Supplementary Information: A statistical shape model for estimating missing soft tissues of the face in a black South African population

Helene Francia Swanepoel MSc [1,*], Harold S. Matthews PhD [2,3,4], Peter Claes PhD [2,3,4,5], Dirk Vandermeulen PhD [3,5], Anna C. Oettlé MD, PhD [1,6]

[1] Department of Anatomy, University of Pretoria, Pretoria, South Africa
[2] Laboratory for Imaging Genetics, Department of Human Genetics, Katholieke Universiteit, Leuven, Belgium
[3] Medical Imaging Research Center, Universitair Ziekenhuis, Leuven, Belgium
[4] Facial Sciences, Murdoch Children's Research Institute, Parkville, Australia
[5] Department of Electrical Engineering, Katholieke Universiteit, Leuven, Belgium
[6] Anatomy and Histology Department, Sefako Makgatho Health Sciences University, Pretoria, South Africa

*Correspondence: Helene Francia Swanepoel, 11 Northfields Avenue, Gwynneville, NSW, 2500, Australia. Email: u27089691@tuks.ac.za

TABLE OF CONTENTS

SUPPLEMENTARY METHODS

Statistical shape modelling

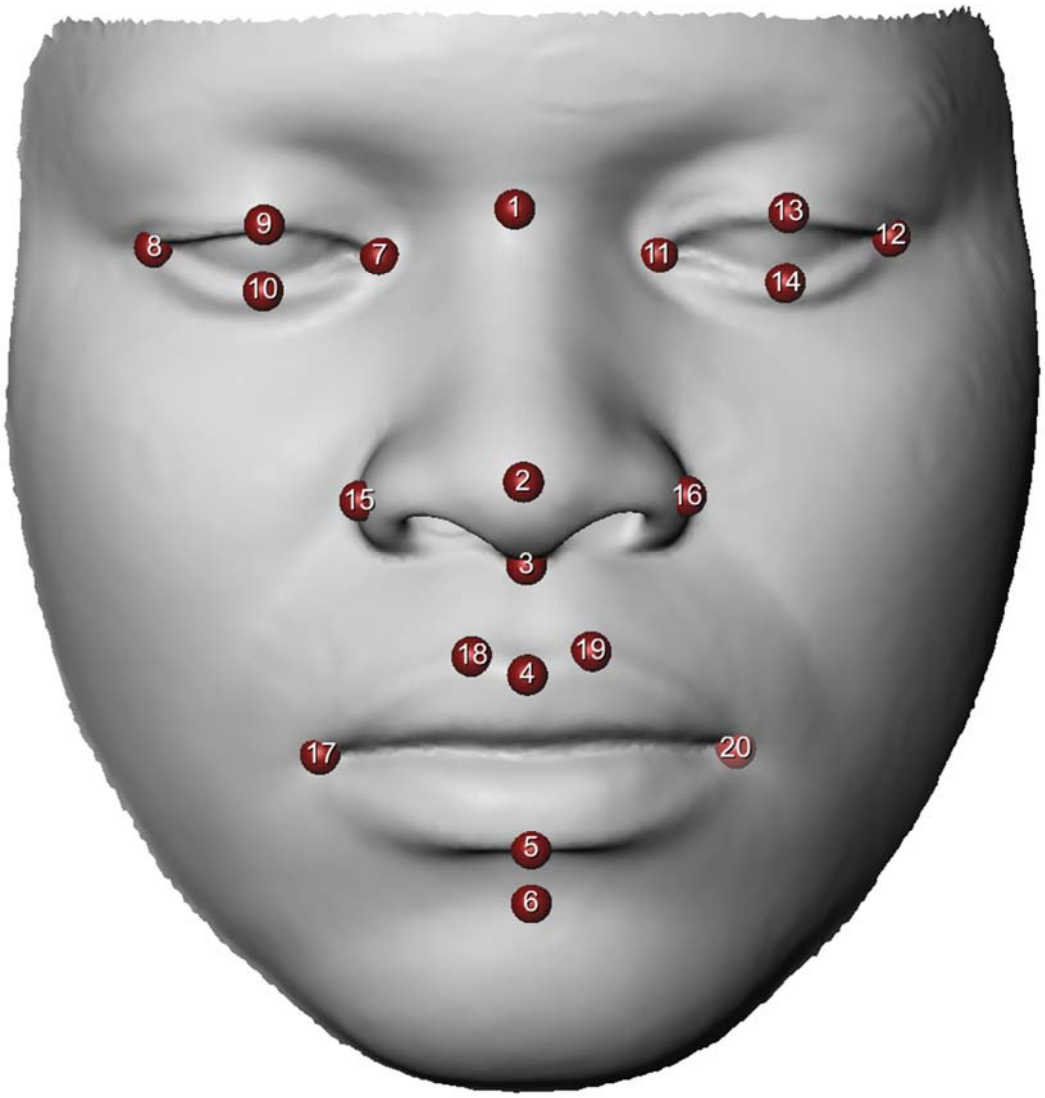
Establishing correspondence among facial meshes.

The CBCT and CT scans were collected as Digital Imaging and Communications in Medicine (DICOM) files. These were first transformed into mesh representations of the facial images by importing the volume into MeVisLab,[1] and applying threshold segmentation to segment the soft tissue surface of the face only. In this way, an iso-surface is generated which is then tessellated to create the triangular mesh and exported as a wavefront object (Supplementary text Fig. 2A). Constructing an SSM and imputing missing parts of a shape from it requires that each face be represented by the same number of vertices and that these vertices should correspond across all instances of the shapes. This can be accomplished via a non-rigid registration of a template face onto each mesh. A standard template face (section 1.1.2) is gradually deformed into the shape of the target, imposing onto the target shape its vertices and topology comprising approximately 20 000 dense quasi-landmarks. In general, we use a non-rigid iterative closest point (ICP) framework,[2] which, over multiple iterations, updates both the estimated corresponding (closest) points on the target to the template as well as the non-rigid deformation from the template to the target. Combined with the gradual relaxation of a regularization parameter on the deformation field, this allows the template to gradually become more flexible in its approach to the target. We employ the non-parametric non-rigid ICP registration implemented in the MeshMonk toolbox,[3] (https://gitlab.kuleuven.be/mirc/meshmonk) in MATLAB version R2021b.[4] The approach is non-parametric in that it incorporates no prior model of allowable deformations. In theory, this allows it to deform to any shape, but another impact is that, if the target is too dissimilar

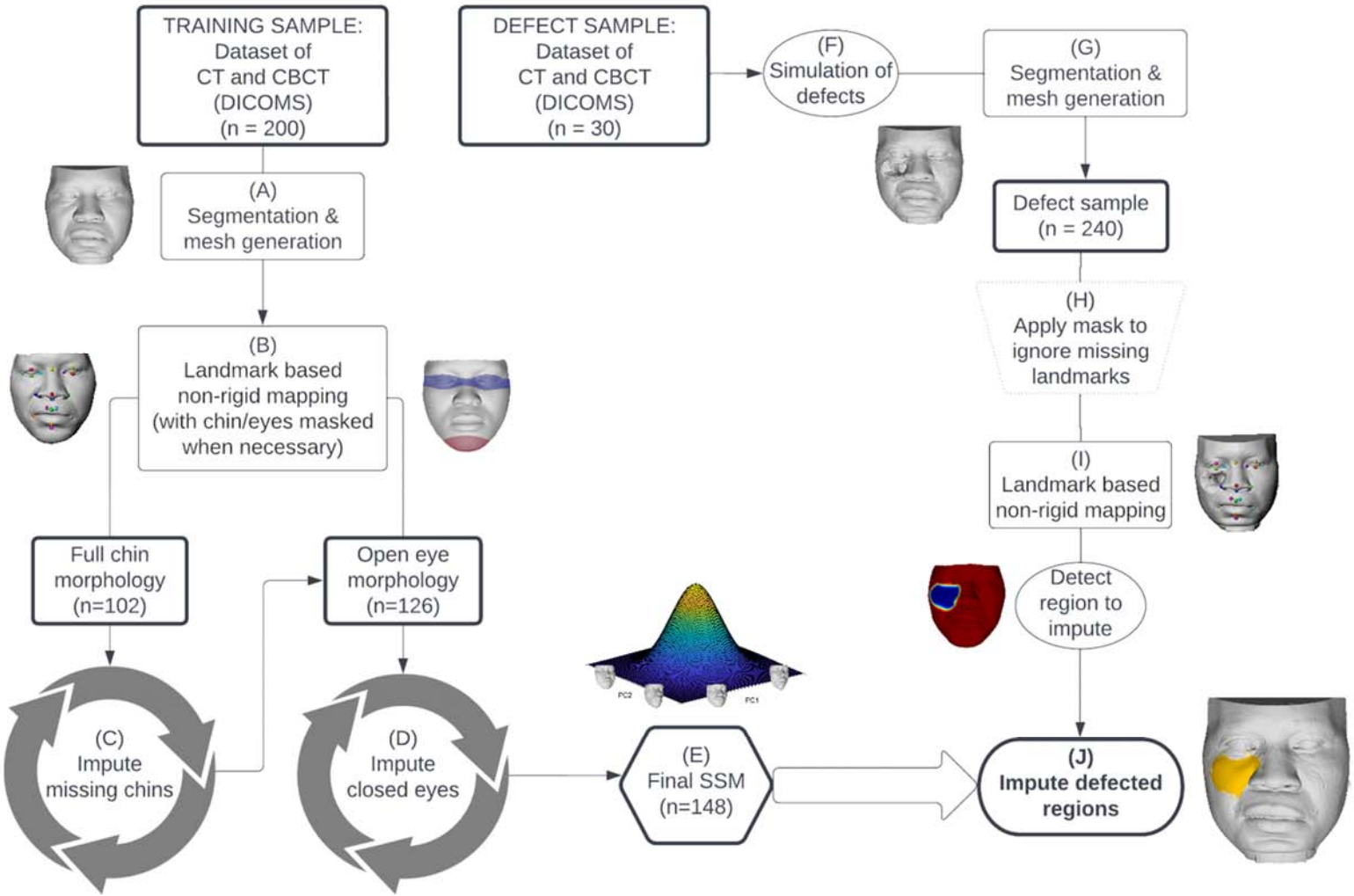
from the template, it can run to anatomically implausible solutions and correspondence is then incorrectly established.

