3D Face Reconstruction from 2D Pictures: First Results of a Web-Based Computer Aided System for Aesthetic Procedures

Thiago Oliveira-Santos (1), Christian Baumberger (1), Mihai Constantinescu (2), Radu Olariu (2), Lutz-Peter Nolte (1), Salman Alaraibi (1), Mauricio Reyes (1)

1 - Institute for Surgical Technology and Biomechanics, University of Bern, Switzerland

2 - Department of Plastic, Reconstructive, and Aesthetic Surgery, Inselspital, Switzerland

Running head: 3D Face Reconstruction from 2D Pictures

Contact: Thiago Oliveira-Santos, ISTB, University of Bern, Stauffacherstr. 78, 3014 Bern, Switzerland, phone: +41 31 631 5955, fax: +41 31 631 5960, thiago.oliveira@istb.unibe.ch

Abstract

The human face is a vital component of our identity and many people undergo medical aesthetics procedures in order to achieve an ideal or desired look. However, communication between physician and patient is fundamental to understand the patient's wishes and to achieve the desired results. To date, most plastic surgeons rely on either “free hand” 2D drawings on picture printouts or computerized picture morphing. Alternatively, hardware dependent solutions allow facial shapes to be created and planned in 3D, but they are usually expensive or complex to handle. To offer a simple and hardware independent solution, we propose a web-based application that uses 3 standard 2D pictures to create a 3D representation of the patient's face on which facial aesthetic procedures such as filling, skin clearing or rejuvenation, and rhinoplasty are planned in 3D. The proposed application couples a set of well-established methods together in a novel manner to optimize 3D reconstructions for clinical use. Face reconstructions performed with the application were evaluated by two plastic surgeons and also compared to ground truth data. Results showed the application can provide accurate 3D face representations to be used in clinics (within an average of 2 mm error) in less than 5 minutes.

Key Terms: 3D Shape Modeling from Images, Planning of Surgical Facial Interventions.

1. Introduction

Along with its functional aspects, the human face is a vital component of our identity and it provides us with many intricate and complex communication channels to society. Many people with desire for change, suffering from low self-esteem, or seeking for an ideal look according to society, undergo medical aesthetics procedures [6]. However, communication between physician and patient is fundamental in order to understand the very often subjective wishes of the patient.

Most plastic surgeons rely on either "free hand" two-dimensional (2D) drawings on picture printouts or computerized picture morphing [19,20] in order to establish the goals of facial aesthetic procedures, i.e. to discuss the feasible procedures according to the wishes of the patient. However, 2D visualization is limited to one point of view and therefore hinders the overview of the procedure outcome. Three-dimensional (3D) models associated with 3D planning tools can overcome such limitations, but the acquisition of these representations of the patient face is not trivial. Computed Tomography (CT)-scans allow creation of 3D facial shapes [14,15], but their radiation exposure and costs are not acceptable for aesthetic procedures. Other hardware dependent solutions such as laser or stereo-photogrammetric scanners [11,16,22] also offer the possibility of creating 3D facial shapes, but such devices are usually expensive or complex to handle. As a result, their use is limited to physicians having the necessary financial or technical resources. Alternatively, other hardware independent methods for creating 3D faces from pictures and video have been proposed including shape from shading [24], structure from motion [13,23], shape from silhouette [18,21], and statistical facial models [3]. While the former three are dependent on the condition of light available, continuous multiple frames acquisitions (e.g. video), or high number of frames respectively, the latter has showed very robust results on reconstructing 3D faces by morphing a statistical facial model to a subject specific face. The use of these methods has been mainly limited to a research environment

and therefore they are not optimized for clinics. Limitations on speed, accuracy and lack of planning capabilities hinder the direct use of these techniques as a physician-patient communication tool.

In order to overcome these drawbacks, we propose in this paper a web-based and hardware independent application which creates a 3D representation of the patient's face from 2D-digital pictures and enables planning of aesthetic procedures in 3D. The proposed application requires three standard 2D-pictures of the patient (one frontal and two profiles) and a few landmarks in each image as input. The web-based software computes a patient-specific virtual 3D face on which physicians can directly show the intended procedural changes to the patient with the 3D planning tools in different points of view. Clinical usability was the main focus of this application, therefore the methods for facial feature detection, 3D reconstruction and texture mapping were carefully chosen in the literature and optimized to enable their use within standard consultation time. Emphasis was given to the link between the different steps of the pipeline in order to allow for a time-effective application with sufficient accuracy for clinical use. Therefore, methods for facial feature contour detection based on training from synthetic data and the semi-supervised feature contour correction are proposed. To evaluate this application, a set of face reconstructions was performed and compared to different ground truth data. Variables include 3D reconstruction time, distance to the ground truth, as well as qualitative evaluation performed by two plastic surgeons (for reconstructions and planning tools).

2. Materials and Methods

2.1 Application

Similarly to a previously developed application for 3D visualization of breast augmentation procedures [12], this application is accessible entirely on the internet. Therefore, it requires no additional hardware apart from a standard digital camera. The application runs on a normal web browser, which allows physicians, dermatologists, and aesthetic professionals from all

over the world to plan aesthetic facial procedure in 3D following a simple workflow without dealing with technical challenges (see Figure 1). First, the user measures the approximate distance between the eyes of the patient with a normal ruler and takes three pictures (one frontal, one from left profile and one from the right profile). The measurement and images are uploaded to the application running on a standard web browser. Subsequently, the physician manually places a set of landmarks in the images (28 in total, see image and landmarks from Figure 1) and uploads them to the web server. These landmarks representing key facial features (pupils, mouth corner, etc) increase the robustness of the automatic detection of further points (around 120 per image) representing contours of facial features (eyes, mouth, face, etc) and necessary for the 3D reconstruction. With the detected contour points, the physician has the opportunity to check and correct them if necessary. Splines facilitate interaction with the contours (see feature contours correction in Figure 1). Once the correction is finished, the splines are converted back to facial contours points that are sent to the web server to reconstruct a 3D textured shape of the patient face based on a 3D statistical shape model. Finally, the 3D representation of the patient face is displayed on the web browser powered with Unity 3 (from Unity Technologies, San Francisco, United States), and the physician may discuss the intended aesthetical procedures with the patients using the included and available planning tools. The set of tools allow manipulation of the patient specific virtual face in 3D considering different aesthetic procedures such as rhinoplasty, skin fillers, and dermabrasion procedures. The following sections explains the different steps of the pipeline.

Input Data Acquisition

The instructions for acquiring the input data are simple and easy to be replicated in every clinic. Colored images should be taken in the portrait orientation with the face occupying most of the image (around 60%). Faces on the order of 500 pixels width have been found to be a good compromise between reconstruction and image quality. The patient should have a neutral expression, with opened eyes and without accessories (e.g. ear rings or glasses). In

case of long hair, a hair band can be used to avoid hair on the facial region. The frontal image should be acquired with the patient facing the camera and with the nose in the center of the image. The profile images should be acquired with the patient ear and shoulder facing the camera and with the cheek in the center of the image. The camera to patient distance should be approximately 2 m. Such distance was found to be a good compromise between distortions caused by perspective projection and optical zoom power of standard digital cameras. The eye distance can be measured by placing a ruler on the nose of the patient and checking the values laying on the center of the eye. This distance is later used to rescale the final reconstructed shape and perform virtual measurements. Additionally, the physician must place 6 landmarks on the frontal image and 11 in each profile image with the instructions presented by an interactive tool that highlights the proper location of the currently selected landmark on a sketch of a face. See Figure 1 for location of the landmarks.

