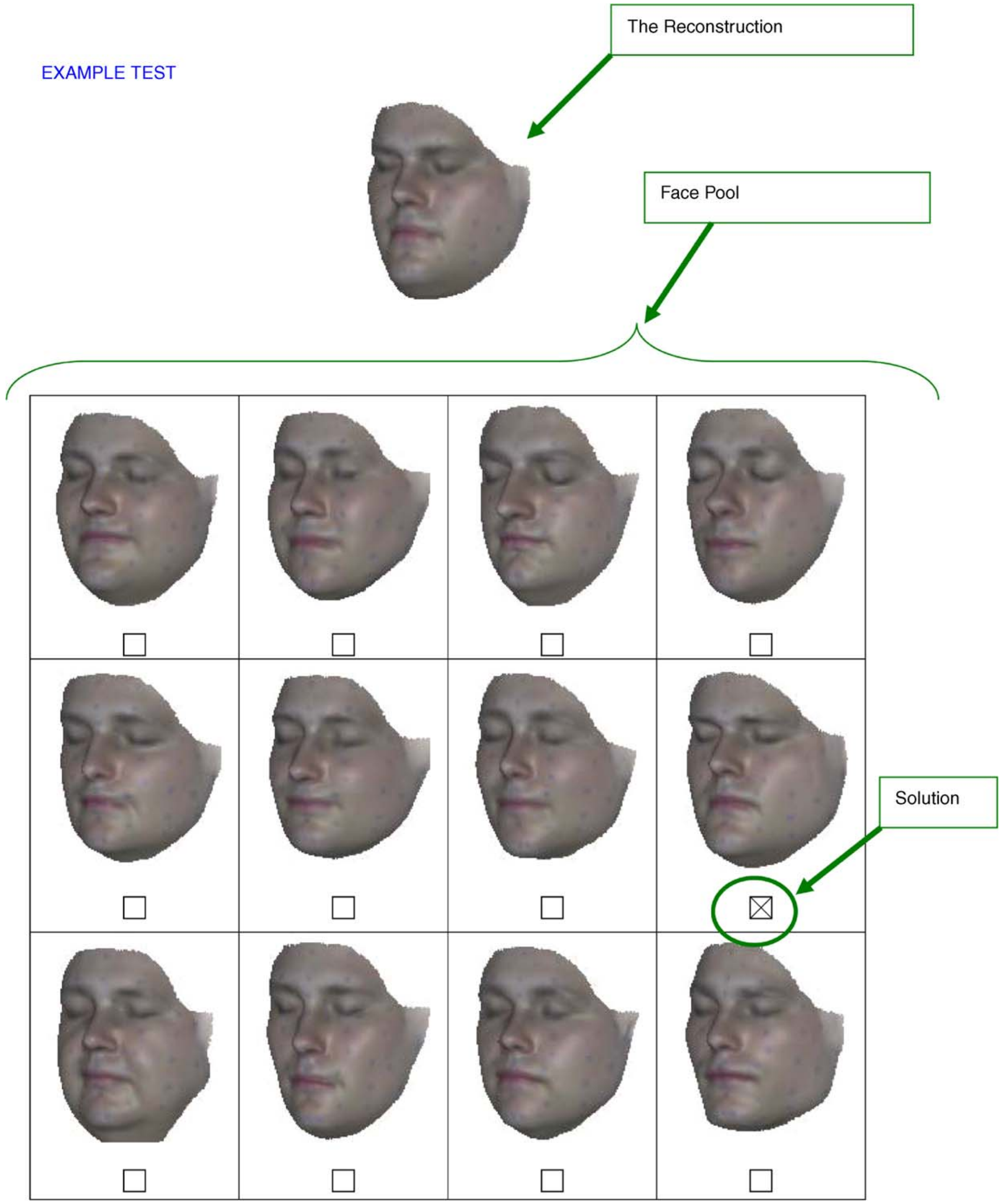

Fig. 5. Example face-pool test. Top image is the image of the reconstruction to be identified with examples in the face pool.

and provides a spatial map of the difficulty of each facial region to be reconstructed.

