MeshMonk: open-source large-scale intensive 3D phenotyping

Julie D. White1\*, Alejandra Ortega-Castrillón3,4, Arslan A. Zaidi1,2, Harry Matthews5,6,7, Omid Ekrami8, Jonatan Snyders9, Stefan Van Dongen8, Mark D. Shriver1, Peter Claes3,4\*

1Department of Anthropology, The Pennsylvania State University, University Park, PA, USA.

2Department of Biology, The Pennsylvania State University, University Park, PA, USA.

3Department of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Belgium

4Medical Imaging Research Center, MIRC, UZ Leuven, Leuven, Belgium

5Mudoch Children’s Research Institute, Melbourne, Australia

6Royal Children’s Hospital, Melbourne, Australia

7Department of Pediatrics, University of Melbourne, Melbourne Australia

8Department of Biology, University of Antwerp, Antwerp, Belgium

9WebMonks, Hasselt, Belgium

**\* Correspondence:**jdw345@psu.edu; peter.claes@kuleuven.be

Keywords: Automated landmarking1, automated phenotyping2, non-rigid registration3, phenomics4, morphometrics5, 3D6, facial variation7.

Abstract

Introduction

In the post-genomics era, emphasis has been placed on disentangling ‘genotype-phenotype’ connections so that the biological basis of complex phenotypes can be understood. However, our ability to efficiently and comprehensively characterize phenotypes lags behind our ability to characterize genomes. Anthropometric studies of morphology have traditionally relied on sparse sets of landmarks manually placed on images, from which linear distances and angles are calculated to be used in genetic association studies. This requires the tedious placement of landmarks on many images and is error prone and sensitive to individual differences among observers. Here, we report a toolbox for fast and reproducible high-throughput phenotyping of 3D images. While we demonstrate the utility of this method using 3D facial images, the procedure can also be applied to 3D scans of other complex morphological structures, such as the human brain and skeletal bones.

Methods

Given a facial image (target) with five crude positioning landmarks, a rigid registration is first used to orient an anthropometric mask (reference) to the target scan. Then, using a weighted k-nearest neighbors and a visco-elastic transformation model, the reference is transformed to fit the specific shape of the target. For facial scans, this results in homologous spatially dense (N=7,160) quasi-landmark configurations for all 3D images. As validation, a dataset (N=41) with 19 manually-placed landmarks was superimposed onto the reference in a leave-one-out approach to identify the closest barycentric coordinate on the mask. These coordinates were then projected back onto the training faces and the manual and automatic landmark placements were compared.

Results and Conclusion

We demonstrate that this method is highly accurate, with an average Euclidean distance between the manual and automatic placements of ~1.2 mm. The process is robust to variation due to scan quality, camera systems, and ancestries. Though validated using 19 landmarks, for comparison with traditional methods, this method allows for automated dense phenotyping, freeing the researcher from the use of a limited number of landmarks and allowing for more comprehensive investigations of facial shape variation. This expansion opens up an exciting avenue of study in assessing genomic and phenomic data to better understand the genetic contributions to complex morphological traits.

# Introduction

# Materials and Methods

## MeshMonk Overview (Alejandra)

### Explanation of process (Alejandra)

### Parameters and tuning (Alejandra)

## Spatially dense quasi-landmarking of 3D facial scans

3D images in wavefront.obj file format are imported into an in-house 3D image-cleaning program for cropping and trimming, removing hair, ears, and any dissociated polygons. Five crude positioning landmarks are placed on the face to establish a rough facial orientation, but not to guide the eventual landmark mask to the face. An anthropometric mask (Claes et al., 2012) is non-rigidly mapped (Snyders et al., 2014) onto all 3D images and their reflections, constructed by changing the sign of the *x* coordinate (Claes et al., 2011), using the MeshMonk software and the parameters described above. This establishes homologous spatially dense (~10,000) quasi-landmark (QL) configuration for all 3D images and their reflections. Facial shape can be symmetrized using generalized Procrustes alignment (Rohlf and Slice, 1990) to eliminate differences in position, orientation and size of both original and reflected quasi-landmark configurations. The average of an original and its reflected quasi-landmark configuration constitutes the symmetric component, while the difference between the two configurations constitutes the asymmetric component.