Statistical Shape Models

Statistical shape modeling is a technique used to represent the shape of objects that possess prior general geometric information, but that may vary among a population. For example, the shape of human faces is different for every person, but follow a general pattern defining the overall location of eyes, nose, and mouth. This method has been explored for 2D and 3D shapes, and its main applications include object detection in images and regression for estimating object shapes [3,5,17]. The first is typically known as active shape model (ASM), while the second as shape or surface reconstruction. We explore statistical shape models in two different ways (2D and 3D), to detect contour points of facial features in the images and to reconstruct the final representation of the patient's face respectively. The 3D statistical shape model used here was created using 200 faces of young adults (100 male and 100 females) [3]. In order to create a statistical model out of shapes represented by vertices, a one-to-one correspondence between each vertex of the shapes had to be established (for further details on establishment of vertex correspondence the reader is referred to [3,5]). Assuming vertex corre-

spondence and a common coordinate system, statistics about variation among the shapes can be collected and used to reconstruct new shape instances. The model used in this work comprises a mean shape mesh and a mean texture along with two matrices of eigenvectors and their respective eigenvalues (for further details on statistical shape generation the reader is referred to [3]).

2D Feature Contour Points Detection

In this work, a 2D ASM is used to identify a set of points representing feature contours in the images [17]. The search for the best contour point is performed according to the smallest Mahalanobis distance between a profile surrounding the current vertex position and the base profile of the corresponding vertex. The base profile is generated out of a training set. The update of the new shape instance is performed by applying a set of weights. The weights are estimated from the difference of the current shape points and their new best positions, as proposed by Cootes *et al.* [5]. Convergence is achieved when there are no more changes between the new points and the new estimated shape.

One of the key challenges with ASM is the gathering of data to create and train the statistical shape model. Several databases with annotated facial images are available, but the number and position of annotated points are typically fixed and not suitable for the 3D reconstruction method used here. In order to obtain accurate 3D reconstructions, the approach presented here demands a ASM generated out of a database of images with frontal and profile images annotated with several facial feature points (e.g. eyes contours, mouth contours, silhouette contour, etc.). Since such a flexible database is not publicly available, we used the 3D statistical shape model to generate artificial data to train our 2D ASM. The artificial data not only eliminates the variability of the annotation process, but also allow for flexible optimization of the set of points used for 3D reconstruction since new images and annotations can be easily re-generated. A set of 6000 shapes were artificially generated by randomly varying shape and texture weights of the 3D statistical shape model. The 3D shapes were subsequently projected

from two different points of view: frontal, to simulate the frontal image; and profile, to simulate the lateral view. The backgrounds of these artificial images were replaced by one of 12 randomly selected images of different uniform walls in order to simulate a real scenario. In addition, a subset of vertices representing manually chosen feature contours of the 3D mean shape were defined. The selected vertices were carefully chosen to match their respective facial feature locations in the images. Finally, the projected 3D vertices representing feature contour points and images were used to train two 2D ASM, one for frontal images and one for the right profile images.

With the trained 2D ASMs, the search of the 2D feature contour points in a new image can be performed. Firstly, the mean shape comprising the points representing the feature contour points and comprising the points representing the manually annotated landmarks is aligned to the face according to the landmarks defined on the images by the physician. Finally, the algorithm iteratively searches for the optimal 2D shape. To ensure a stable search, the manually annotated landmarks are considered as ground truth. Therefore, points in the 2D shape representing a initial landmark are set to their respective manually annotated location at each iteration. The left profile image feature contour points are found by mirroring the left image along the vertical axis, applying the ASM for the right profile, and mirroring the contour points back. [See Appendix for additional feature search information].

2D to 3D Face Reconstruction

One of the challenges with dense 3D shape reconstruction is the computational time required to estimate the set of weights that best represent the desired face. For this reason iterative approaches [3,5], such as the one used in the previous section, do not provide a clinical acceptable processing time. To overcome such time constraints, Blanz *et al.* [2] have proposed a method for reconstructing dense 3D shapes from sparse data. The method presents a time efficient closed form solution for reconstructing 3D faces out of a set of points defined on a 2D image, but is limited to one image. To cope with the time requirements of the dynamical clinical

environment while also increasing the information for reconstruction, we adopt a similar approach based on multiple views, presented by Faggian *et al.* [7]. The multiple views reconstruction method allows for fast reconstruction of a 3D face out of a set of points representing facial features defined in images acquired from different points of views. Basically, an energy function averaging the contribution of points in each image and the prior knowledge of the shape of a face helps to find an optimal set of weights for the reconstruction in one step. [See Appendix for additional information on the energy function]. The set of weights is finally used to obtain the desired shape of the patient's face.

Texture Mapping

Original statistical shape model approaches [3] estimate the texture of the shape from a statistical texture model. However, faster and more realistic shape texture can be achieved by mapping real images of patients. With one frontal and two profile images, a good corresponding texture value can be found for each vertex of the shape representing the patient's face. Therefore, shape texture was mapped from the images acquired during the consultation.

Firstly, a surface parameterization algorithm (Floater Mean Value Coordinates [8]) was applied to the mean shape to define a offline transformation establishing correspondence between the shape vertices and the texture image. Such transformation is used afterwards to generate two intermediate texture images for each patient, one derived from the frontal image and one derived from the profile images. Finally, the two textures are blended in one texture image using a multiband filter [4]. [See Appendix for the formulation of the texture mapping].

3D Visualization and Planning Tools

The 3D visualization and planning is very important to the physician because it is where he (she) will continuously interact with the system to discuss with the patient. Therefore, a clear and responsive tool is essential to maintain clinical usability. In order to cope with these requirements with a web based application, the visualization and planning tools are implemented with Unity 3 (a environment for high-end 3D web-based game development). As a plug-

in, Unity 3 enables 3D rendering such as lightning, and mesh manipulation on standard web browsers with performance comparable to state-of-the-art platform applications.

Four main planning tools were developed: one for rhinoplasty, one for skin fillers, one for dermabrasion (skin cleaning) and one for comparing before and after planning. The tool for rhinoplasty allows the nose to be manipulated using pre-defined points that are typically changed during plastic interventions. The pre-defined points are used as control points for local interpolation of the 3D mesh and deformation of the nose. The tool for skin filling allows regions on the skin to be delineated and filled with a certain volume that is evenly distributed along the selected region. The injected volume is not intended to represent the volume to be injected in reality because of difficulties related to absorption and other factors, but rather to illustrate the difference between pre- and post-procedure. The dermabrasion tool allows wrinkles and undesired marks to be removed, as well as rejuvenation of the skin. Basically, a 3D brush selects a circular region to be smoothed on the skin. The region is then mapped to the textured image where a gaussian filter is applied. With the comparison tool, physicians and patients can visualize the intended effect of the intervention with pre- and post-planning situations displayed side by side. Additional tools enable measurements of distances between two points considering straight lines or along the surface (geodesic paths).

2.2 **Experiments**

In order to evaluate our application, three types of data have been used as ground truth (see Table 1): in-model data (IMD), out-model registered data (OMRD) and out-model non-registered data (OMNRD). For IMD, the initial landmarks were automatically generated using their ground truth location with a gaussian noise (sigma equals to 2 and cropped at 4 pixels) and reconstruction was performed automatically without considering the feature contour correction step (illustrated in Figure 1). For OMRD and OMNRD, reconstruction was performed by an expert for each case. The time required to perform each step of the pipeline was measured. Finally, the reconstructed faces were compared to the ground truth for each case as

follow. Firstly, ground truth and reconstructed shapes were aligned considering eye, nose and mouth segments of the 3D statistical shape model [3]. Secondly, distances for all three datasets were measured from vertices of the reconstructed face to their corresponding point in the ground truth surface. For IMD and OMRD (with one-to-one vertex correspondence between ground truth and reconstruction), the shapes were aligned with Procrustes [10]. For OMNRD (without vertex correspondence), the shapes were aligned with iterative closest point (ICP) [1]. The vertex correspondence of IMD and OMRD were not directly used for distance measurement because the correspondence cannot be ensured in flat areas such as cheeks and forehead after reconstruction. Therefore, two different methods were used to find vertex correspondence in all three datasets [14]: closest point matching (CPM), which considers the closest point in the ground truth surface as corresponding point; and thin plate spline plus closest point matching (TPS+CPM), which first warps the reconstructed face with a TPS transformation and a set of landmarks, and subsequently finds the closest point on the ground truth surface. The former is a direct method that is not influenced by human error, nevertheless it does not ensure correct anatomical correspondence. The latter relies on the manual definition of landmarks, but presents a better anatomical correspondence. Since the distance measured from corresponding points found by TPS+CPM is not necessarily to the closest point between the two surfaces, but a more anatomically relevant distance, it should result in higher values than the ones found by CPM. A total of 15 validation landmarks were defined in the reconstructed and in the ground truth shapes (see Figure 2 in the Electronic Supplementary Material). In addition to the distance measurements, a visual analysis was performed by two plastic surgeons in each of the cases from OMNRD to support the qualitative results. The surgeons rated each of the reconstruction according to the values presented in Table 2 while comparing to the ground truth and to the pictures. In a last step, the 3D planning tools were evaluated qualitatively on the reconstructed cases.