However, the final goal of craniofacial reconstruction is not reconstruction accuracy, but rather recognition or identification success. In order to examine the performance of the proposed statistical reconstruction method for 3D face identification, we make use of Euclidean distance matrix (EDM) descriptors for 3D surfaces [19]. The EDM of a 3D surface that is represented as a set of landmark point coordinates, is a matrix consisting of all pair-wise distances between the landmarks. An EDM of a 3D surface is a kind of signature for that surface and can be compared with signatures of other surfaces by means of correlation: the higher the correlation between the signatures, the more alike the surfaces are. We calculate the EDM for every reconstructed 3D facial surface and for every 3D facial surface in the database. Then, for every reconstruction we compare its EDM to the EDM of every face in the database. The face corresponding to the identification is defined as the one with the most similar EDM. The performance of the identification process based on the reconstructions obtained from the leave-one-out tests is an indicator for the quality of these reconstructions in the context of 3D face recognition and identification.

A more realistic, human subjective, identification process is simulated by generating face-pool tests. An example test is shown in Fig. 5. Given an image of a reconstruction and a set of possible candidate images extracted from the database, a human observer is asked to indicate the face from the face-pool most similar to the given reconstruction. The images of the faces in these tests are normalized according to pose and colour (the same texture is projected on every facial surface). The pose is defined as a ‘‘three-quarter’’ view (a 45° rotation around a vertical axis of the head) which is an advantageous pose in terms of face recognition [20]. A total of 18 different tests are created with different difficulty, based on the EDM signatures. A first test difficulty criterion is the quality of the reconstruction to be identified. Based on the correlation values between the EDM of the reconstructions and the EDM of their corresponding ground truth facial surfaces, we rank the quality of the reconstructions into superior, average and inferior. A second test difficulty criterion is based on similarity of the face-pool images to the face reconstruction to be identified. This similarity is based on comparing the EDM signatures of face-pool and reconstructed faces as well: tests with a face-pool set of similar faces are much harder to correctly identify than with dissimilar, say randomly selected, face-pool sets. In this way, face-pools can also be divided into three groups: good look-alikes, average look-alikes and bad look-alikes. A final test difficulty criterion is the closed versus open face-pool comparison: in a closed face-pool comparison test, the face to be identified with is an element of the face-pool set, while in the open face-pool comparison test, the option that the face to be identified is in the face-pool is left open. An overview of the individual test difficulties is given in Table 1.

Finally, to show the differences between our proposed statistical model and the more traditional static models in combination with a generic interpolated deformation, we compare the different reconstruction results on a real-case skull found in Belgium. In order to make the reconstructions comparable, all models used have 52 skull-landmarks and an estimate of the nose tip as additional skin landmark to determine the model deformation. A 3D digitized model of the skull, acquired with CT scanning, and the landmarks are depicted in the schematic overview (Fig. 4). The statistical model is compared with a generic (the average face from our database) and a specific face model (a particular face from our database), which are deformed towards the skull by solely making use of a TPS-based smooth deformation.

Table 1
Individual face-pool test difficulties

TEST	Reconstruction	Face pool	Open/closed	Difficulty (1–10)
1	S	BL	Closed	1
2	A	BL	Closed	2
3	I	BL	Closed	3
4	S	AL	Closed	4
5	A	AL	Closed	5
6	I	AL	Closed	6
7	S	GL	Closed	7
8	A	GL	Closed	8
9	I	GL	Closed	9
10	I	GL	Open	10
11	A	GL	Open	9
12	S	GL	Open	8
13	I	AL	Open	7
14	A	AL	Open	6
15	S	AL	Open	5
16	I	BL	Open	4
17	A	BL	Open	3
18	S	BL	Open	2

S, superior reconstruction; A, average reconstruction; I, inferior reconstruction; BL, bad look-alike face-pool; AL, average look-alike face-pool; GL, good look-alike face-pool.