To improve the precision of the final registration, the non-rigid ICP was initialized with a landmark-guided non-rigid deformation. Twenty landmarks (Supplementary text Fig. 1) were carefully placed on each target mesh and the template mesh (Supplementary text Fig. 2B). This initial non-rigid deformation was modelled as three thin-plate spline interpolants each predicting the scalar x, y, or z coordinates of the landmarks on the target, from the x, y and z coordinates of the landmarks on the template. Evaluating the interpolants for all points on the template yielded their deformed coordinates. For registration of the defect scans not all landmarks could be placed so the full set of 20 landmarks was reduced to only those that were not on a defective region.

Following initialization, the non-rigid ICP was performed. The MeshMonk framework allows for user-specified regions of the face to be ignored while computing the deformation towards the target. For those participants where the chin was supported by a chin strut or had closed eyes, these regions were ignored in the calculation of the deformation field. Any participants that still failed the registration were excluded and a total of 148 samples were included in the final model.



Supplementary text figure 1: Anatomical landmarks used for initialization during mapping. 1. Nasion 2. Pronasale 3. Subnasale 4. Labiale superius 5. Labiale inferius 6. Sublabiale 7. Endocanthion (right) 8. Exocanthion (right) 9. Upper lid (right) 10. Lower lid (right) 11. Endocanthion (left) 12. Exocanthion (left) 13. Upper lid (left) 14. Lower lid (left) 15. Alare (right) 16. Alare (left) 17. Cheilion (right) 18. Christa philtre (right) 19. Christa philtre (left) 20. Cheilion (left)


TRAINING SAMPLE:
Dataset of
CT and CBCT
(DICOMS)
(n = 200)
DEFECT SAMPLE:
Dataset of
CT and CBCT
(DICOMS)
(n = 30)
(F)
Simulation of
defects
(G)
Segmentation &
mesh generation
(A)
Segmentation &
mesh generation
Defect sample
(n = 240)
(H)
Apply mask to
ignore missing
landmarks
(B)
Landmark based
non-rigid mapping
(with chin/eyes masked
when necessary)
(I)
Landmark based
non-rigid mapping
Full chin
morphology
(n=102)
Open eye
morphology
(n=126)
Detect
region to
impute
(C)
Impute
missing chins
(D)
Impute
closed eyes
(E)
Final SSM
(n=148)
(J)
Impute defected
regions

Supplementary text figure 2: Methodology overview. The left side of the figure shows the steps involved in generating the statistical shape model used to estimate missing regions of the defective face. The right side of the figure shows the steps involved in creating a sample of simulated defects, mapping the defective sample, and using the SSM to estimate the defective regions. (A) A training set of DICOMS was segmented and the resulting iso-surface of the soft tissue face was tessellated into meshes for further processing. (B) To achieve correspondence, a set of 20 landmarks was placed on the meshes and a standard facial template. Non-rigid mapping was conducted in the MeshMonk toolbox to ensure a standardized topology across all meshes. (C) and (D) An iterative bootstrapping approach was followed to estimate chins and eyes for samples without complete chin morphology or with closed eyes and sequentially added to the SSM. (E) The final SSM was generated by calculating the modes of variation (through Principal Component Analysis) for the total sample including open-eye and chin estimations. (F) Six classes of facial defects were simulated using the Avizo software. (G) Defect faces were segmented and tessellated into meshes for further processing. (H) Missing landmarks due to defect regions were masked from the subsequent mapping during (I) by assigning 'target flags'. (J) The defect regions were estimated using the SSM by the same process used to estimate the chins and eyes in (C) and (D) and visualized in the context of the original unprocessed defective mesh.

Construction of the standard template

One of the sample meshes was selected that showed clear and crisp details and had open eyes and full chin morphology. The mesh was trimmed to shape and resampled to ensure an even distribution of vertices using the isotropic explicit re-meshing tool in Meshlab.[5] Each mesh in the training sample was brought into correspondence with the resampled template mesh using the same method described in section 1.1.1 above. Once correspondence was achieved, the average face of an SSM was generated as described in section 1.1.3 below. Three iterations of mapping for each of the faces in the training sample with complete chin morphology and open eyes (n = 41) were completed using the average face as template, with the template being updated continuously. The final average face was used as the standard template for all further steps. Figure 3 shows the final standard template used in this study.

Supplementary text figure 3: Standard facial template obtained by using a bootstrapping approach and used for correspondence throughout the study.

Model building

The mixed nature of the data (i.e., some with complete chin morphology, some with chin struts, some with open eyes and some with closed eyes) presented a particular challenge.

This was addressed by following a bootstrapping approach in which those with incomplete morphology were first estimated and added to the model sequentially. Given a sample of shapes represented by the same landmarks, the SSM was built as follows. Firstly, only samples with complete chin morphology were included to generate the first version of the SSM. Non-shape-related variation is removed by generalized Procrustes analysis (GPA) which aligns all landmark configurations onto the sample mean and scales all configurations to unit size. Over multiple iterations, each face is aligned to the mean configuration via a rigid scaled least-squares Procrustes alignment. Following this, the mean configuration is re-calculated. To minimize the influence of the chin region for those scans including the chin support we used a weighted least-squares Procrustes superimposition with weights of zero assigned to the chin region and ones assigned to all other points. Similarly, the mean configuration is the weighted mean configuration, employing the same weights. The first version of the model was used to estimate (see next paragraph below) the chins of samples where these features were missing or unusable (Supplementary text Fig. 2C). A second version of the SSM was then generated including all the complete and estimated chins, as well as all samples with open eyes (Supplementary text Fig. 2D). We again used a weighted least-squares Procrustes superimposition, this time with weights of zero assigned to the eye region and ones assigned to all other points. This version of the SSM was used to estimate the eyes of the samples with closed eyes. Finally, the SSM was updated to include all samples (Supplementary text Fig. 2E). Each SSM was created by principal components analysis (PCA) of the GPA-aligned landmarks. An SSM comprises modes of variation or PCs, which each correspond to a linear transformation of facial shape, as well as the normal range variance along each mode together defining multinormal parameterization of shape variation. The modes of variation

were calculated (through Principal Component Analysis) by a singular value decomposition of the n (observations) by 3k(k landmarks) matrix.

To estimate user-identified missing parts of the face, the vertices to be estimated are assigned weights of zeros and those to remain unchanged are assigned a value of one. To avoid discontinuities in the final result the weights are smoothed with 10 iterations of Laplacian smoothing. Alignment and scaling to the model average, followed by a weighted fit to the SSM is accomplished as described by Matthews et al.[6] to estimate the linear combination of the modes of variation that most closely approximates the face in a weighted least-squares sense. In essence, the linear combination is such that it aims to correctly reconstruct the parts of the face with high weighting while ignoring the regions of low weighting, as a result, the regions of low weighting are simply filled in with the most likely shape given the regions of high weighting. The estimated 'weighted fit face' can then be reconstructed by evaluating the linear combination. The weighted fit face is then returned to the coordinate system and the size of the face prior to the estimation. The final estimated face is created by blending the face before estimation with the weighted fit face as the weighted sum of the two landmark configurations, with the original face weighted according to the weights described above and the weighted fit face weighted according to 1 minus the weights.

Model evaluation

We evaluated the model's representation of the target population by calculating its generalization and specificity. In principle, both of these properties contribute to the ability of the model to realistically estimate missing parts of the face. All these computations require the calculation of inter-shape distances. To avoid the chin and eye regions influencing these

calculations where it is not informative, weights of zero were assigned to these regions. The inter-shape distance was then calculated as the weighted root mean squared (RMS) distance between the two shapes. Generalization represents how well the model can represent realistic faces not used in training and can also be interpreted as the mean of the average reconstruction error between the model and an unseen shape,[7] or the out-of-sample reconstruction error of the training data. The reconstruction error is calculated as the difference between the face and the reconstruction of that face from their projections onto the modes of variation. It was calculated by sequentially holding each face out of the training of the SSM and then estimating their 'weighted fit face' from the model and calculating the inter-shape distance between the two. The average of all these inter-shape distances is the model generalization. In-sample accuracy is computed identically to the generalization, except that the face is not held out from training the model. The in-sample accuracy for a given number of modes constitutes the lower bound of what is the possible generalization error and calibrates the interpretation of the generalization.

Model specificity concerns the ability of the SSM to represent only realistic or valid faces. This is calculated by randomly simulating 1000 faces by randomly sampling linear combination coefficients from their multivariate Gaussian distribution and reconstructing the corresponding faces. The verisimilitude of each simulated face is calculated as the inter-shape distance to the most similarly shaped face from the training sample. Both generalization and specificity were calculated using SSMs trimmed to only the modes of variation that explained up to 96% of the variation in the sample. To express model specificity, generalization and in-sample accuracy in mm units, before calculating the inter-shape distance, the simulated shapes and weighted fit shapes were scaled to the size of the face to which they were being

compared. This is weighted to ignore estimated areas such as the chin and eyes where applicable.

Whether the model could be improved by collecting further data was assessed by evaluating how generalization evolves when gradually more and more data are added up to the available amount. Specifically for 10 repetitions, 15 samples were randomly chosen to train an SSM and iteratively 15 subjects are randomly selected and added to the model training data up to the total sample size. At each iteration, the distribution of the average error values over the 10 repetitions was plotted as boxplots.