3. **Results**

The average time necessary to obtain the 3D face once the 2D images were uploaded to the application was 297.79 ± 90.49 seconds. This time has been divided among different individual steps of the application: manual definition of the facial landmarks (94.32 ± 36.45 seconds), 2D feature contour points detection (8.50 ± 3.99 seconds), manual correction of the feature contours (191.52 ± 70.59 seconds), 3D face reconstruction (0.71 ± 0.39 seconds), and texture mapping (0.83 ± 0.23 seconds).

The average reconstruction error over all cases measured with CPM was below 2 mm for all datasets type. Peaks of up to 2.1 mm per region were noticed for the individual case errors (except for the 299th worst case that presented 2.88 mm error for the mouth region before manual correction of the feature contours, but below 2 mm after manual correction). The average reconstruction error over all cases measured with TPS+CPM was below 2 mm for subjects of IMD, OMRD, and below and 3 mm for OMNRD. The distances calculated with CPM and TPS+CPM vertex correspondence representing the reconstruction error are presented in different graphs according to dataset type, see Figure 3 and Figure 4 respectively. Figure 5 and Figure 6 show respectively the good and bad examples of reconstructions of each dataset type. The distance maps present small errors around the eyes and chin region. There exist larger errors in the face region around the cheek and forehead since the current method does not use information from those regions for reconstruction. The neck and ears region also showed large errors, but are not considered in the analysis since such regions have only been used to complete the appearance of the face. It is worth mentioning that the errors of the IMD cases could still be improved since no manual corrections were performed for the feature contours. Visual inspection of the 2D feature contour points detection showed that automatic detection of the feature contour failed considerable in 9 % of the IMD cases. Therefore, the reconstruction of such cases could be significantly improved after manual the corrections, see Figure 7 for two examples.

According to the visual analysis performed by 2 surgeons, all reconstructed cases from the OMNRD could be used for communicating with the patient, although some of them presented sub-optimal reconstruction. Out of the 28 real cases reconstructed, 1 and 2 cases were evaluated as a "Bad" reconstruction by surgeon 1 and 2 respectively. No cases were evaluated as "Very Bad" or "Excellent". The average evaluation were in between "Good" and "Very Good" with values of 3.54 and 3.32 for surgeon 1 and 2 respectively. Examples of reconstruction paired with the respective grade attributed by the surgeons are presented in Figure 8. According to the surgeons, the reconstruction of some cases gave different impressions when analyzing from different points of view (e.g. frontal and profile). For example, case 1 from Figure 8 (graded as "Very good" by surgeon 2) gives better frontal impression than profile, while case 3 (graded as "Very good" by surgeon 2) gives better profile impression. From the reconstructions, only case 5 diverged significantly between the surgeons ("Very good" by surgeon 1 and "Bad" by surgeon 2). While for surgeon 1 the overall appearance of the face was very well captured, for surgeon 2 it did not replicate the nose very well from profile and it did not capture the face appearance from frontal view. Another example of a "Bad" reconstruction for surgeon 2 can be seen in case 6. The same case was considered "Good" by surgeon 1 with a better profile view than frontal (face appeared thinner than subject). Case 3 was considered only "Good" by both surgeons because of differences on the facial curves around the cheek region.

The planning tools enabled emulation of various aesthetic procedures. The rhinoplasty arrows that are located in crucial points typically considered for intervention allowed for easy manipulation of the nose in 3D. Simple emulation of filling procedures could be achieved by delineating the region to be filled and varying the amount of filling to be injected. Wrinkles could be quickly removed from the patient skin considering certain regions of the face. The planned procedure could be directly visualized on the 3D face from different angles. Additionally, pre- and post procedure emulation could be compared side by side in order to em-

phasize the modifications achieved. An illustration of the results of the planning tools validation on a random case are displayed in Figure 9.

4. **Discussion**

This paper presents the first results of a web-based computer assisted system for aesthetic procedure consultations that enables physicians to emulate different procedures on a virtual 3D representation of the patient's face. The application aims to facilitate communication between physicians and patients. The simple workflow requiring no additional expensive and complicated hardware is a great advantage for plastic surgeons, dermatologists, and aesthetic professionals not having the resources to acquire current available approaches, or reluctant to adopt such complex technologies. Current hand-held scanner are still optimized for accuracy and lack on usability for clinics. Web based applications provides not only world wide access allowing for online discussions between physicians but also simplify upgrades and maintenance (considered a highlight in the clinical community) since doctors only need to login and use the application. The pipeline is based on standard 2D images that are already part of the standard clinical workflow requiring therefore no additional steps. The automatic steps of the application are performed within a few seconds and are of no concern in this scenario. Among the automatic methods, this work proposes a feature contour detection that takes advantage of synthetic data generated by the 3D statistical model to facilitate the engineering of applications similar to the presented here. Our results have showed that physicians are able to reconstruct faces of patients in less than an average of five minutes, which allows the application to be used within standard consultation time. Since the application is going to run on a server and the input images are scaled to a standard size (e.g. interpupillary distance) before processing, the processing time of those steps is expected to be similar for different cases. the manual Currently, the most time consuming parts of the procedure are the manual definition of the facial landmarks and correction of the facial feature contours (averaging around two and three minutes respectively). No difficulties with those steps were reported from volun-

teers who tested the application since they are facilitated by semi-supervised methods (e.g. spline contours). The experiments with IMD cases, showed that the manual correction of the facial feature contours can improve the reconstruction results (see Figure 7) but they were not necessary for most of the cases. Therefore, faster less accurate reconstructions can be obtained without manual correction depending on the need of each user. Furthermore, automatic detection of manually defined landmarks is part of future improvements of the system.

In this study, two distance measures were used (considering CPM and TPS+CPM point correspondence) to compare reconstruction and ground truth. The comparison with the ground truth allowed for identification of the distribution of error in the face. The graphs showed that from the three regions the mouth has usually higher error. Distance maps showed that the cheek and the forehead regions concentrate most of the errors. However, as it can be seen in Figure 5, Figure 6 and Figure 8, the texture mapping plays an important role on the overall perception of the face. The texture seems to minimize perception of small errors of the shape reconstruction. From our experiments, it was noticed that imperfections are less perceived when casually examining the reconstructed face rather than thoroughly analyzing it, which is usually the case when communicating a certain procedure to the patient in a given dynamic clinical scenario. The reconstructions in our evaluation presented a very stable texture mapping. None of the cases showed major texture problems such as background as part of the face or stitching effect even for the IMD cases without manual correction of the feature contours (example in Figure 7). The qualitative analysis performed by two surgeons on the reconstructions showed that most of the cases were evaluated as "Good" or "Very Good", which supports the use of the application in clinics. Although there were sub-optimal reconstructions, the surgeons would still use it to discuss with the patient. According to the surgeons, "Bad" reconstructions would reduce the visual impact of the application, but not hinder its use as a communication tool. From the clinical point of view, some reconstructions gave different impression when analyzed from different sights, which could make them less

suitable for discussing certain procedures than for others. For example, a case with better frontal than profile impression could be used for communicating skin clearing or rejuvenating better than for other procedures. Although texture could reduce some of the perceived reconstruction errors, large errors in the shape can still affect the overall appearance of the face. For example, case 6 of Figure 8 is also illustrated as sub-optimal reconstruction in Figure 6 with large errors on the cheek bone region. Such errors in shape made the subject look thinner than he actually is and reduced the score gave by surgeon 2. According to the surgeons, rhinoplasty or other profile altering surgical procedures rely more on profile view and therefore on shape reconstruction accuracy. Hence, accurate shape reconstructions (illustrated on Figure 3 and Figure 5) can facilitate discussions on rhinoplasty increasing the power of the application as well as giving a better overall impression to patients and physicians. None of the cases were evaluated as "Excellent", showing that there are limitations on the actual face appearance reproduction accuracy when compared to 3D scanner devices that require more complex setup and post-processing. On our results, errors seemed to be higher in subjects with features not belonging to the population used to create the 3D statistical shape model used in this work (200 young Caucasians). Therefore, future work includes extending the range of face representations of our application by expanding the current 3D statistical shape model and by creating similar models for different races. Additionally, wrinkles are typically not reconstructed on the shape in the current version since the model was mostly created with young subjects. Therefore, wrinkles are represented as texture only.