3. Results and discussion

3.1. Cross-validation-based validations

The leave-one-out cross-validation procedure was performed on the currently available database resulting in 118 test cases. One of the test cases was the face in Fig. 1. The reconstruction made with the proposed statistical model, constructed from all the other faces in the database and based on the landmarks and attribute values (BMI 35, age 25 and male) of the face in Fig. 1, is shown in Fig. 6a. More quantitatively, the local absolute reconstruction errors in every

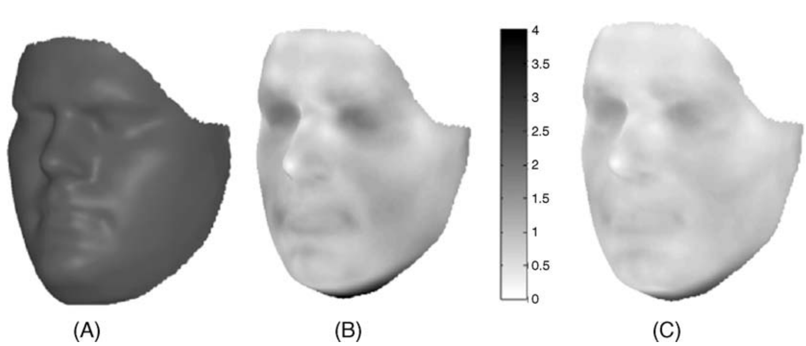

Fig. 6. Accuracy results from the cross-validation tests. (a) The reconstruction of the face in Fig. 1(a) based on the skull-landmarks in Fig. 1(b); (b) the average local absolute reconstruction error in every point of the skin-surface visualized on the average face; (c) the standard deviation of the error in every facial surface point.

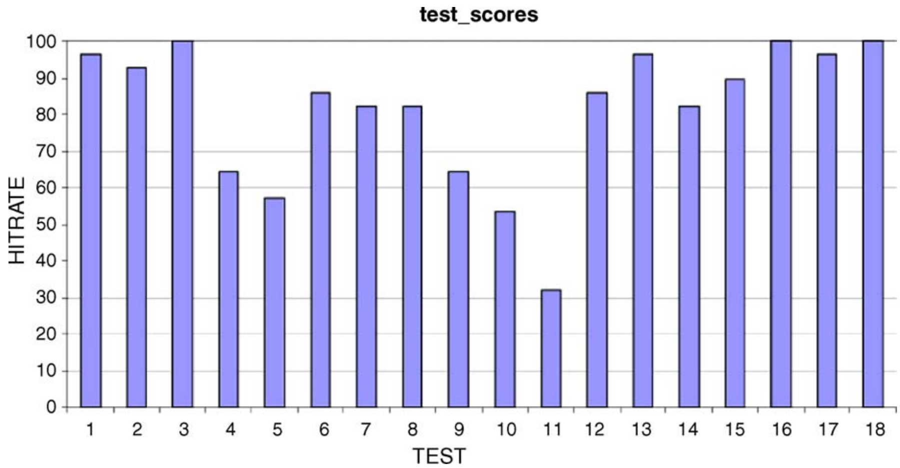

Fig. 7. The test scores (percentage of correct identifications) for the 18 face-pool tests.

point of the reconstructed skin surface (Fig. 6a) compared with the real test skin surface (Fig. 1a) were determined. This was repeated for all 118 faces in the database resulting in an average (Fig. 6b) and standard deviation (Fig. 6c) of the absolute local reconstruction error in every point of the skin surfaces. We used the average face surface in order to visualize the average errors in every point with a grey colour code. The range of the colour code is between 0 (white) and 4 mm (black). The overall average absolute reconstruction error and standard deviation over all the points and test cases is 1.14 mm and 1.04 mm, respectively. The facial areas with the highest reconstruction error are the eyes and the region below the chin. This is mostly due to the fact that no skull-based landmarks are present in these areas.

In order to analyze the identification success rate, we determined the EDM signatures for the 118 reconstructions and the 118 original faces from the database and compared the signature of every reconstruction with the signature of every original face by means of correlation. The corresponding or identified face from the database belonging to a reconstruction was defined as the one with the highest correlation between their signatures. By doing so we reached a 100% correct identification rate: every reconstruction was linked to the correct face in the database! This indicates that the quality of the reconstructions is good enough for identification based on EDM signatures of 3D facial surfaces. The statistical power of this test, however, is also limited by the relatively small size of the current database. A larger database has a higher probability of containing some faces that are even more similar based on their EDM signatures, which could result in false identifications. Furthermore, a 3D database of missing persons, to compare a craniofacial reconstruction with, does not exist, so in practice this 3D identification method is not applicable.