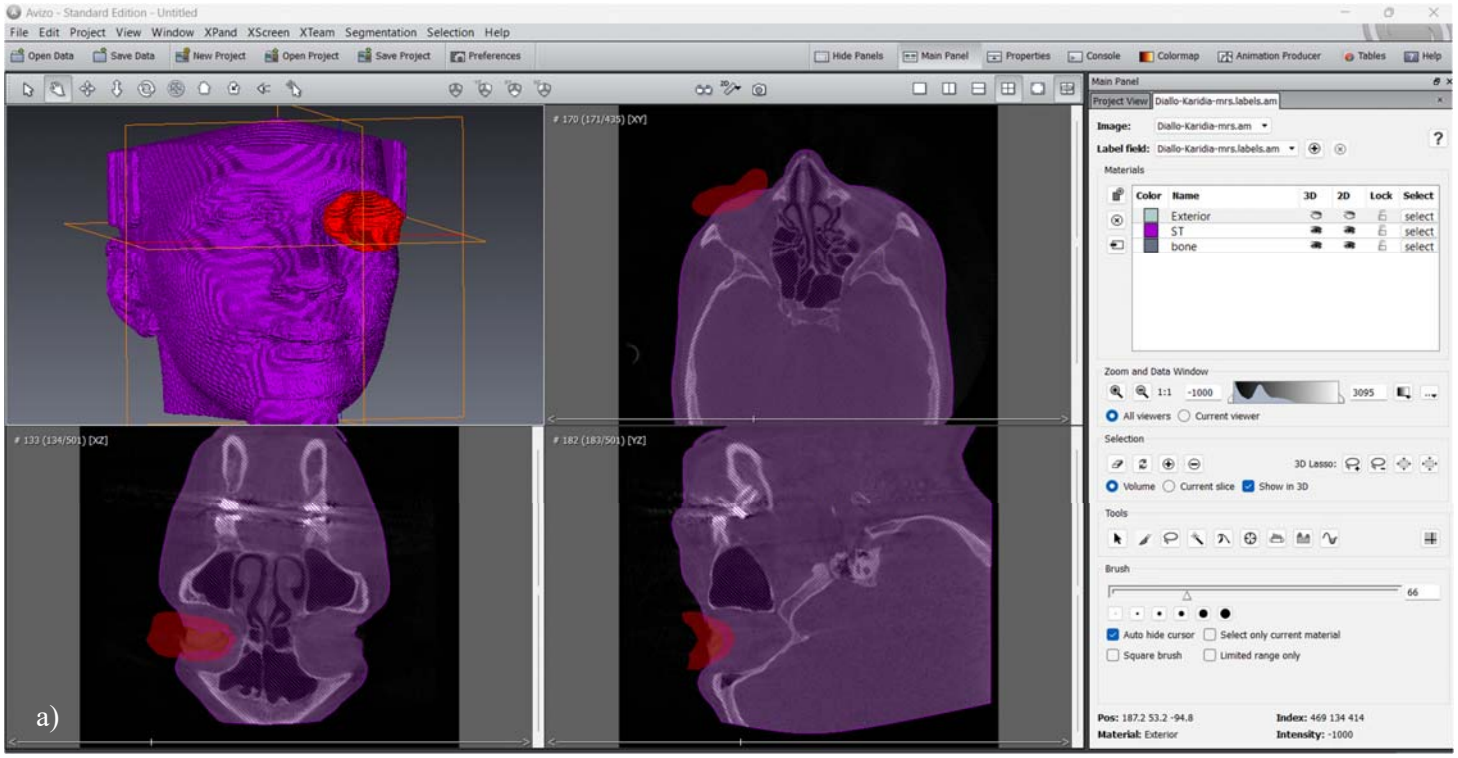
Simulation of facial defects

Facial defects included only extra-oral defects. These were classed into six groups and simulated on a sample of 30 open-eyed faces using the Avizo® v. 8.0.0 software.[8] A total of 240 defect instances were simulated (30 faces x 8 simulations per face). The classes and their descriptions are represented in *Supplementary Table 1.*

Sample volumes were imported into the Avizo and visualized in three planes. Essentially, the voxels corresponding to each defect must be deselected (Supplementary text Fig. 2F). Firstly, the entire head was selected by thresholding and exported as a DICOM stack. The labels were then edited manually in Avizo to deselect the defective regions to be removed (Supplementary text Fig. 4 a) and each set of labels, corresponding to each defect class 1-5 was again exported as a DICOM stack. For class 1-5 defects, the labelled DICOM stacks were each individually imported into MeVisLab, an iso-surface was generated, tessellated into a mesh and exported as a wavefront object (Supplementary text Fig. 2G) (Supplementary text Fig. 4 b). For class 6 combination defects, we imported labels for the entire head and for all

class 1-5 defects into MeVisLab at the same time. Defects were combined for an individual by taking the intersection of labels from a selection of class 1-5 defects and tessellating the resulting isosurface of the intersection. Three selections were done per individual. The first selection simulated bi-orbital defects, where both the right and left orbits were involved, the second selection (composite 1) simulated a large defect involving three facial features (e.g., orbital, cheek, and lips etc.), while the third selection (composite 2) simulated smaller defects only involving two features (e.g., orbital & full nasal, or partial nasal and cheek etc.).

To establish correspondence for the meshes in the defect sample a modification of the process as described in section 1.1.1 above was used. Landmarks that could not be placed on the defect scan were not used to estimate the landmark-guided non-rigid initialization (Supplementary text Fig. 2I). Points corresponding to the defect region of the scan (section 1.3) were ignored by assigning 'target flags' to MeshMonk's ShapeMapper (Supplementary text Fig. 2H).

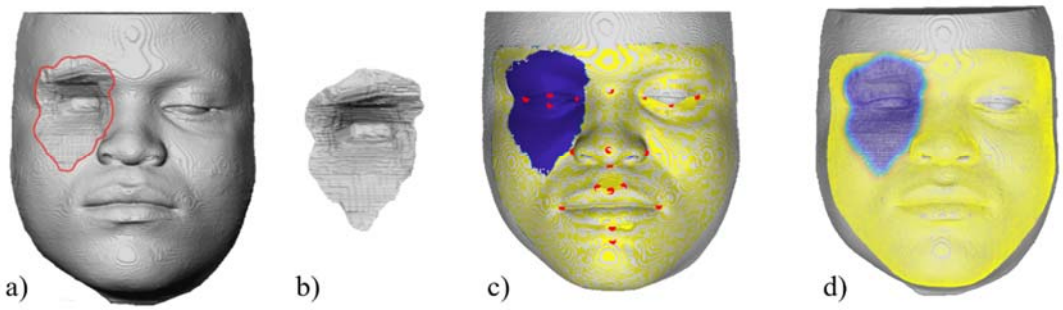


Supplementary text figure 4: Simulating the defective region on Avizo 8.0.0. a) A label field was created from which voxels corresponding to a defect were deselected and the remainder exported as a DICOM stack, which was then b) imported into MeVisLab to visualize the isosurface of the defective face.

Estimation of facial defects

The same process to estimate defects was followed as described above for imputing the chins and eyes used in the SSMs, the only difference being that the weights were applied to the defective region, instead of the chin or eye regions. Supplementary text Fig. 5 shows how the defective regions were identified by isolating the vertices corresponding to the defect on the unmapped isosurface of the defective face using MeVisLab. The defective regions were then transferred onto the registered version of the face by flagging the closest points on the mapped mesh that corresponds to the defective region. The final SSM was used to estimate the linear combination of the modes of variation that most closely approximates the intact regions of each face and estimated the missing regions using a weighted projection onto the modes of variation (section 1.1.3). The face reconstructed from these projections was blended with the mapped version of the defective face by smoothing the flagged points to blend in with the surrounding areas. This result is the mapped version of the defective face with vertices of the defective region substituted with those of the weighted fit face. The estimated region was then visualized and interpreted in the context of the original unprocessed defective mesh (Supplementary text Fig. 2J).

In the future, the manual selection of points to be estimated can be executed via several methods, for example, using an interactive brush tool to select points (https://github.com/harrymatthews50/MeshEditor).