### Facial quality control

Outlier faces, due to quasi-landmark mapping errors, are detected by measuring the Mahalanobis distance for each face to the overall average face in the symmetrized shape space spanned by an orthogonal basis of principal components that captures 98% of the total variation in face shape. From the distribution of Mahalanobis distances, a *z* score for each facial shape is established, and each face with a *z* score equal to or larger than 2 is manually inspected for quasi-landmark errors. Identified erroneous faces are removed, and the whole process starting from the generalized Procrustes superimposition of original and reflected quasi-landmark configurations is repeated.

## Validation

### Sample and data curation

Over many years, our collaborative group has recruited study participants through several studies at the Pennsylvania State University and sampled in the following locations: State College, PA (IRB 44929 and 4320); New York, NY (IRB 45727); Urbana-Champaign, IL (IRB 13103); Dublin, Ireland; Rome, Italy; Warsaw, Poland; and Porto, Portugal (IRB 32341).

Digital facial stereophotogrammetry has been used to capture 3D facial surfaces of N~6,000 participants using the 3dMD Face 2-pod and 3-pod systems (3dMD, Atlanta, GA). This well-established method uses digital photography to generate a dense 3D point cloud representing the surface geometry of the face from multiple 2D images with overlapping fields of view. During photo capture, participant volunteers were asked to adopt a neutral facial expression with their mouth closed and gaze forward, following standard facial image acquisition protocols (Heike et al., 2010). 3D images were immediately stitched together by the camera system and visually checked to make sure that no major holes or artifacts existed.

### Manual placement of validation landmarks

Of the larger sample, N=50 images were chosen at random for validation. This number was then reduced by excluding images from participants that reported major facial injury or surgery and excluding images which did not pass the quality control measures reported above. This resulted in N=41 images for validation. Images were diverse with respect to sex (F=29, M=12), age (18-79, µ=32.7), height (149.86-184.00 cm, µ=167.13 cm), weight (43.00-103.80 kg, µ=67.62 kg), and 3D camera system used (SI Table 1). Most participants reported being of European descent. 3dMDpatient was used to record the 3D coordinates of 19 standard landmarks (7 midline and 12 bilateral) from each unaltered image in wavefront.obj format (Fig. X; Table X). Two independent observers placed these landmarks three times each, with at least 24 hours in-between landmarking sessions, resulting in 6 total landmark iterations for each facial scan. For each individual, we checked for gross landmark coordinate errors (e.g. mislabeling right and left side landmarks) before analysis.

**Table X. Description of landmarks used in validation.** Landmark descriptions from the Richtsmeier Lab (http://www.getahead.la.psu.edu/).

|  |  |  |  |
| --- | --- | --- | --- |
| Landmark | Abvn. | Location | Definition |
| Glabella | g | Midline | The most prominent midline point between the eyebrows. |
| Nasion | n | Midline | The point in the midline of both the nasal root and the nasofrontal suture. This point is always above the line that connects the two inner canthi. |
| Pronasale | prn | Midline | The most protruded point of the apex nasi. |
| Subnasale | sn | Midline | The midpoint of the angle at the columella base where the lower border of the nasal septum and the surface of the upper lip meet. |
| Labiale superius | ls | Midline | The midpoint of the upper vermillion line. |
| Labiale inferius | li | Midline | The midpoint of the lower vermillion line. |
| Pogonion | SPg | Midline | The most anterior point of the chin. |
| Endocanthion | en | Bilateral | The point at the inner commissure of the eye fissure. |
| Exocanthion | ex | Bilateral | The point at the outer commissure of the eye fissure. |
| Alar curvature | ac | Bilateral | The most lateral point in the curved base of each ala. Indicating the facial insertion of the nasal wingbase. |
| Subalare | sbal | Bilateral | The point at the lower limit of each alar base, where the alar base disappears into the skin of the upper lip. The landmarks indicate the labial insertion of the alar base |
| Crista philtri | cph | Bilateral | The lower point on each elevated margin of the philtrum just above the vermillion line. |
| Chelion | ch | Bilateral | Point located at each labial commissure at the most lateral intersection of upper and lower lip. |