More realistic identification success rates were obtained with the face-pool tests. Indeed these tests more or less simulate the real situation by generating 2D images from the 3D models and showing them to human observers. Twenty eight human observers with no prior training in this identification experience, solved the 18 tests with an overall average identification rate of 81.15%. The separate test-scores (percentage of correct identification) are shown in Fig. 7. As can be seen the test-scores correlate well with the test

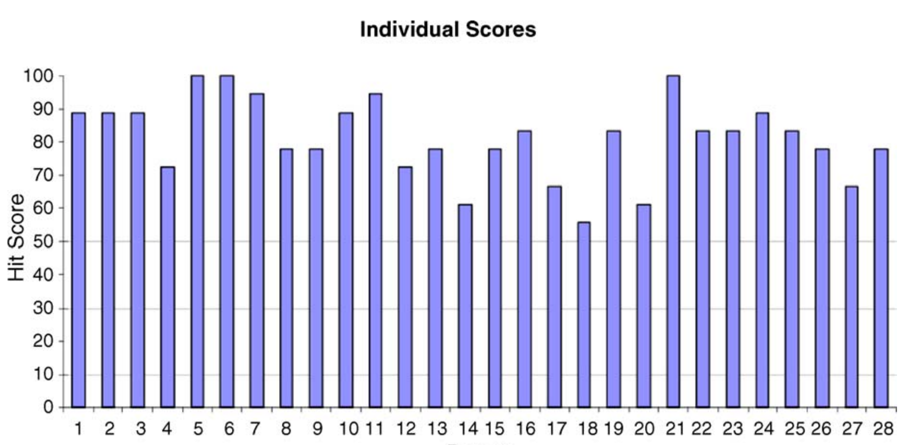

Fig. 8. The individual scores (percentage of correct identifications) for the 28 human observers.

difficulties tabulated in Table 1. The lowest test score is 32.14% for test number 11, which is indeed difficult according to the criteria we defined before. Moreover, the faces in the face pool shown for this case did not contain the face of the corresponding individual. The observer errors for this case were all false candidate (false positive) indications. Surprisingly, the test scores of tests 4 and 5 were rather low, indicating that the predicted difficulties of tests 4 and 5 were too low. The separate individual human scores are depicted in Fig. 8, where the lowest identification score is 55%. Three out of the 28 observers had a perfect identification score of 100%. Upon analyzing these results, it should be noted that the face-pool tests are still a simplified version of reality. In practice, images of missing persons to compare with are not all of the same quality and the poses of the faces are not as well normalized either. Furthermore, no hair or background was shown in the face-pool tests. Including this will probably lower the identification rate to a more realistic identification rate. However, it is very difficult to generate such realistic face-pool tests. The simplified face-pool test version used in this experiment is interesting in se to illustrate the performance of a certain craniofacial reconstruction method.

The accuracy and identification success rates are also overly optimistic still, since the input data (the skull-based landmarks (52 on the skull as well as the nose tip determined from the skull) of the test faces) for the reconstruction method were simulated to be defined in an identical way as during the model building. Further tests are to be performed that include some amount of error on the skull-landmark indications in order to analyze their effect on the reconstruction accuracy and identification success. We also want to notice that the proposed statistical model for craniofacial reconstruction can be applied with a different choice of skull-landmarks than the ones used by De Greef et al. [12]. In fact, one could use a dense set of skull-points (acquired with CT scanning, e.g.) in combination with a dense set of face-points. However, the question then arises whether a dense set of skull-points is really necessary in terms of improving the identification success. The advantage of using our limited set of landmarks is the fact that they were measured in a non-invasive way (with ultrasound) and in an upright position. Also, fewer landmarks reduce both computational and algorithmic complexity.

Based on the leave-one-out cross-validations, we can conclude that the proposed craniofacial reconstruction method performs well on these simplified, but controlled, test data. Note that, until now, to the best of our knowledge, no validation framework for craniofacial reconstruction methods has been defined or used in literature. The proposed validation procedure is the first attempt to set-up a proper validation which will hopefully substantiate on a scientific basis the added value of craniofacial reconstruction methods during crime-scene investigations.