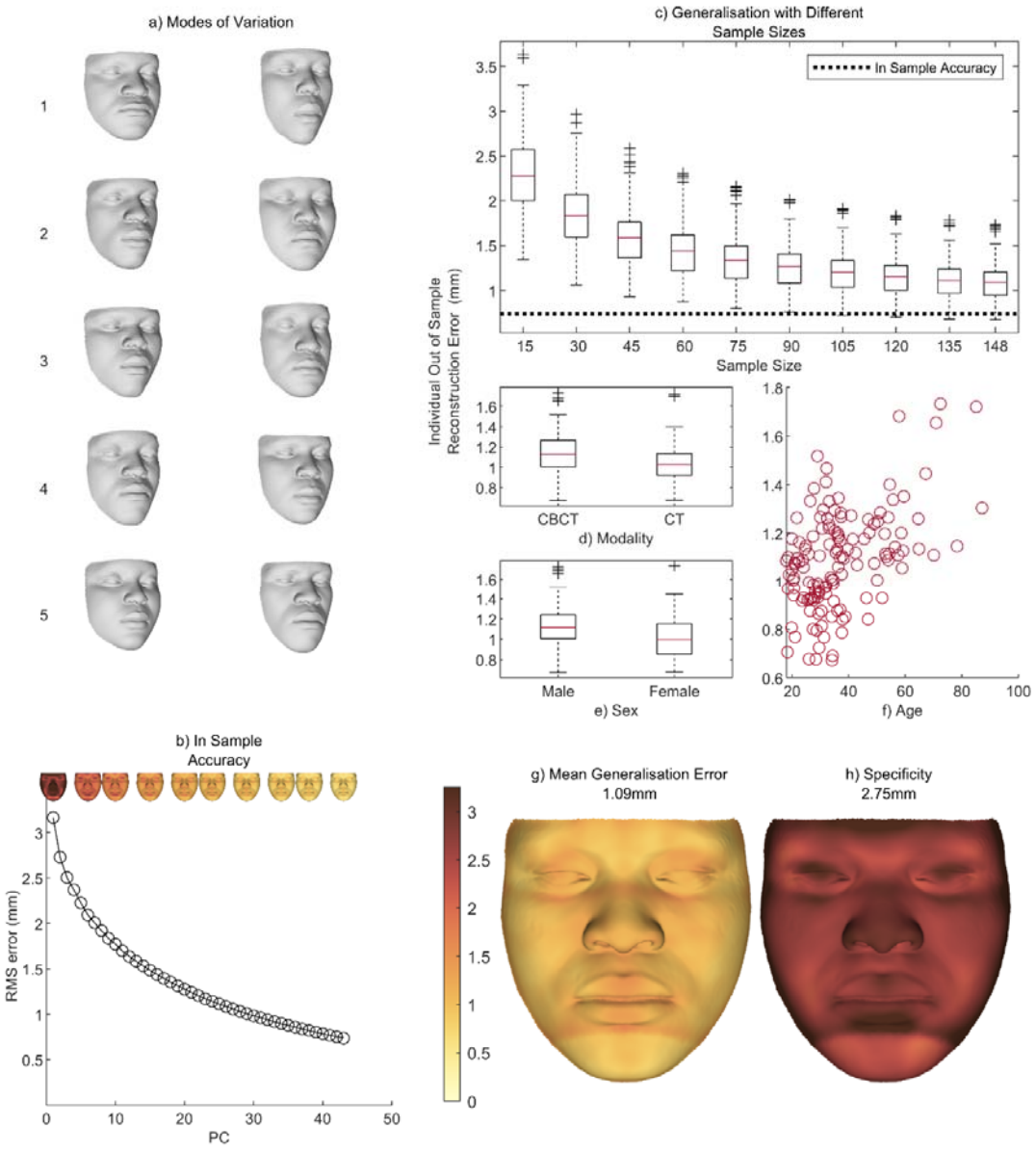
Supplementary text figure 5: a) Identifying the defective region. b) The vertices corresponding to the defective region were isolated by removing the rest of the isosurface (areas not part of the defective region) in MeVisLab. c) The closest points corresponding to the defective region are transferred onto the mapped version of the face and weighted 0 during the estimation (dark blue). d) Selected points are smoothed to blend in with the surrounding area.

SUPPLEMENTARY RESULTS

Model evaluation

Forty-three modes of variation were required to model 96% of the shape variation and were retained. The first 5 modes of variation are shown in Supplementary text Fig. 6 a. Video representations of the first 5 modes of variation can be viewed in the Supplementary Materials (supplementary video files 1 to 5). The first mode of variation, responsible for the most variation, predominantly represents total facial height and width, with changes in the length of the maxillary alveolar processes, mode 2 relates to midfacial projection with changes in the zygomatic width, mode 3 represents primarily variability in upper facial height and nose height. In mode 4, the depth of the eyes is influenced by the zygomatic width and frontal area bossing and mode 5 relates to maxillary and mandibular protrusion.

Supplementary text Fig. 6 c displays the distributions of out-of-sample reconstruction errors corresponding to various training sample sizes. As the model generalization error decreases with increasing sample size, the model better represents the population. As the curve is reaching a plateau, the errors as a function of the sample size are decreasing very slowly. This indicates that beyond the current sample size, a large number of additional participants would result in only an incremental improvement to the model. No difference in out-of-sample reconstruction error was observed between imaging modality (Supplementary text Fig. 6 d) or sex (Supplementary text Fig. 6 e) and age (Supplementary text Fig. 6 f). In Supplementary text Fig. 6 g and h, the weighted RMS per point is also shown as a color map for mean generalization and specificity. The mean generalization error was 1.09 mm (Supplementary text Fig. 6 g). Specificity, the model's ability to represent only valid or realistic faces, resulted in a mean error of 2.75 mm (Supplementary text Fig. 6 h). Regions of slightly higher errors include the eyelids, nose bridge, and lips (Supplementary text Fig. 6 g and h).

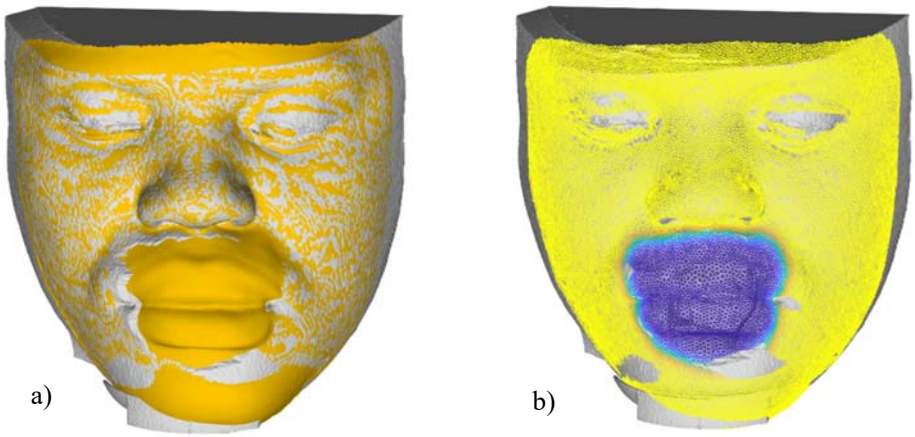


Supplementary text figure 6: Statistical shape model evaluation. a) The first 5 modes of variation are shown. b) The In-sample accuracy in terms of the RMSE shows the lower bound of the possible generalization error and calibrates the interpretation of the generalization for a given number of modes. c) The individual Out of Sample error for different sample sizes (generalization) indicates that beyond the current sample size, a large number of additional participants would result in only an incremental improvement to the model. d) e) and f) indicate the Out of Sample error for modality, sex

and age and showed no differences in the reconstruction error. g) A color map of the mean generalization error shows an acceptable error of 1.09 mm and h) a color map of the specificity of the model indicates a mean error of 2.75 mm.

Qualitative defect evaluation

In approximately 38% of cases, visual inspection of the defect estimations showed that the defect was not smoothly blended with the surrounding tissue. Supplementary text Fig. 7 b shows the weightings used in one case for blending the weighted-fit face to the defective face. Deep blue regions were completely estimated whereas yellow regions were not, and are expected to match the target face perfectly. For areas colored in-between yellow and deep blue, the shape is a weighted combination of the face estimated from the SSM and the target face and as such may not match the target face perfectly. Essentially the mismatch between the two surfaces is because the region selected for estimation extends beyond the true defective region. The scope of the estimated region is determined by both the manual selection of points to define initial binary weights and the number of smoothing passes applied to these initial weights. This smoothing is necessary for even blending, but it also blurs and effectively extends, the boundary of the selected region (section 1.3).



Supplementary text figure 7: Example of a poor defect estimation of the lips. a) The corners of the mouth and parts of the chin (in grey) are cut off due to the flagged defective region seen in blue in b) being too small.

SUPPLEMENTARY TABLES

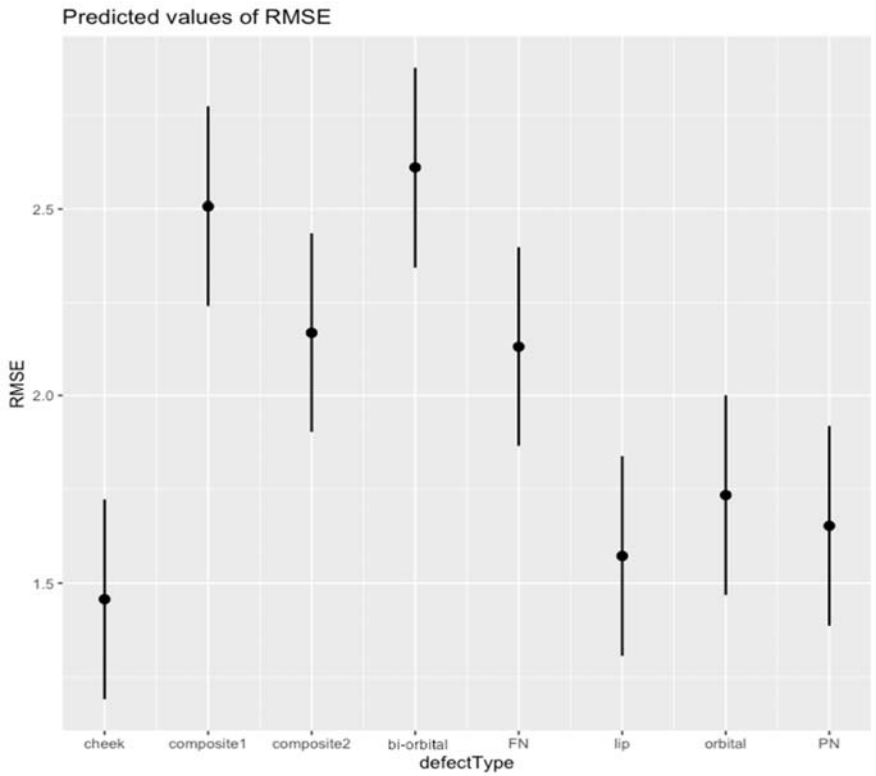
Supplementary Table 1: *Classification of facial defects according to facial feature involvement*

Class	Description
1. Orbital	Defects of the eye and orbit: including the globe, extraocular muscles, eyelashes, and at least part of the eyelids, orbital fat, and periorbita as seen in orbital exenteration
2. Cheek and upper lip	Defects of the soft tissue and bony tissue, or soft tissue only, of maxillary and zygomatic regions, including the upper lip. May extend into the maxillary sinuses.
3. Lips or isolated lower lip	Defects of the soft tissue and bony tissue, or soft tissue only, of mental and mandibular regions, involving the lower lip alone, or both upper and lower lips
4. Full nasal	Defects of the entire soft tissue (and cartilaginous) nose
5. Partial nasal	Defects involving only parts of the soft tissue (and cartilaginous) nose e.g., ala and columella only
6. Combined	Defects involving more than one facial feature in 3 assorted combinations (bi-orbital, large and small), e.g., both eyes; orbital, cheek, and full nose; partial nose and lips etc.