The planning tools were created using feedback from plastic surgeons and optimized for fast and intuitive 3D operations. Mainly, the proposed operations can be performed on real time on the web browser of common personal computers. The application offers possibilities of emulating filling, skin clearing or rejuvenation, and rhinoplasty procedures. Additionally, visualization tools allow the user to compare pre, and post-intervention scenarios in a syn-

chronized way, which enriches the decision making of the physician and the communication to the patient.

In summary, we have presented the first results of a developed web-based 2D to 3D facial reconstruction tool which provides sufficiently high precision for communication between physician and patients for visualization of facial treatment options. Patient understanding about the aesthetic procedure, and consequently satisfaction with the consultation, is expected to increase with the use of 3D virtual face representation and procedure planning. The current results warrant further evaluation of the application in clinical setting with evaluation of this novel method at large scale by physicians, aesthetic professionals and patients.

5. Acknowledgments

We acknowledge the support of the Swiss KTI Promotion Agency (grant: 12892.1 PFLS-LS).

6. References

1. Besl, P. J., and H. D. McKay. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14:239–256, 1992.
2. Blanz, V., A. Mehl, T. Vetter, and H. –. Seidel. A statistical method for robust 3D surface reconstruction from sparse data. *In int. symp. on 3D data processing, visualization and transmission* 293–300, 2004.
3. Blanz, V., and T. Vetter. A morphable model for the synthesis of 3D faces. *Proceedings of the 26th annual conference on Computer graphics and interactive techniques* 187–194, New York, NY, USA, 1999.
4. Burt, P. J., and E. H. Adelson. A multiresolution spline with application to image mosaics. *ACM Trans. Graph.* 2:217–236, 1983.
5. Cootes, T. F., C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models - Their training and application. *Comput. Vis. Image Underst.* 61:38–59, 1995.

6. Darisi, T., S. Thorne, and C. Iacobelli. Influences on Decision-Making for Undergoing Plastic Surgery: A Mental Models and Quantitative Assessment. *Plastic and Reconstructive Surgery* 116:907–916, 2005.
7. Faggian, N., A. Paplinski, and J. Sherrah. 3D Morphable Model fitting from multiple views. *8th IEEE International Conference on Automatic Face & Gesture Recognition, 2008. FG '08* 1–6, Amsterdam, Netherlands, 2008.
8. Floater, M. S., and K. Hormann. Surface Parameterization: a Tutorial and Survey. In: Advances in Multiresolution for Geometric Modelling, edited by N. A. Dodgson, M. S. Floater, and M. A. Sabin. Springer Berlin Heidelberg, 2005, pp. 157–186.
9. Flury, B., and H. Riedwyl. Multivariate Statistics: A Practical Approach. Chapman and Hall, 1988, 296 pp.
10. Gower, J. Generalized procrustes analysis. *Psychometrika* 40:33–51, 1975.
11. Heike, C. L., K. Upson, E. Stuhaug, and S. M. Weinberg. 3D digital stereophotogrammetry: a practical guide to facial image acquisition. *Head & Face Medicine* 6:18, 2010.
12. De Heras Ciechomski, P., M. Constantinescu, J. Garcia, R. Olariu, I. Dindoyal, S. Le Huu, and M. Reyes. Development and Implementation of a Web-Enabled 3D Consultation Tool for Breast Augmentation Surgery Based on 3D-Image Reconstruction of 2D Pictures. *Journal of Medical Internet Research* 14:, 2012.
13. Jebara, T., A. Azarbayejani, and A. Pentland. 3D structure from 2D motion. *Signal Processing Magazine, IEEE* 16:66 –84, 1999.
14. Kim, H., P. Jürgens, S. Weber, L.-P. Nolte, and M. Reyes. A new soft-tissue simulation strategy for cranio-maxillofacial surgery using facial muscle template model. *Progress in Biophysics and Molecular Biology* 103:284–291, 2010.
15. Koch, R. M., M. H. Gross, F. R. Carls, D. F. von Büren, G. Fankhauser, and Y. I. H. Parish. Simulating facial surgery using finite element models. *Proceedings of the 23rd an-*

nual conference on Computer graphics and interactive techniques 421–428, New York, NY, USA, 1996.

16. Lin, S. J., N. Patel, K. O'Shaughnessy, and N. Fine. A New Three-Dimensional Imaging Device in Facial Aesthetic and Reconstructive Surgery. *Otolaryngology -- Head and Neck Surgery* 139:313–315, 2008.
17. Milborrow, S., and F. Nicolls. Locating Facial Features with an Extended Active Shape Model. *Proceedings of the 10th European Conference on Computer Vision: Part IV* 504–513, Berlin, Heidelberg, 2008.
18. Moghaddam, B., J. Lee, H. Pfister, and R. Machiraju. Model-based 3D face capture with shape-from-silhouettes. *Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on* 20 – 27, Nice, France, 2003.
19. Ozkul, T., and M. H. Ozkul. Computer simulation tool for rhinoplasty planning. *Computers in Biology and Medicine* 34:697–718, 2004.
20. Rabi, S. A., and P. Aarabi. Face Fusion: An Automatic Method For Virtual Plastic Surgery. *Information Fusion, 2006 9th International Conference on* 1 –7, Florence, Italy, 2006.doi:10.1109/ICIF.2006.301579
21. Salzmann, M., J. Pilet, S. Ilic, and P. Fua. Surface Deformation Models for Nonrigid 3D Shape Recovery. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29:1481 –1487, 2007.
22. See, M., M. Foxton, N. Miedzianowski-Sinclair, C. Roberts, and C. Nduka. Stereophotogrammetric measurement of the nasolabial fold in repose: a study of age and posture-related changes. *European Journal of Plastic Surgery* 29:387–393, 2007.
23. Torresani, L., A. Hertzmann, and C. Bregler. Nonrigid Structure-from-Motion: Estimating Shape and Motion with Hierarchical Priors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 30:878 –892, 2008.

24. Zhang, R., P.-S. Tsai, J. E. Cryer, and M. Shah. Shape from Shading: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21:690–706, 1999.

7. **Appendix**

Let $s = (s_1, ..., s_m)$ be a set of m shapes represented by p corresponding vertices $s_i = v = (v_1, ..., v_p)^T$ where $v_i \in \mathbb{R}^3$ and represent x, y and z coordinates. New shape instances v where $v \notin s$, can be created with a linear combination of weights

$$v = \bar{v} + P^v \, diag(\sigma^v) b^v, \qquad 1$$

where, $\bar{v} = \frac{1}{m}\sum_{i=1}^{m} s_i$ is the mean shape, $P^v = (P^v{}_1, ..., P^v{}_m)$ is a matrix of eigenvectors and $diag(\sigma^v)$ is a diagonal matrix with the respective eigenvalues that can be obtained by applying principal component analysis (PCA) [9] to the m shapes, and $b^v = (b^v{}_1, ..., b^v{}_m)^T$ is a vector of weights where $b^v{}_i \in \mathbb{R}$.

With an analogous approach, the texture values of the vertices of these shapes $s^t{}_i = t = (t_1, ..., t_p)^T$, where $t_i \in \mathbb{R}^3$ and represent r, g and b values, can modeled as a linear combination of weights, $b^t = (b^t{}_1, ..., b^t{}_n)^T$. New texture instances can be estimated as

$$t = \bar{t} + P^t diag(\sigma^t) b^t, \qquad 2$$

where, $\bar{t} = \frac{1}{m}\sum_{i=1}^{m} s^t{}_i$ is the mean texture, $P^t = (P^t{}_1, ..., P^t{}_n)$ is a matrix of eigenvectors and is $diag(\sigma^t)$ is a diagonal matrix with the respective eigenvalues that can be obtained by applying PCA to the m shape textures.