3.2. Comparative study

Two different static models were used to make a facial reconstruction of the skull in Fig. 4 based on a smooth generic TPS deformation. The first is the average face of our database, being a generic facial model. The second is a subject-specific face from the database. The models and reconstruction results are shown in Fig. 9. The model bias in the reconstructions is clearly visible. The result from the average face is very smooth and contains no specific facial characteristics, similar to the average face. When using the subject-specific facial template, facial characteristics of the template remain in the final reconstruction (the eyebrows and the chin region, e.g.). Furthermore, large differences between several skull-based landmarks on the model and the target skull (the nose tip e.g.) generate caricature-like final outlooks of the reconstructions due to the large amount of non-affine TPS deformation that has to be applied. Especially, when using the subject-specific face, the nose of the reconstruction looks unrealistic.

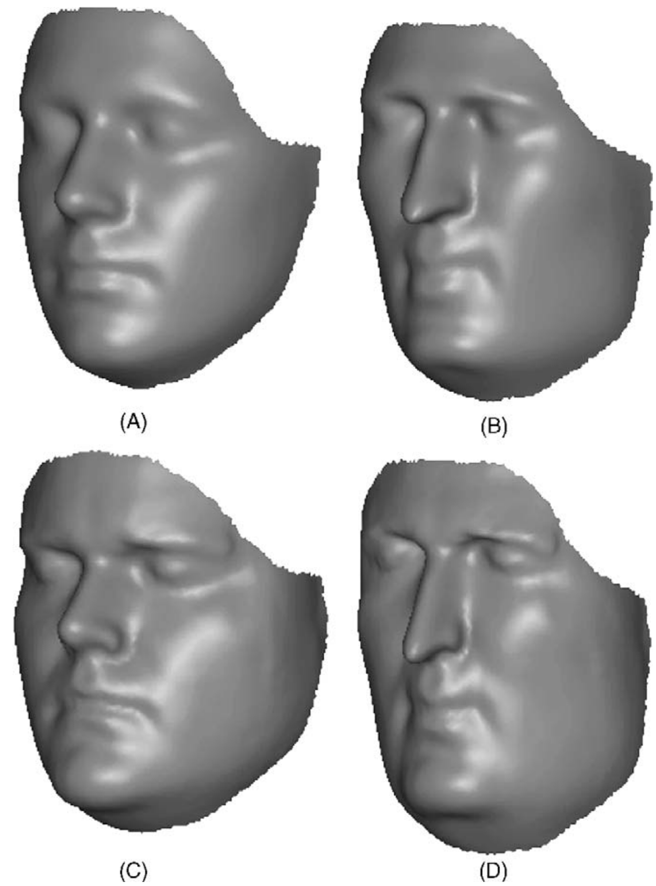

Fig. 9. Craniofacial reconstructions of the skull used in Fig. 4 made with two different static models in combination with a generic TPS-based deformation. Top row: a generic facial template (a) being the average face of our database and the corresponding reconstruction result (b). Bottom row: a specific facial template (c) identical to the face shown in Fig. 1 and the corresponding reconstruction result (d).

A facial reconstruction of the same skull was also made with our proposed flexible statistical model. In a first test, the property values were set to BMI = 20 (thin person), age = 25 and gender = male. The model was initialized with the average face of the property normalized database. The results after model fitting and TPS deformation are shown in Fig. 10(a) and