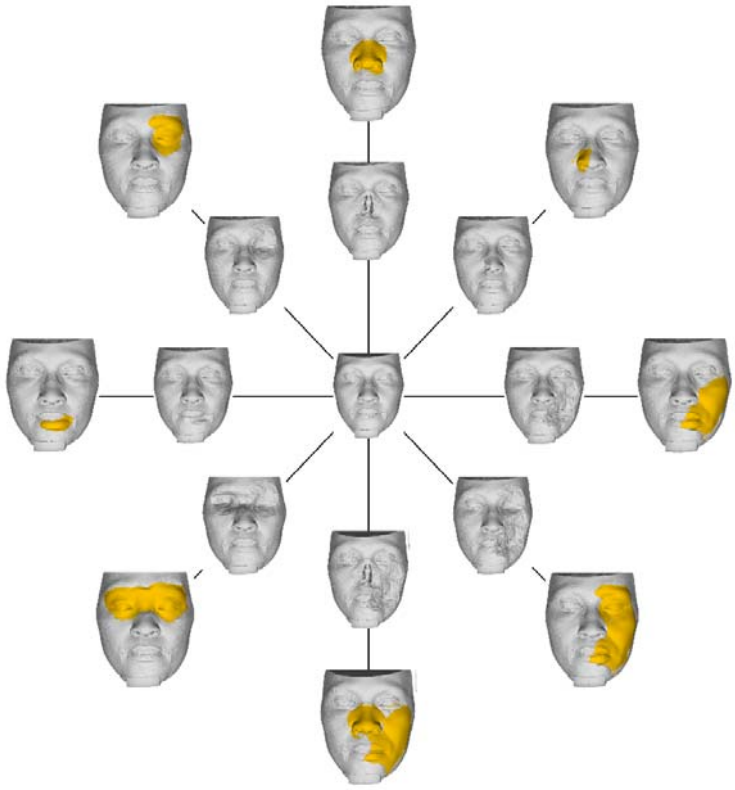
Supplementary Table 2: ANOVA for the effects of defect type, imaging modality, sex and age on the RMSE

	Sumsq	Meansq	NumDF	DenDF	Statistic	p-value
Defect type	40.194	5.742	7	203	27.030	<0.001
Modality	0.433	0.433	1	26	2.036	0.165
Sex	0.200	0.200	1	26	0.942	0.341
Age	3.601	3.601	1	26	16.950	<0.001

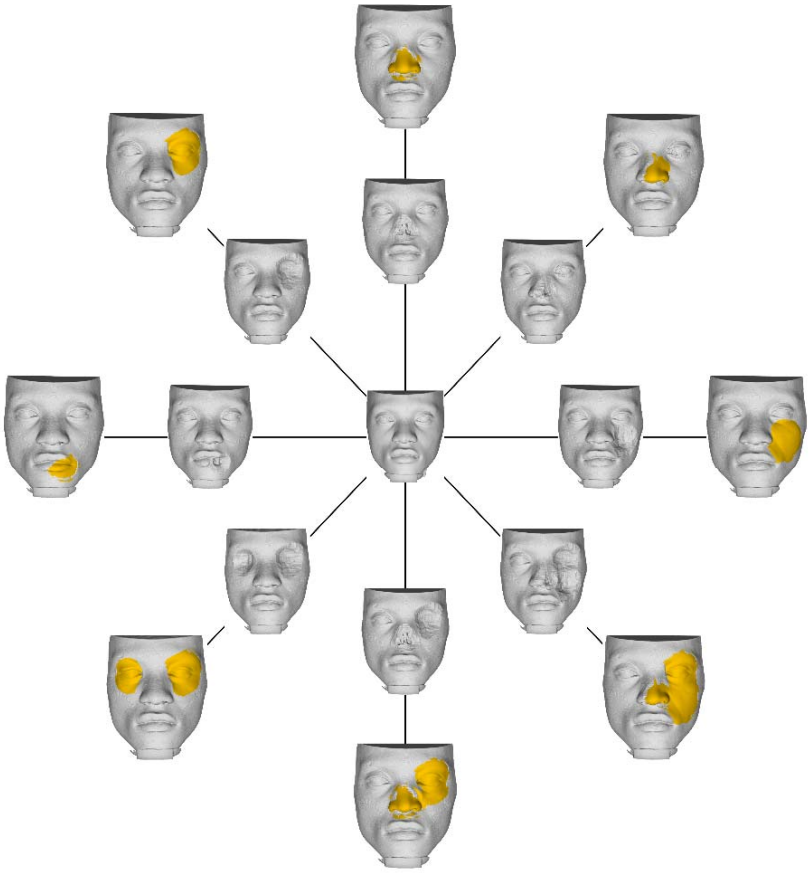
SUPPLEMENTARY FIGURES



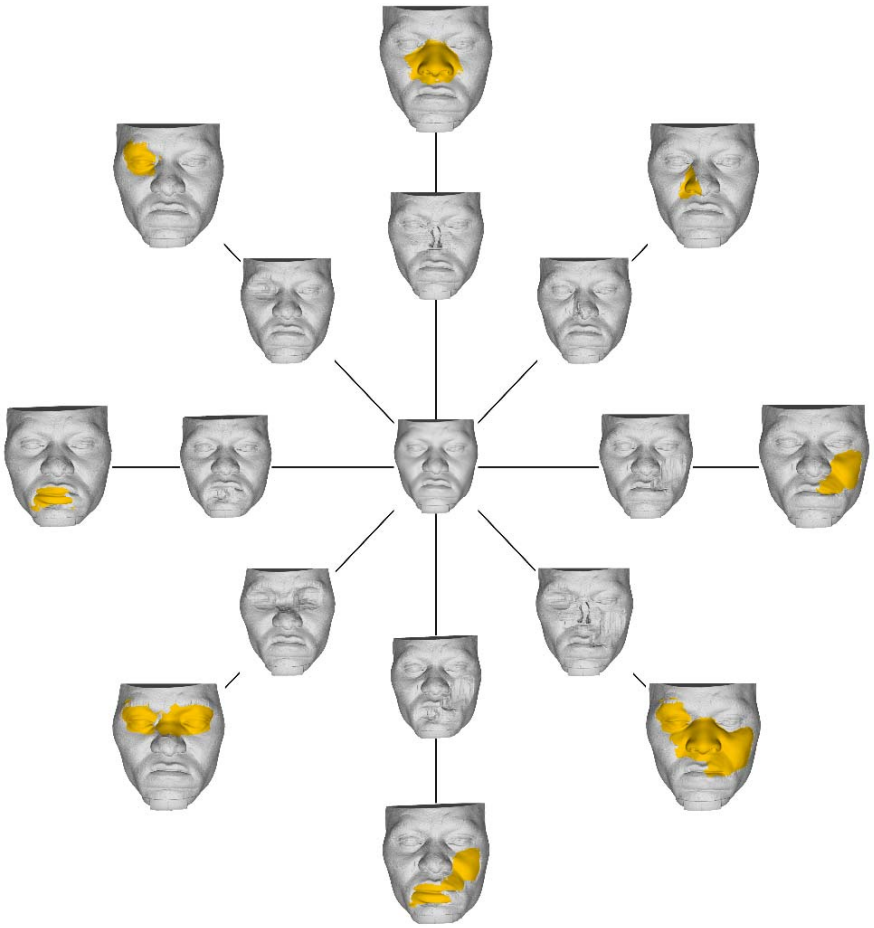
Supplementary figure 1: The expected RMSE values and 95% confidence intervals of the expectation for each defect type. As can be seen, these can be grouped into two distinct groups: composite, bi-orbital and full nose defects; and individual feature defects (cheek, lip, orbital and partial nose defects).