A 2D statistical shape model can be created with a similar approach, but considering a set of h feature contour points $f^a = (f_1, ..., f_h)^T$, where $f_i \in \mathbb{R}^2$ and represents x and y coordinates, and $a \in$ {"*fi*", "*ri*", "*li*"} indicating frontal, right profile and left profile images respectively, to be automatically detected in the images. Let $l^a = (l_1, ..., l_k)^T$, where $l_i \in \mathbb{R}^2$, $l_i \subset f_i$, and represents x and y coordinates, be a set of k manually defined landmarks. The 2D ground truth location of the facial feature contour points f^a used for training the 2D ASM is calculated as

$$f^{fi} = T^{pfi}\ T^{sfi}v,\ f^{ri} = T^{pri}\ T^{sri}v, \quad 3$$

where, T^{pfi} and T^{pri} represent the frontal and right profile 3D to 2D projections respectively, T^{sfi} and T^{sri} are transformations (manually defined offline) to select a subset of vertices representing facial feature on the 3D mean shape (i.e. the landmarks l^{fi}, l^{ri} and l^{li}, and additional features such as eyes contour, mouth contour, etc). Random shapes and texture used for training of the 2D ASM were generated by varying the weights b^v and b^t in equations 1 and 2 according to a normal distribution.

The 2D facial feature contour points search was performed in two steps. Firstly, an initial alignment of the mean shape, $\overline{f^{fi}}$ or $\overline{f^{ri}}$, with the manually defined landmarks, l^{fi} or l^{ri}, was performed following

$$f^{fi*} = TPS(PROC(\overline{f^{fi}}, l^{fi}), l^{fi}),\ f^{ri*} = TPS(PROC(\overline{f^{ri}}, l^{ri}), l^{ri}), \quad 4$$

where, f^{fi*} and f^{ri*} are the initial position of the shape search for the frontal and right profile respectively, $TPS(m,n)$ is a thin plate spline transformation of the point set *m* to a subset of control points *n*, and $PROC(m,n)$ is a Procrustes transformation of the point set *m* to a subset of points *n*. Secondly, an iterative process [5] searches for the optimal 2D shape, i.e. f^{fi}, f^{ri} and f^{li} for frontal, right profile and left profile respectively.

With the set of points representing 2D facial feature contours from different point of views, the optimal weights b^v can be found by minimizing the energy function [7]:

$$\begin{aligned} E = & \frac{1}{3}\left\| T^{pfi}\ T^{sfi}\left(P^v\, diag(\sigma^v) b^v - \bar{v}\right) - (f^{fi} - T^{pfi}\ T^{sfi}\bar{v})\right\|^2 + \\ & \frac{1}{3}\left\| T^{pri}\ T^{sri}\left(P^v\, diag(\sigma^v) b^v - \bar{v}\right) - (f^{ri} - T^{pri}\ T^{sri}\bar{v})\right\|^2 + \\ & \frac{1}{3}\left\| T^{pli}\ T^{sli}\left(P^v\, diag(\sigma^v) b^v - \bar{v}\right) - (f^{li} - T^{pli}\ T^{sli}\bar{v})\right\|^2 + \\ & \eta\|b^v\|^2, \end{aligned} \quad 5$$

where, the first three lines represent the contribution of contour points in each image (frontal, right profile and left profile respectively) and the last line represents a prior to keep the de-

sired shape close to the shape of a human face, in this case the mean shape. The last line is necessary because the energy function can converge to different minimums, and the optimum minimum when assuming an error on locating the facial feature contour points might result on a non-face-like shape. A closed form solution to solve the equation above in one step can be found in [2,7]. Finally, b^v is replaced in the in equation 1 to reconstruct the 3D shape of the patient's face.

After the final patient's face shape is reconstructed, the texture mapping is performed as follows. Firstly, two intermediate textures are generated, t^{fi} frontal and t^{pi} profile. See equations 6 and 7.

$$t^{fi} = RGB(PROC(\ T^{pfi}v, f^{fi}), I^{fi}), \tag{6}$$

where, I^{fi} is the frontal image of the patient, and *RGB*(*m*, *n*) is a function that gets the list of *r*, *g*, and *b* values out of the image *n* at the locations *m*.

$$t^{pi} = RGB_{lt}(PROC(\ T^{pri}v, f^{ri}), PROC(\ T^{pli}v, f^{li}), I^{ri}, I^{li}), \tag{7}$$

where, I^{ri} and I^{li} are the right and left profile images of the patient respectively, and $RGB_{lt}(m_r, m_l, n_r, n_l)$ is the function that gets the list of *r*, *g*, and *b* values out of the image, n_r or n_l, at the locations m_r or m_l depending whether v_i is located on the left or right side of the shape. Finally, the two intermediate textures are blended in one texture image I^{tx}

$$I^{tx} = MULTIBAND(\ T^{\bar{v}tx}t^{fi},\ T^{\bar{v}tx}t^{pi}, I^{msk}), \tag{8}$$

where, I^{msk} is a mask (generated offline) separating frontal and profile portions of a face in the I^{tx} space, $T^{\bar{v}tx}$ is a transformation to map shape vertices considering the mean shape to a texture image (surface parameterization [8] obtained offline using the mean shape), and *MULTIBAND*(n_f, n_l, n_m) is a multi band multiband filter [4] that blends images n_f and n_l according to the mask image n_m.