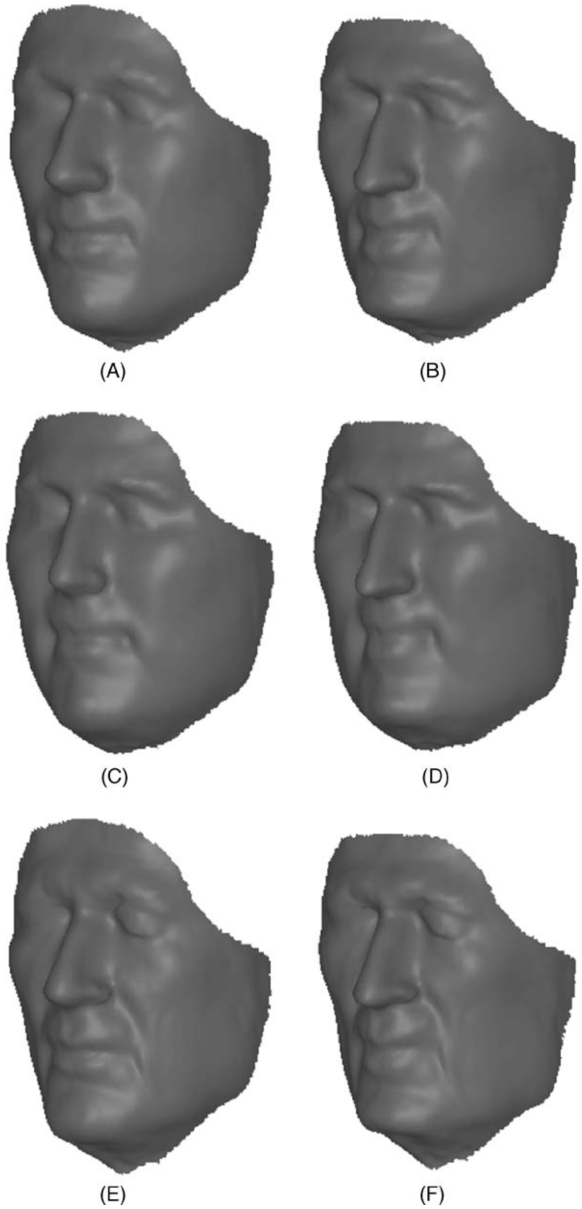

Fig. 10. Three different reconstructions of the skull used in Fig. 4 based on the proposed statistical model and fitting procedure. Top row: Statistical model fit (a) and final TPS-based reconstruction (b) for BMI = 20, age = 25 and gender = male. Middle row: statistical model fit (c) and final TPS-based reconstruction (d) for BMI = 30, age = 25 and gender = male. Bottom row: statistical model fit (e) and final TPS-based reconstruction (f) for BMI = 20, age = 70 and gender = male.

(b), respectively. In a second test, we used a subject-specific face of the normalized database as initialization for the model fitting and we observed that the final model fit and final reconstruction were exactly the same as in Fig. 10(a) and (b). In fact, any face from the property normalized database could have been used to initialize the model leading to the same results. When observing the reconstruction in Fig. 10(b), more characteristic facial features are apparent in the reconstruction than the one based on the average model in Fig. 9(b), but they cannot be attributed to a subject-specific face in the database, unlike the reconstruction based on the subject-specific face model in Fig. 9(d). In other words, no bias towards a specific face in the database is introduced into the reconstruction. Finally, the plausibility of the reconstruction being a real human face in Fig. 10(b) is much higher than the caricature ones in Fig. 9, especially when observing the nose. This can be explained by the fact that the differences between the model skull-based landmarks and the given skull-based landmarks are reduced during the model fitting by applying face-specific variations so that a smaller amount of final TPS deformation is needed eventually. Indeed, the difference between the final statistical model fit in Fig. 10(a) and the final reconstruction in Fig. 10(b) is much smaller than the differences between static models and reconstructions in Fig. 9(a, c) and (b, d), respectively.

Two more reconstructions were made based on different property values to illustrate the possibilities of the proposed reconstruction method. When a different set of property values is suggested, manual reconstructions by means of modeling plasticine need to be completely redone, while reconstruction methods based on a static model require a new facial template with the proper properties to start from. With our proposed method, no new template is required and a new facial reconstruction is made within a few seconds (on a Pentium4 2.2 GHz CPU). In Fig. 10 the reconstruction results are shown for an increase of BMI (30) (Fig. 10(c) and (d)) and an increase in age (70) (Fig. 10(e) and (f)). When observing the reconstruction results in Fig. 10, one can see that the major facial characteristics remain visible over the different reconstructions, while the property specific characteristics are clearly different.