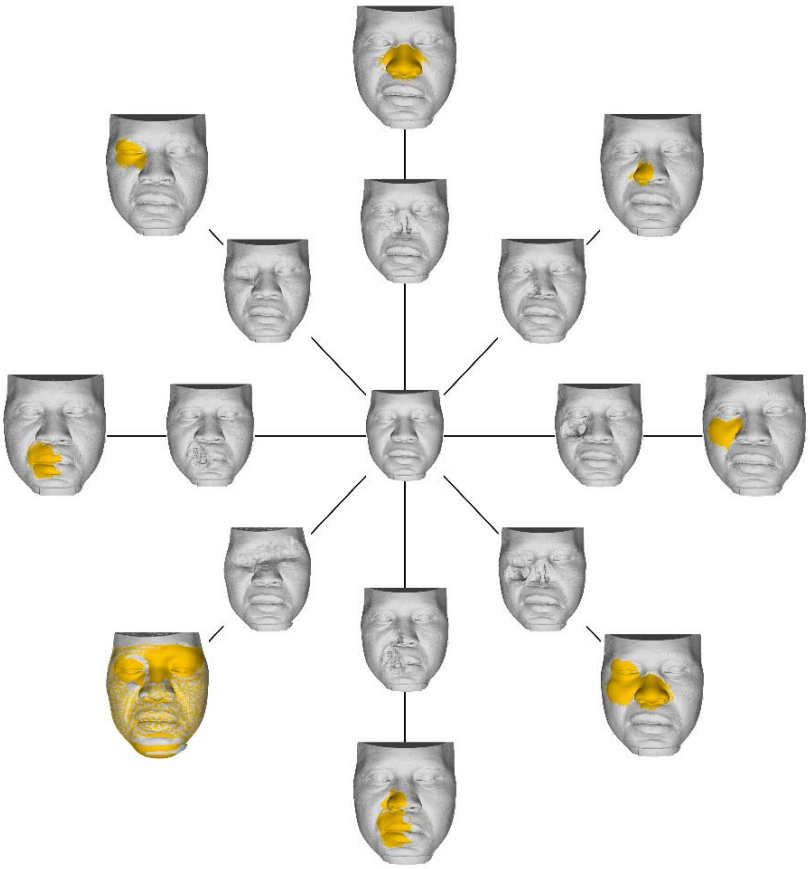
Supplementary figure 2



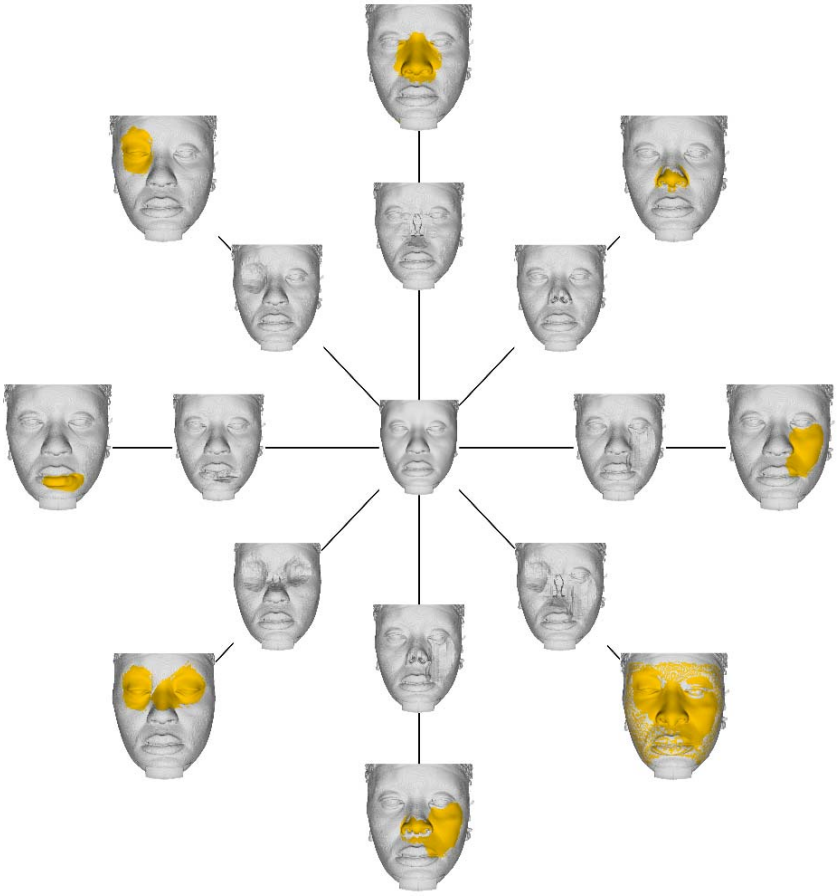
Supplementary figure 3



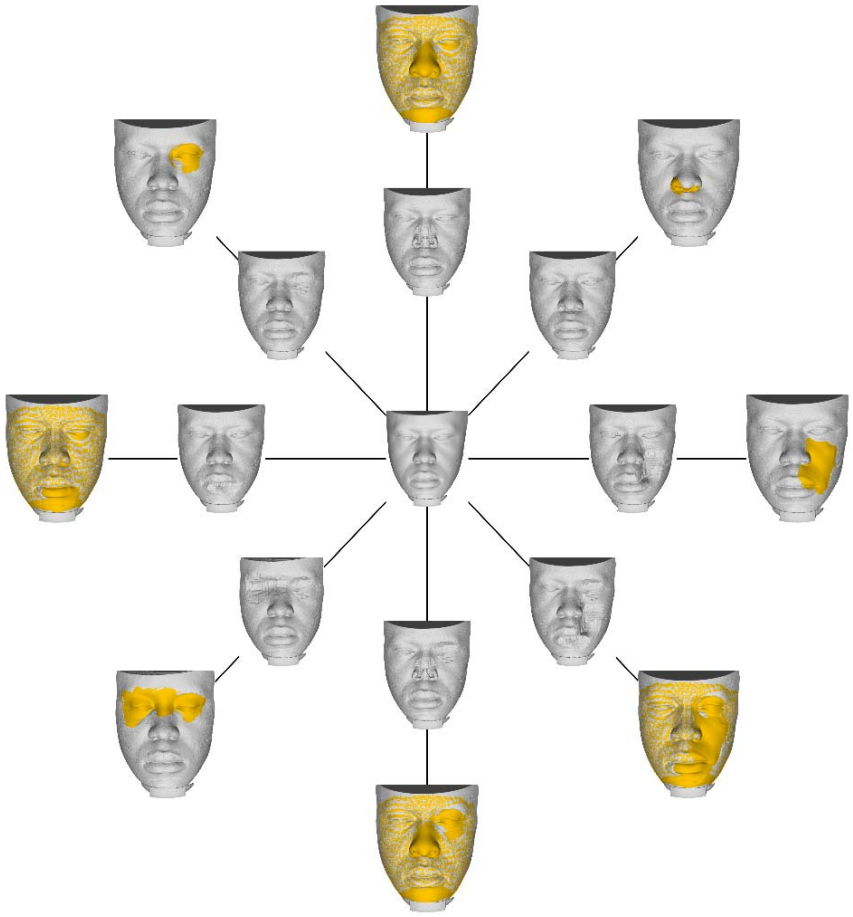
Supplementary figure 4



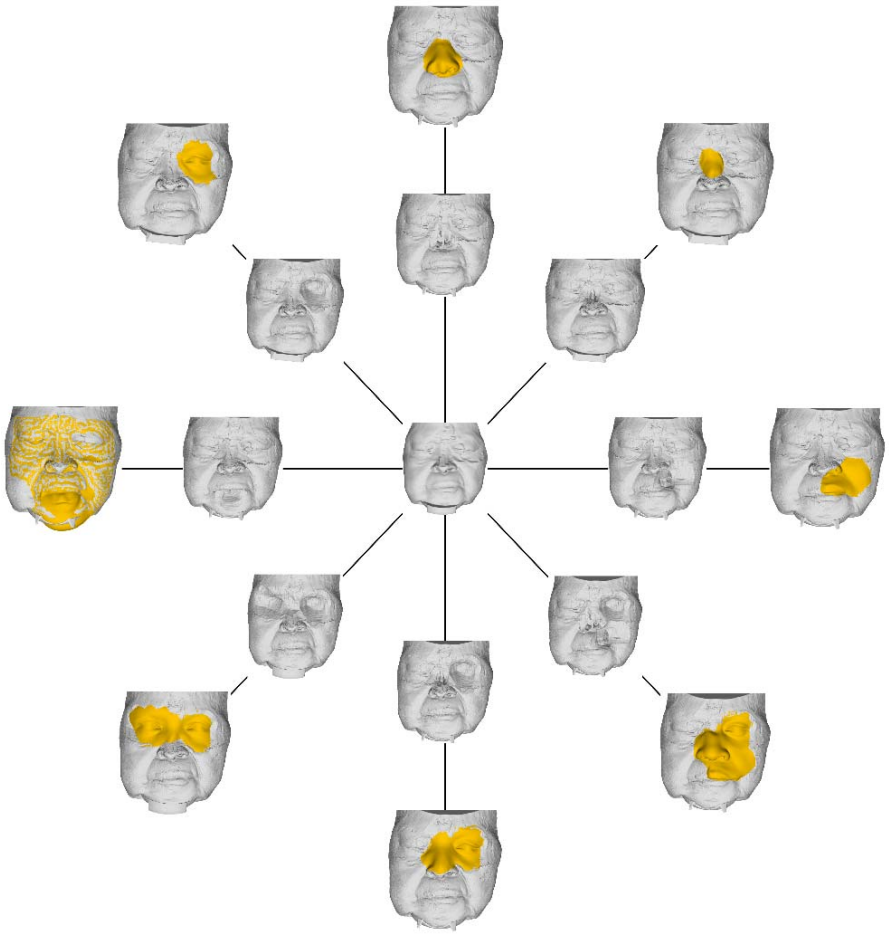
Supplementary figure 5



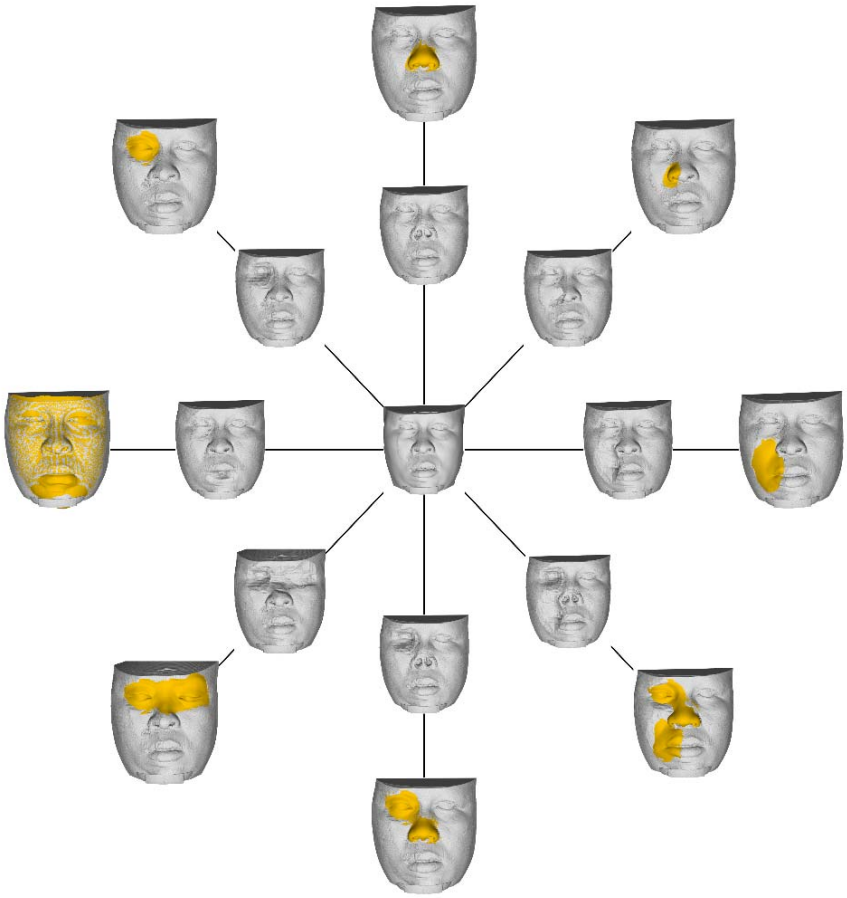
Supplementary figure 6



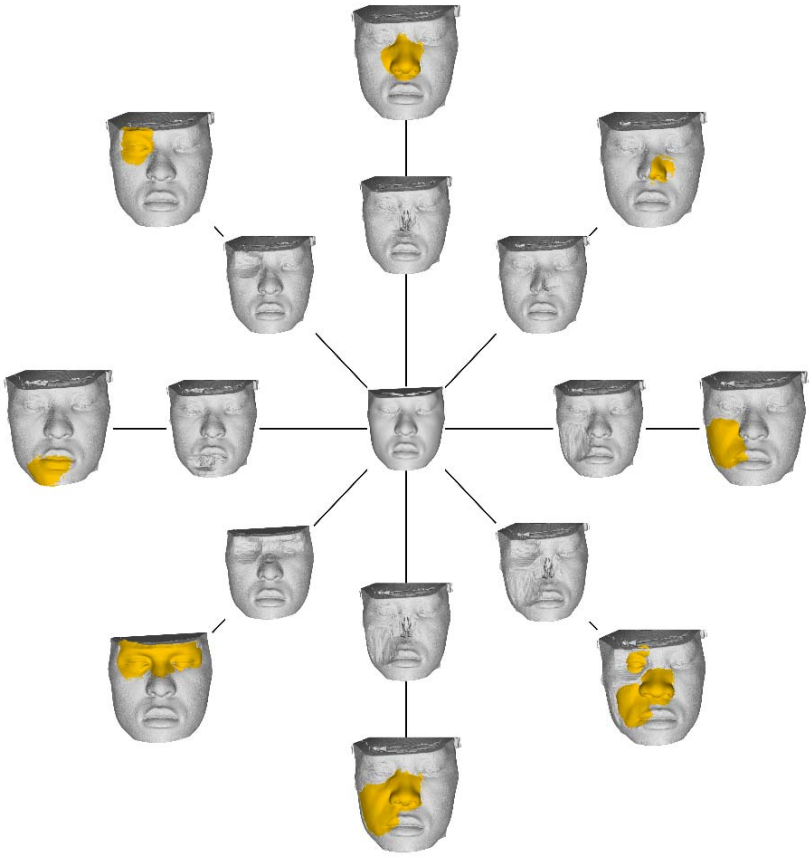
Supplementary figure 7



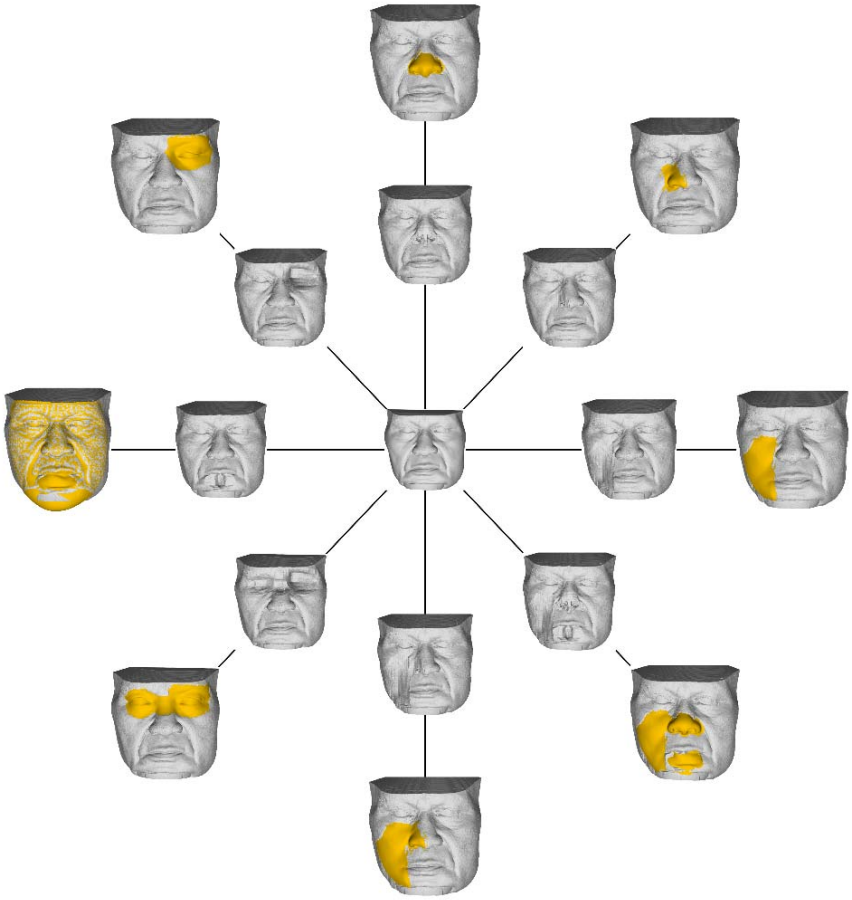
Supplementary figure 8



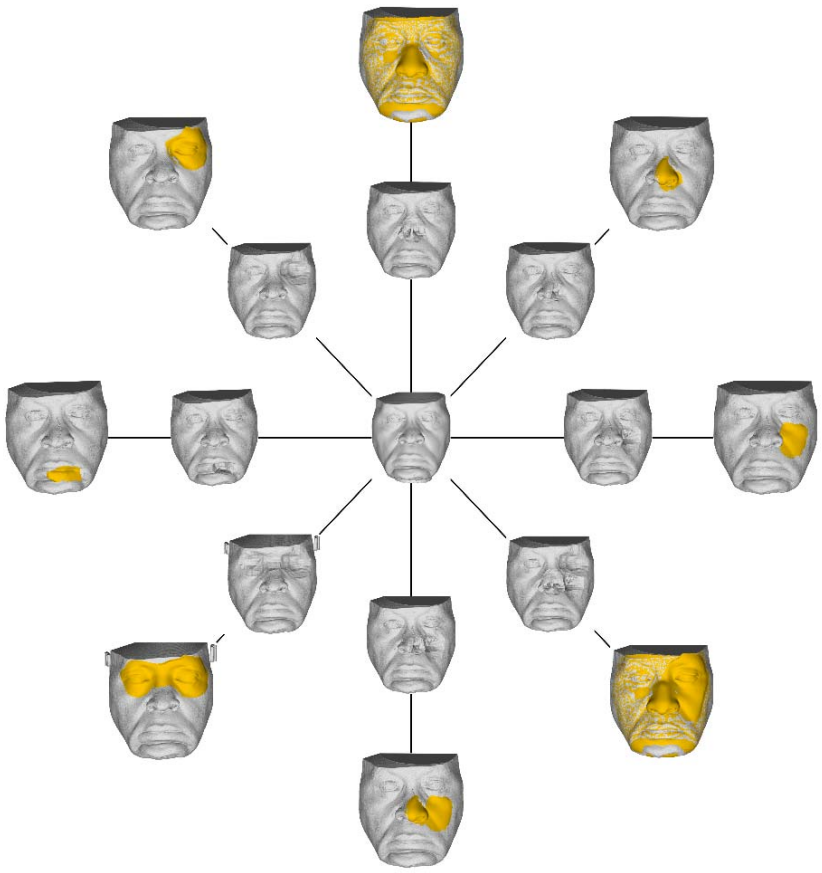
Supplementary figure 9



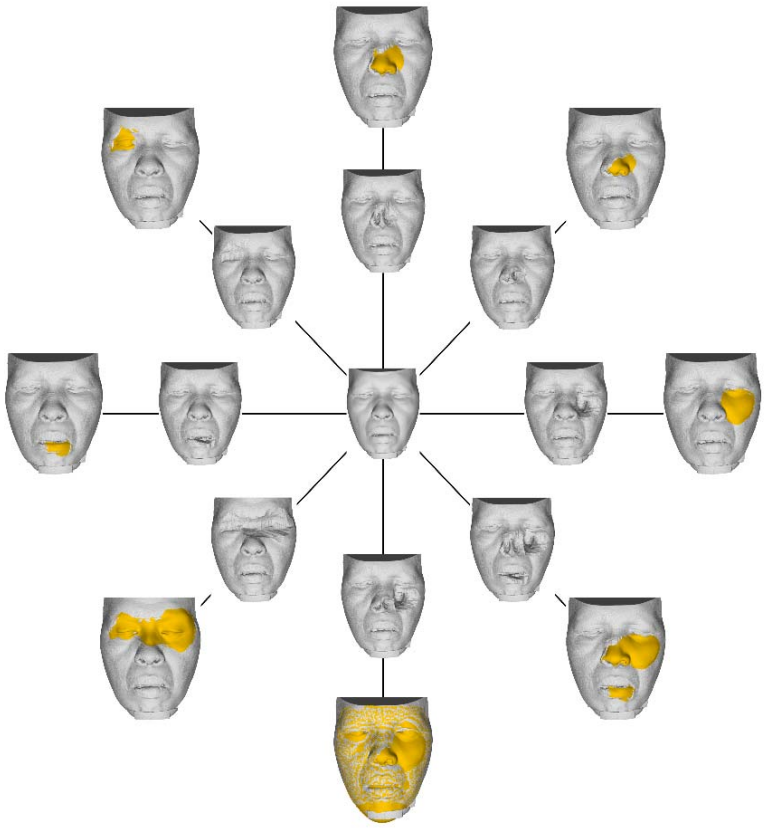
Supplementary figure 10



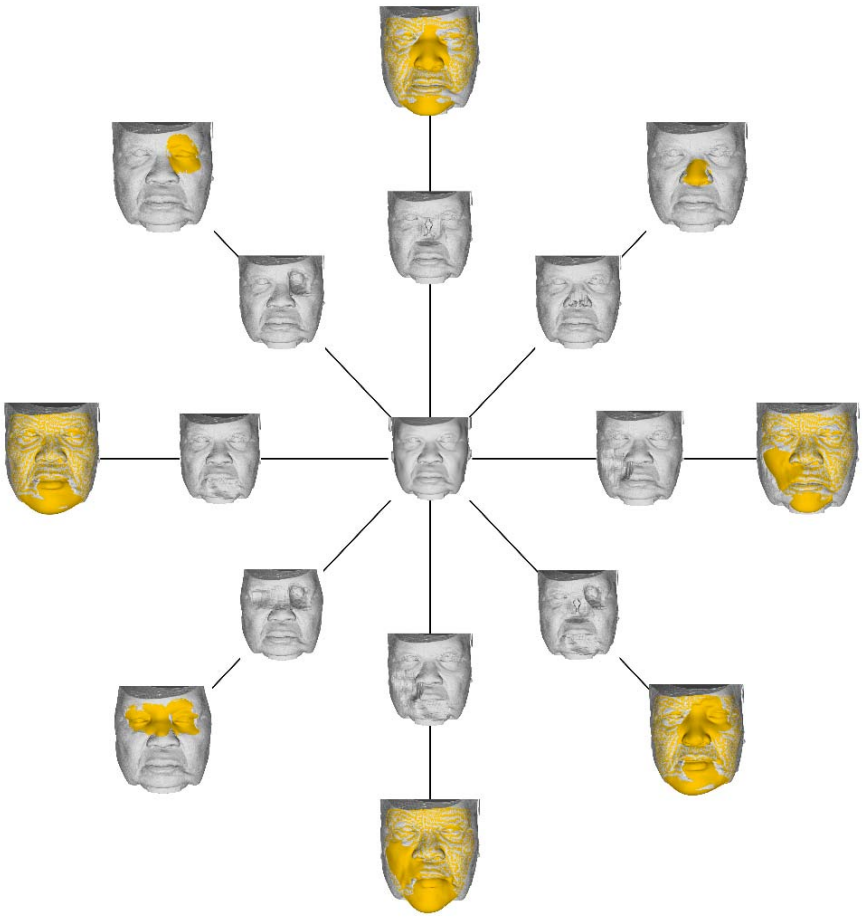
Supplementary figure 11