8. Tables, figure legends

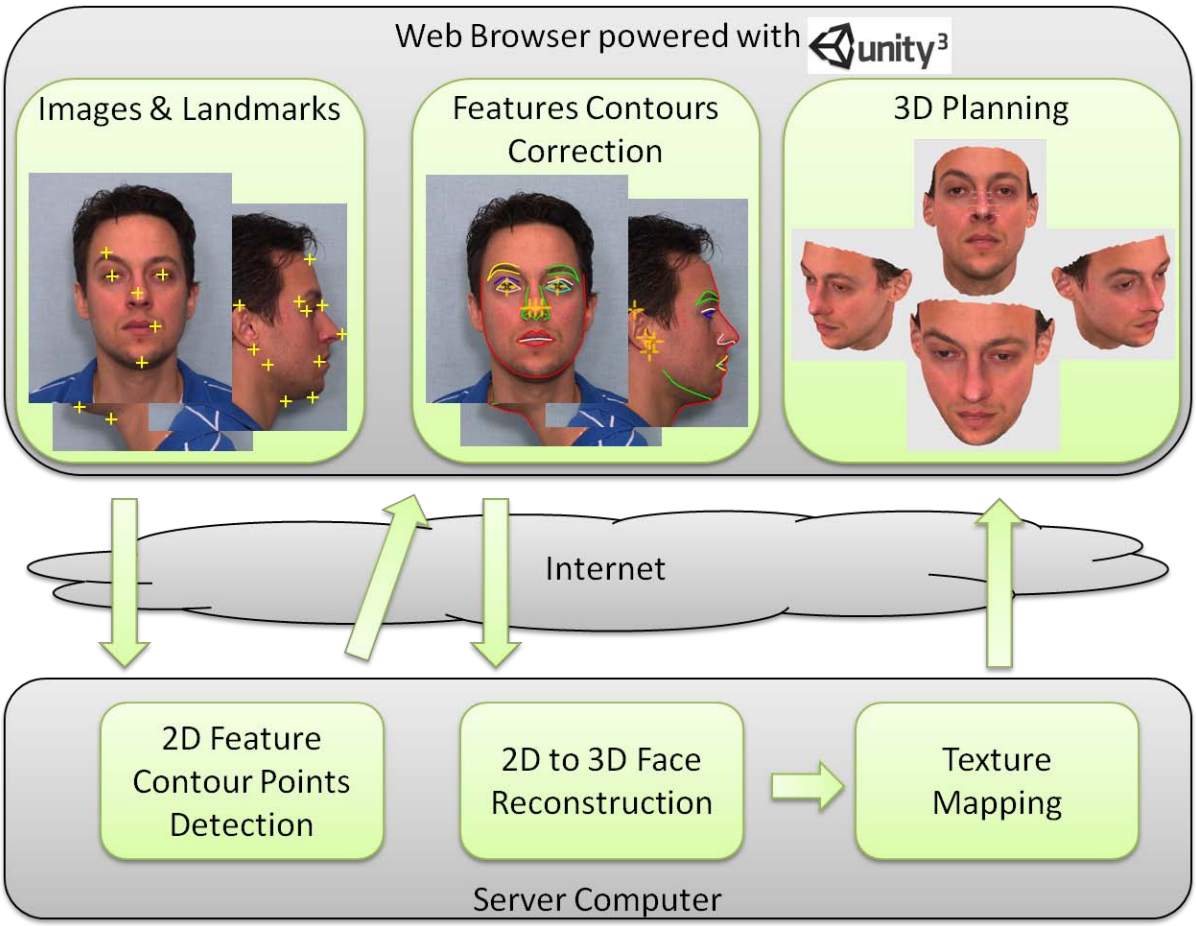


Figure 1: Overview of the data flow in the different steps of the application that is divided in two layers separated by the internet cloud: the web browser powered with "unity 3" at the physician's computer and server computer providing the web service. The image & landmarks box highlights the landmarks to be manually defined (yellow crosses): 6 frontal (right eyebrow, eye centers, nose tip, left mouth corner, and chin) and 11 in each profile image (top of the forehead, inflection of the nose with forehead, end of the eyebrow, eye corner, tip of the nose, corner of the mouth, connection of the chin with the neck, back part of the jaw, bottom and top of the ear, and neck inflection). The feature contours correction box highlights the splines for correction of the feature contour points. The 3D Planning box highlights the final 3D shape for planning.

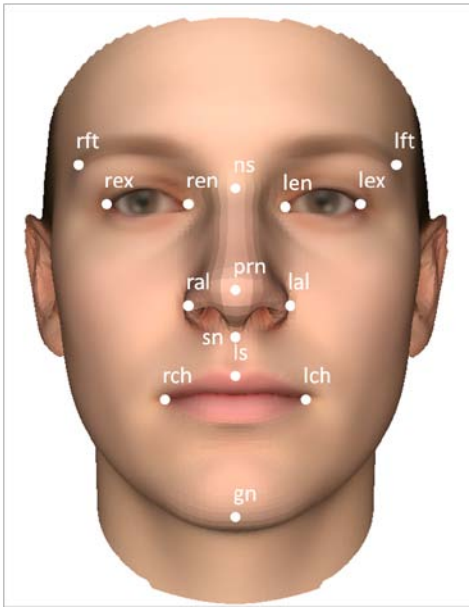


Figure 2: Location of the 15 landmarks used for the TPS+CPM metric: rft and lft, right and left frontotemporale; rex and lex, right and left exocanthion; ren and len, right and left endocanthion; na, nasion; prn, pronasale; ral and lal, right and left, alare; sn, subnasale; ls, labiale superius; rch and lch, right and left cheilion; and gn, gnathion. {To be available as Electronic Supplementary Material}

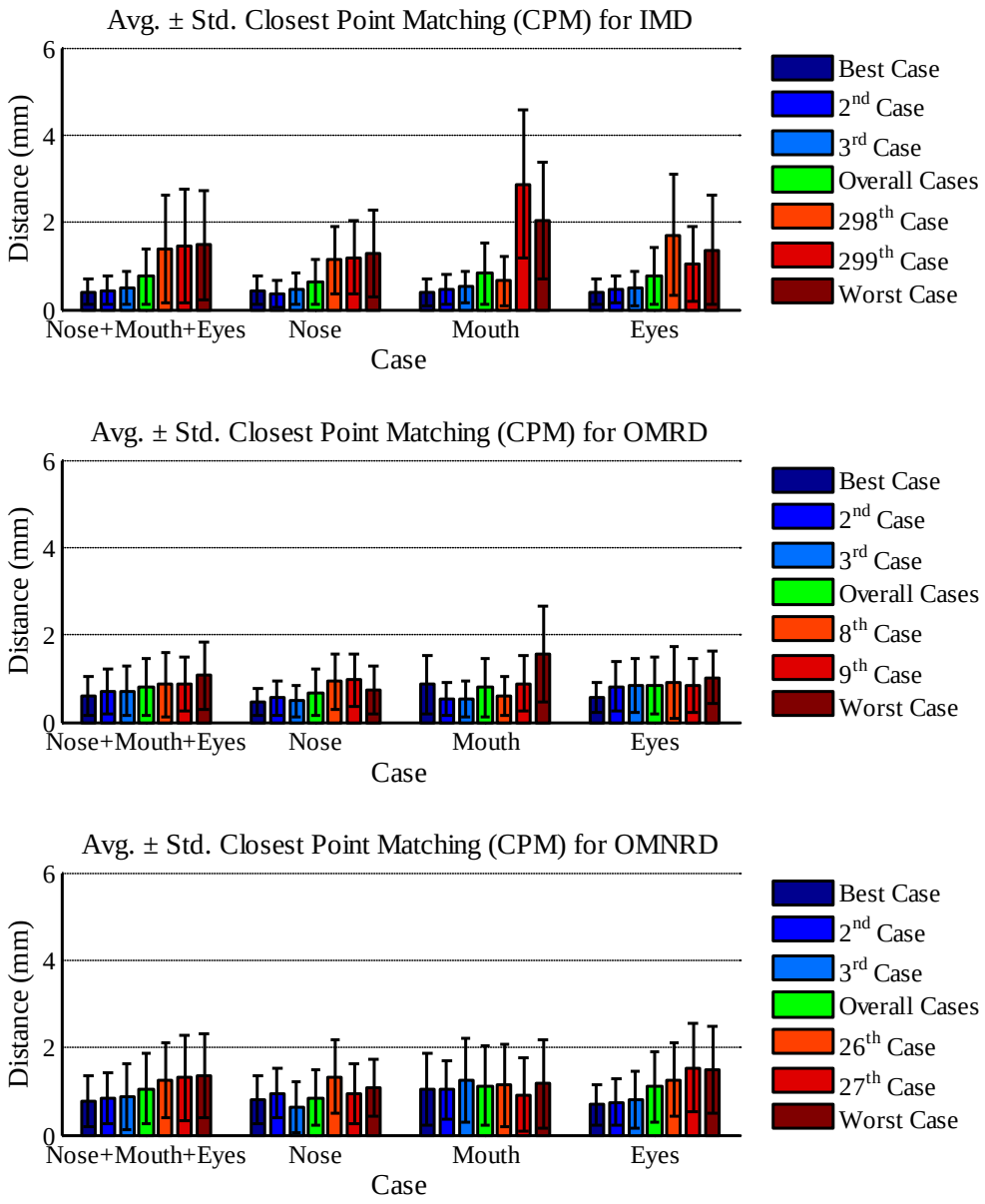


Figure 3: Three graphs (one per dataset type) illustrate the average CPM distance and standard deviation for the three best cases (blue bars with one bar per case), overall cases (green bar with average considering all cases) and for the three worst cases (red bars with one bar per case) grouped by segments. Classification of the best and worst cases is according to the Nose+Mouth+Eyes region. Nose+Mouth+Eyes represents the error considering vertices from nose, mouth and eyes segments. Nose, Mouth and Eyes represents the error considering vertices from each of the segments individually (nose, mouth and eyes respectively). Different colors represent different subjects.