4. Conclusion

We proposed a statistical model of combined soft tissue-depths and complete facial surfaces, which can be used for 3D computerized forensic facial reconstructions. The main difference with currently used facial models is the automatic adjustment or improvement of the model by making use of face-specific modes of variation, which in combination with a TPS-based final deformation results in unbiased and more realistic reconstructions.

The statistical face model was learned from a database of complete 3D facial surfaces with 52 skull-landmarks and the tip of the nose, by applying PCA analysis. The resulting model consists of a geometrically averaged facial template together with a correlation-ranked set of modes of principal variations or face-specific deformations that capture the major changes or differences between facial outlooks and their skull-based landmarks in the database. This (statistically) elastic mask is subsequently fitted to the external surface of the individual craniofacial skeleton such that all the 53 landmarks of the mask fit the corresponding target skull-landmarks and the estimate of the nose tip. The elasticity of the mask and the dowels is defined as the statistically allowed variation learned from the database. In order to overcome a lack of elasticity of the mask due to a limited database or limited variation in the database, we combine the face-specific deformations with a more generic

TPS-based deformation, resulting in a flexible but constrained deformation of the mask towards a given skull. The average face from the database is used as initialization of the mask and in combination with the face-specific constraints on the deformation, no severe bias is introduced into the final reconstruction. Before the fitting process, property values of a given skull specimen are imposed as hard constraints by removing the variability in the database and the statistical model due to differences in property values. As a result, multiple reconstructions of the same skull but with different property conditions can be made within a few seconds.

Validation of reconstructions made with the statistical face model was performed by making use of leave-one-out cross-validation tests. Both reconstruction accuracy and identification success were analyzed, showing an overall average accuracy of 1.14 mm and an identification rate of 100% based on EDM signatures of the facial surfaces in the current database of 118 individuals. Identification accuracy by 28 untrained human observers on a face-pool test of 18 randomly selected reconstructions, showed an identification success rate of 81.15%. Despite the fact that these tests were simplifications of the real reconstruction case, they are a first and necessary step in a controlled validation study for craniofacial reconstructions. This will, hopefully, improve a correct appreciation of the value of craniofacial reconstructions in forensic practice.

A comparison was made between a reconstruction with the proposed model and reconstructions with a generic and a subject-specific static facial model on a real-case skull. The result of this might be interpreted as an improvement in terms of facial plausibility of the reconstruction. Different initializations of the model were used, leading to the same template and reconstruction results, which indicate that no bias towards a specific face was introduced into the reconstruction. Finally, multiple reconstructions were made in order to illustrate the possibilities of the proposed statistical method for altering the skull property values.

Some extensions can be proposed to the reconstruction based on the combined statistical model in combination with a TPS deformation. First of all, having a larger database, with samples that better cover the population spectrum, increases the flexibility of the model. The more the model can be adjusted towards the skull in a face-specific way during model fitting, the smaller the amount of the remaining generic, face-unspecific deformation required. The current database of 118 faces is rather limited in this respect, and is currently being expanded. Secondly, the indication of the skull-landmarks is currently done manually and needs a certain level of expertise. Furthermore, it introduces additional errors in the fitting procedure that reduce the quality of the craniofacial reconstruction. It would be interesting to relax the model fitting procedure according to the measured error ranges on the landmark indications. Another solution could be the automatic extraction of the skull-landmarks from the skull, such that the indication would be less error prone and such that the complete reconstruction pipeline would be automatic.

Acknowledgements

This work is supported by the Flemish Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT, project IWT/GBOU/020195), by the K.U. Leuven (project/OF/GOA/3E020776 01.01.2004-31.12.2008) and by the Fund of Scientific Research—Flanders (FWO-Vlaanderen projects FWO/G.0258.02N and FWO/G.0566.06).

References

[1] D. Vandermeulen, P. Claes, P. Suetens, S. De Greef, G. Willems, Volumetric deformable face models for cranio-facial reconstruction, in: S. Lončarić, H. Babić, M. Bellanger (Eds.), Proceedings of the Fourth International Symposium on Image and Signal Processing and Analysis—ISPA2005, Zagreb, Croatia, 15–17, (2005), pp. 353–358.