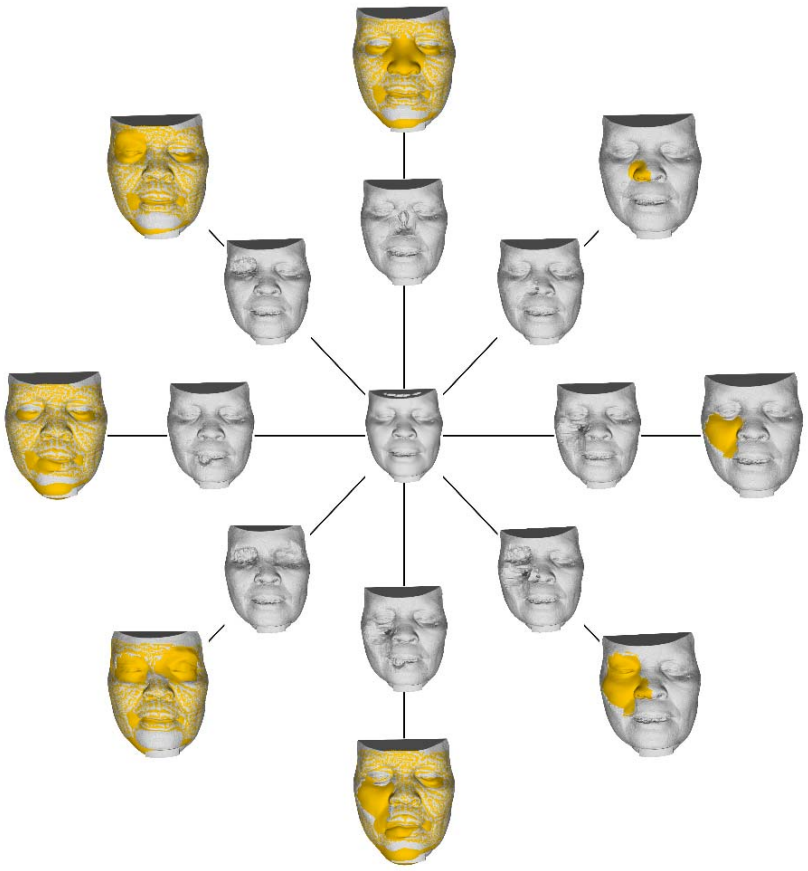
Supplementary figure 12



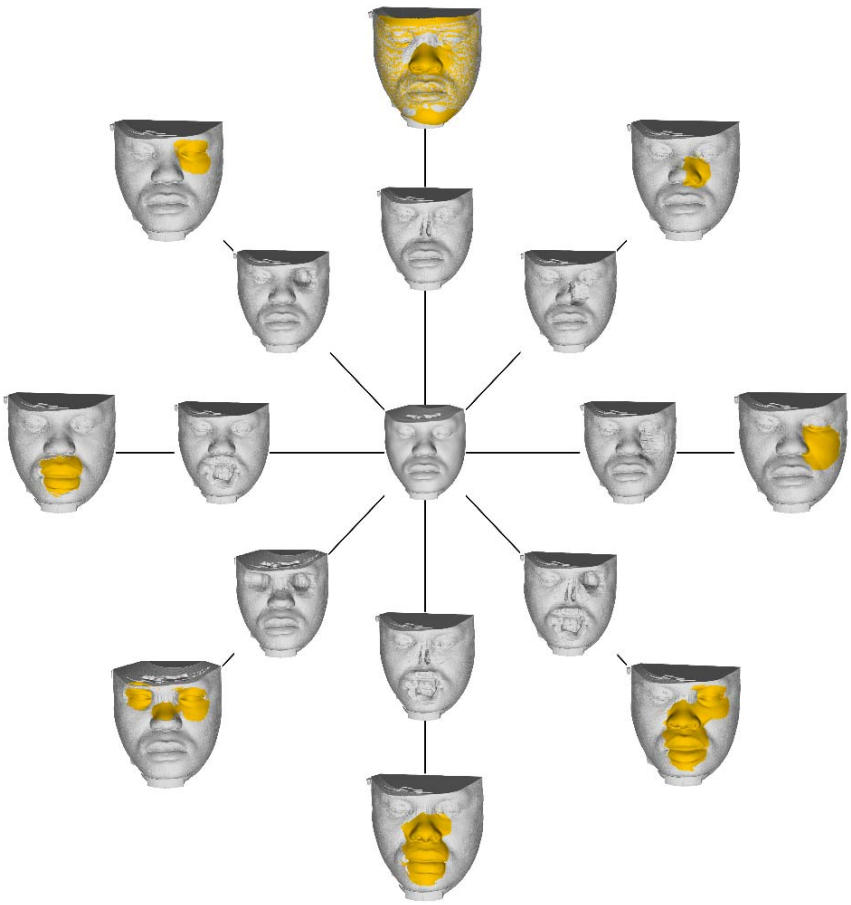
Supplementary figure 13



Supplementary figure 14



Supplementary figure 15



Supplementary figure 16

References

1. MeVisLab. (ed 2.7.1). Bremen, Germany, MeVis Medical Solutions AG, 2015.

2. Amberg B, Romdhani S, Vetter T. Optimal step nonrigid ICP algorithms for surface registration. 2007 IEEE conference on Computer Vision and Pattern Recognition. 2007 17-22 June. Minneapolis, USA.

3. White JD, Ortega-Castrillon A, Matthews H, et al. MeshMonk: Open-source large-scale intensive 3D phenotyping. Sci Rep. 2019;9(1):6085

4. MATLAB. (ed 9.11.0.1837725 (2021b)). Natick, Massachusetts, The Mathworks, Inc., 2021.

5. Cignoni P, Callieri M, Corsini M, et al. Meshlab: an open-source mesh processing tool. Eurographics Italian chapter conference. 2008

6. Matthews HS, Palmer RL, Baynam GS, et al. Large-scale open-source three-dimensional growth curves for clinical facial assessment and objective description of facial dysmorphism. Sci Rep. 2021;11(1):12175

7. Marzola A, Robilotta C, Volpe Y, et al. Statistical Shape Model: comparison between ICP and CPD algorithms on medical applications. Int J Interact Des Manuf. 2021;15:85-89

8. Amira-Avizo. (ed 8.0.0), Thermo Fisher Scientific, 2013.