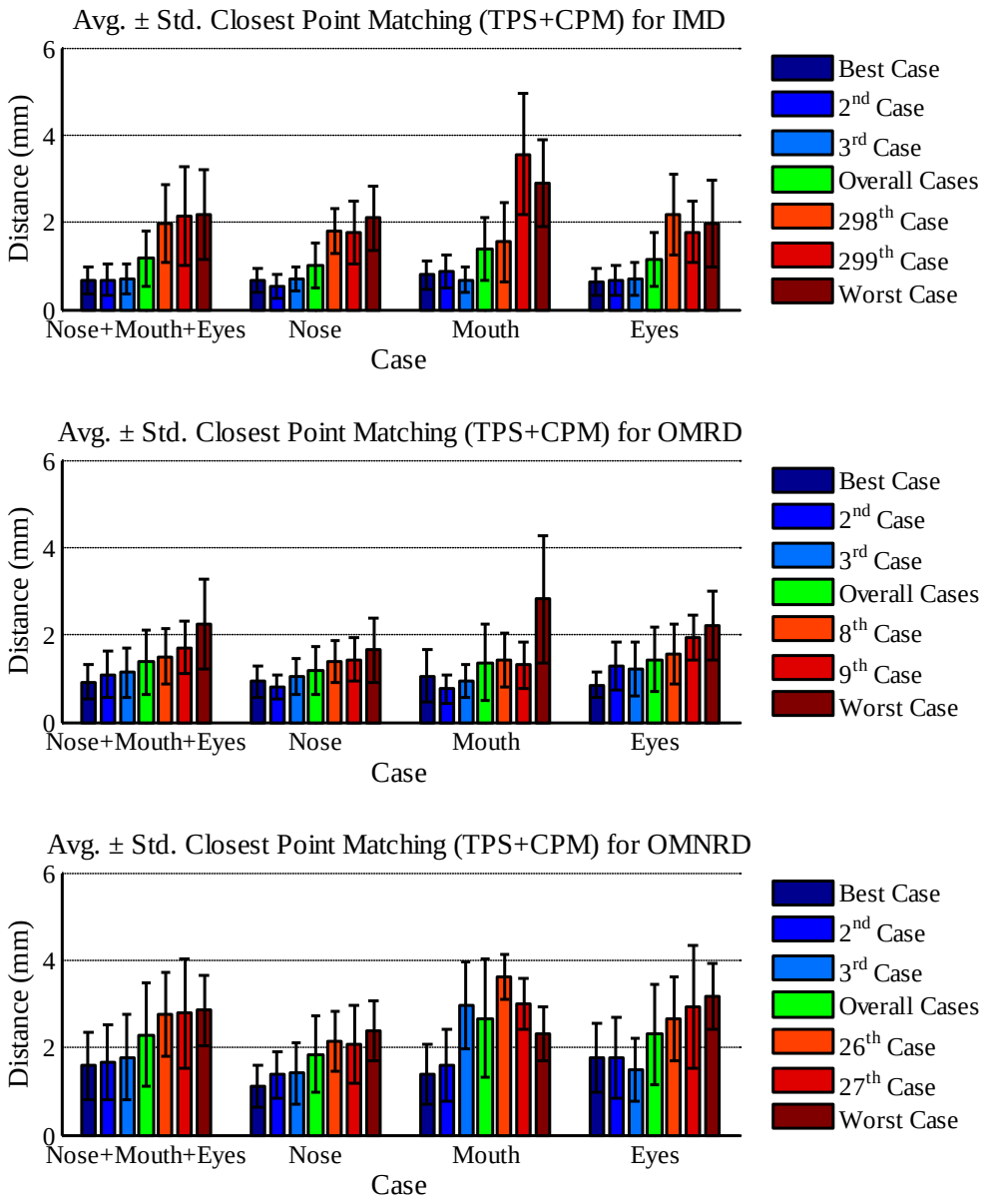


Figure 4: Three graphs (one per dataset type) illustrate the average TPS+CPM distance and standard deviation for the three best cases (blue bars with one bar per case), overall cases (green bar with average considering all cases) and for the three worst cases (red bars with one bar per case) grouped by segments. Classification of the best and worst cases is according to the Nose+Mouth+Eyes region. Nose+Mouth+Eyes represents the error considering vertices from nose, mouth and eyes segments. Nose, Mouth and Eyes represents the error considering vertices from each of the segments individually (nose, mouth and eyes respectively). Different colors represent different subjects.

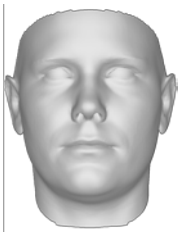
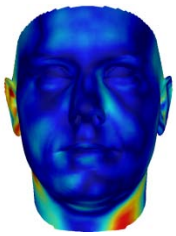
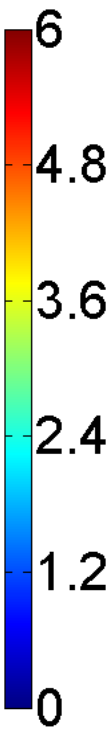
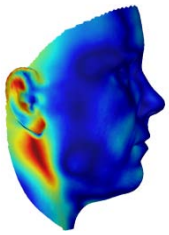
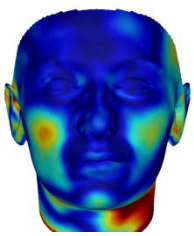
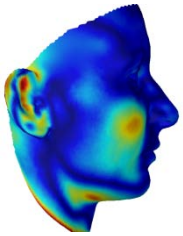
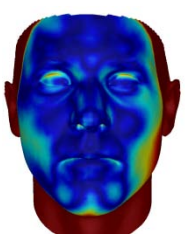
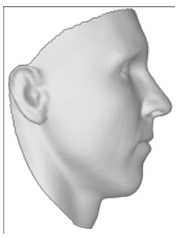
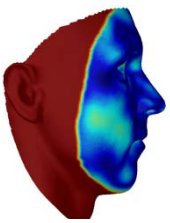
Dataset	2D Input	3D Reconstruction Shape & Texture	3D Reconstruction Shape	Distance Map (CPM)	Color Bar (mm)
IMD					
					
OMRD					
					
OMNRD					
					

Figure 5: Example of good reconstruction results showing the input 2D pictures, the reconstructed 3D face with and without texture, and the distance map.

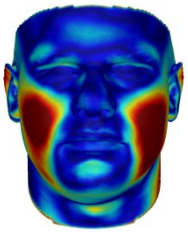
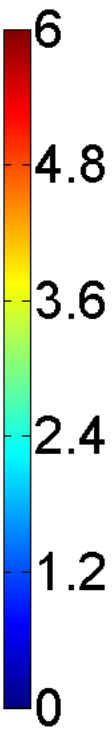
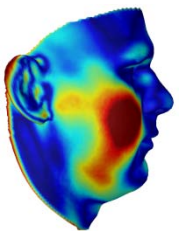
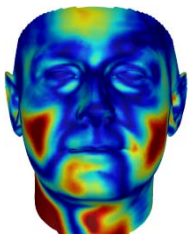
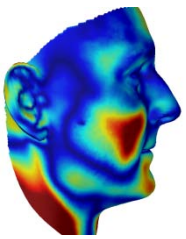
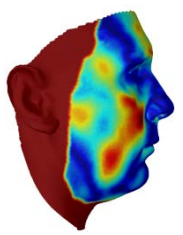
Dataset	2D Input	3D Reconstruction Shape & Texture	3D Reconstruction Shape	Distance Map (CPM)	Color Bar (mm)
IMD					
					
OMRD					
					
OMNRD					
					

Figure 6: Example of bad reconstruction results showing the input 2D pictures, the reconstructed 3D face with and without texture, and the distance map.