[2] R. Tu, R. Hartley, W.E. Lorensen, A. Alyassin, R. Gupta, L. Heier, Face reconstruction using flesh deformation modes, in: J.G. Clement, M.K. Marks (Eds.), Computer Graphic Facial Reconstruction, Elsevier Academic Press, 2005, pp. 145–162.

[3] P. Vanezis, Application of 3-D computer graphics for facial reconstruction and comparison with sculpting techniques, Forensic Sci. Int. 42 (1989) 69–84.

[4] P. Vanezis, M. Vanezis, G. McCombe, T. Niblet, Facial reconstruction using 3-D computer graphics, Forensic Sci. Int. 108 (2000) 81–95.

[5] R. Evenhouse, M. Rasmussen, L. Sadler, Computer-aided forensic facial reconstruction, J. Biocommun. 19 (1992) 22–28.

[6] A.W. Shahrom, P. Vanezis, R.C. Chapman, A. Gonzales, C. Blenkinsop, M.L. Rossi, Techniques in facial identification: computer-aided facial reconstruction using laser scanner and video superimposition, Int. J. Legal Med. 108 (1996) 194–200.

[7] A.J. Tyrell, M.P. Evison, A.T. Chamberlain, M.A. Green, Forensic three-dimensional facial reconstruction: historical review and contemporary developments, J. Forensic Sci. 42 (1997) 653–661.

[8] J. Muller, A. Mang, T.M. Buzug, A template-deformation method for facial reproduction, in: S. Lončarić, H. Babić, M. Bellanger (Eds.), Proceedings of the Fourth International Symposium on Image and Signal Processing and Analysis—ISPA2005, Zagreb, Croatia, September 15–17, (2005), pp. 359–364.

[9] G. Quatrehomme, et al., A fully three-dimensional method for facial reconstruction based on deformable models, J. Forensic Sci. 42 (1997) 649–652.

[10] L.A. Nelson, S.D. Michael, The application of volume deformation to three-dimensional facial reconstruction: a comparison with previous techniques, Forensic Sci. Int. 94 (1998) 167–181.

[11] M.W. Jones, Facial reconstruction using volumetric data, in: T. Ertl, B. Girod, H. Niemann, H.-P. Seidel (Eds.), Proceedings of the Sixth International Vision Modelling and Visualisation Conference—VMV-01, Stuttgart, Germany, November 21–23, (2001), pp. 135–150.

[12] S. De Greef, P. Claes, W. Mollemans, M. Loubele, D. Vandermeulen, P. Suetens, G. Willems, Semi-automated ultrasound facial soft tissue depth registration: method and validation, J. Forensic Sci. 50 (2005) 1282–1288.

[13] H. Chui, A. Rangarajan, A new algorithm for non-rigid point matching, in: Proceedings Conference on Computer Vision and Pattern Recognition (CVPR 2000), IEEE Computer Society, Hilton Head, SC, USA, June 13–15, (2000), pp. 44–51.

[14] H. Chui, A. Rangarajan, A feature registration framework using mixture models, in: Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA'00), IEEE Computer Society, Hilton Head, SC, USA, June 13–15, (2000), pp. 190–197.

[15] E. Huot, I. Cohen, H. Yahia, I. Herlin, Matching structures by computing minimal paths on a manifold, J. Visual Commun. Image Representation 13 (2002) 302–312.

[16] J. Hug, C. Brechbühler, G. Székely, Tamed snake: a particle system for robust semi-automatic segmentation, in: MICCAI '99: Proceedings of the

Second International Conference on Medical Image Computing and Computer-Assisted Intervention, 1999, pp. 106–115.
[17] K.T. Taylor, Forensic Art and Illustration, CRC Press, New York, 2001.
[18] F. Bookstein, Principle warps: thin plate splines and the decomposition of deformations, IEEE Trans. Pattern Anal. Mach. Intell. 11 (1989) 567–585.
[19] S.R. Lele, J.T. Richtsmeier, An Invariant Approach to Statistical Analysis of Shapes, Chapmann & Hall/CRC, 2001.
[20] H. Hill, Using laser scans to study face perception, in: J.G. Clement, M.K. Marks (Eds.), Computer Graphic Facial Reconstruction, Elsevier Academic Press, 2005, pp. 285–303.