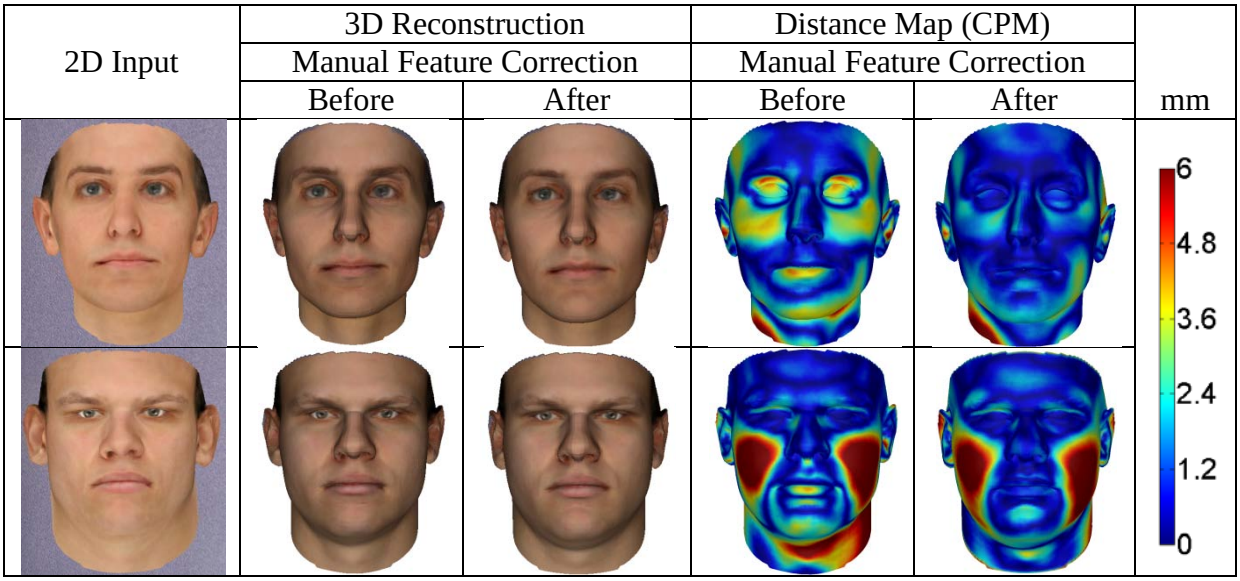


Figure 7: Example of reconstruction improvement with manual feature contour correction of two IMD cases. Case one (first row) show most improvements in the eyes and lips region, while case two (second row) show most improvements in the mouth region.

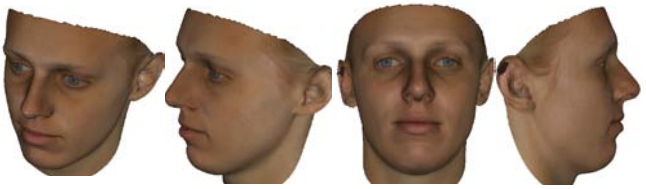
#	Original Images	3D Reconstruction	Surg. Eval.	
			1	2
1			4	4
2			4	3
3			3	4
4			3	3
5			4	2
6			3	2

Figure 8: Example of reconstructed cases. From left to right, the figure shows the original images (part of the input), the respective reconstructed 3D face from 4 points of view, and the evaluation of surgeons 1 and 2 according to Table 2.

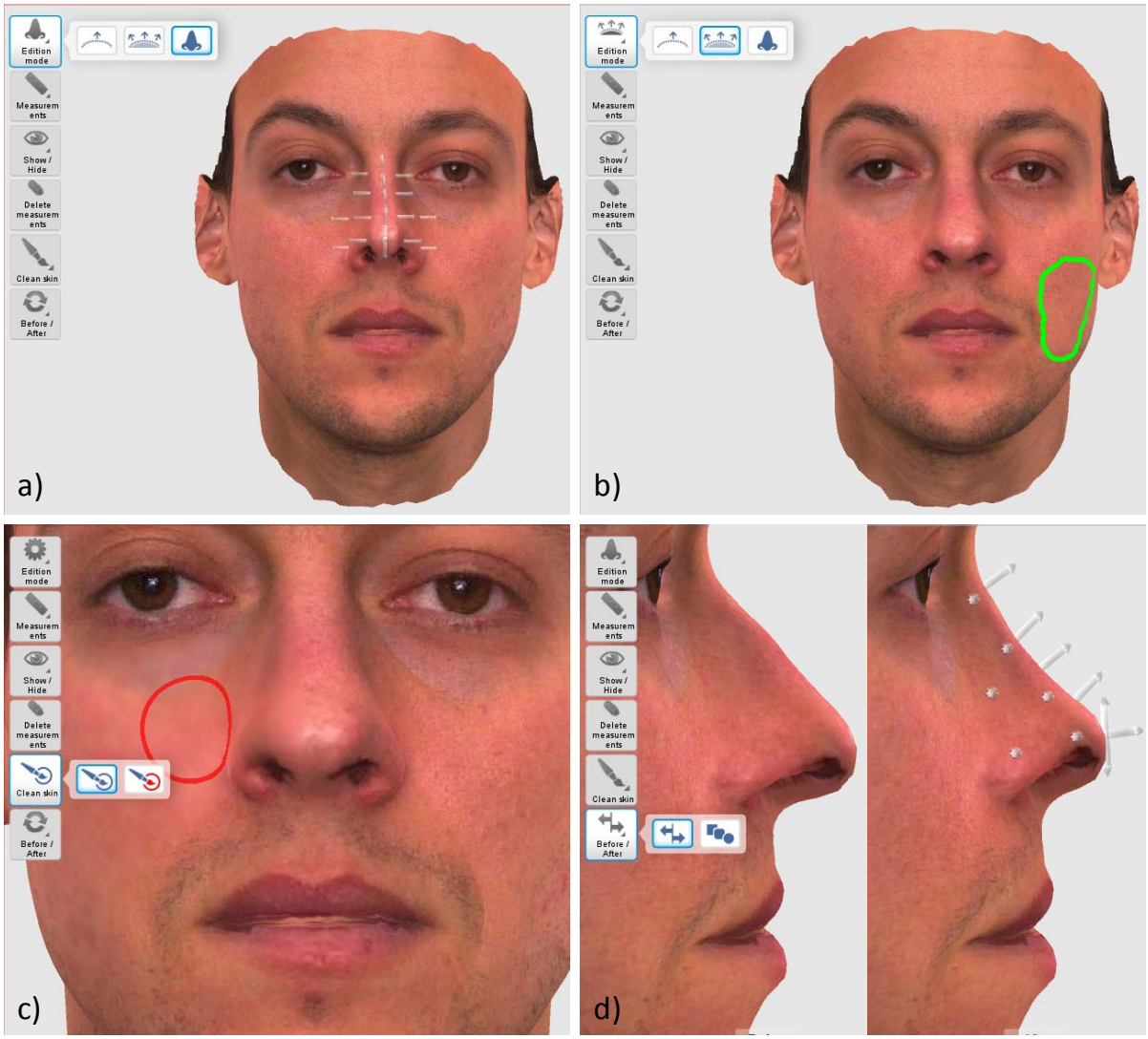


Figure 9: Illustration of the 3D planning tools. a) shows tools for nose correction that can be used for rhinoplasty. b) shows tools for emulating filling procedures in which physicians define a region and an amount to be filled. c) shows a tool for cleaning the skin. d) shows a comparison of pre- and post procedure planning.

Table 1: Description of datasets used for evaluation.

Dataset	Data Description	2D Input Images	Aim
IMD	A set of 300 face shapes artificially generated by randomly varying the parameters of the 3D statistical shape model using a normal distribution. For further details the reader is referred to [3].	Artificially created frontal, right profile and left profile images (972 x 1280 pixels and eye distance of around 240 pixels) .	Allow verification of the capability of the application to reproduce a well known face.
OMRD	A set of 10 face shapes provided with the 3D statistical shape model [3] and therefore with one-to-one vertex correspondence to the mean shape, but out of the faces used for creating the model.	Artificially created frontal, right profile and left profile images (972 x 1280 pixels and eye distance of around 240 pixels) .	Allow verification of the capability of the application to reproduce an unknown face.
OMNRD	A set of 28 newly scanned faces (five males and five females scanned with an Artec MHT 3D scanner Artec Group, Moscow, Russia), therefore out of the faces used for creating the model and without vertex correspondence.	Acquired with a standard digital camera (Nikon Coolpix L23 with 5x Optical Zoom, 2736 x 3648 pixels).	Allow verification of the capability of the application to reproduce an unknown face.

Table 2: Rating system used for the qualitative evaluation of the reconstructed 3D faces.

Value	Meaning	Description
5	Excelent	No apparent differences from the patient can be noticed.
4	Very Good	It is very similar to the patient apart from minor details.
3	Good	It is similar to the patient and good for communication with the patient.
2	Bad	Large differences to the patient can be noticed, but it can still be used for communication of certain procedures.
1	Very Bad	It has no resemblance with the patient and cannot be used for any kind of communication.