### Automatic placement of validation landmarks (Need some sort of image flow chart for this)

To obtain automatic indications of the 19 validation landmarks, a leave-one-out approach was used to identify the placement of the landmark on the anthropometric mask, then the landmarks were projected back on to the left-out face. Specifically, the average manual landmark configurations of 39 faces were aligned to the anthropometric mask and forced to lie on the surface of the mask, if the landmark and the mask differed in the *z-*dimension. The nearest barycentric coordinate (Hille, 1982) on the anthropometric mask was identified using the average manual landmark coordinates and the barycentric coordinates were then placed on the left-out face. This resulted in the automatic placement of the validation landmarks using a “training” set that did not include the test face. The placement of automatic landmarks was performed three times, once using the average of observer AZ’s three landmark iterations, again using the average of observer JW’s three landmark iterations, and a final time using the average of all six iterations from both observers. This process resulted in three placements of automatic landmarks for comparison.

### Statistical analysis

#### Intra- and inter-observer error of manual landmarks

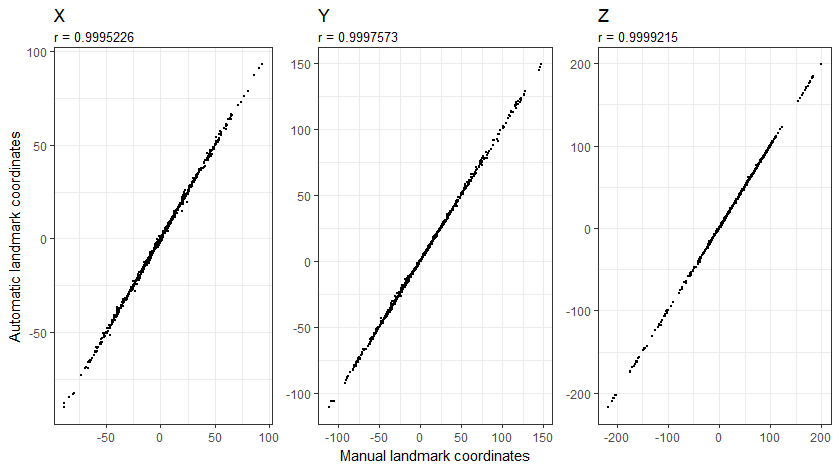
We calculated the intra-observer error as the standard deviation between the *x*, *y*, and *z* coordinates of each observer’s landmarking iterations. Each observer’s landmarking iterations were then averaged to create a centroid landmark configuration for that observer. The standard deviation of the x, y, and z coordinates of each observer’s centroid configurations were taken as the inter-observer error of the manual landmarks. Measures are averaged across dimensions and images (Table X) as well as averaged only across images (SI Table X).

**Table X.** **Intra- and inter-observer error of manual landmarks.** Average of the standard deviation between observer AZ and observer JW’s landmarking iterations and between the centroid of observer AZ and observer JW’s landmark iterations across *x*, *y*, and *z* dimensions and images.

|  |  |  |  |
| --- | --- | --- | --- |
| *Landmark* | *Average SD (mm)* | | |
| *Observer AZ* | *Observer JW* | *Inter-observer* |
| *Alar curvature left* | 0.6020 | 0.4339 | 0.3067 |
| *Alar curvature right* | 0.6304 | 0.3773 | 0.3287 |
| *Chelion left* | 0.6080 | 0.4472 | 0.4182 |
| *Chelion right* | 0.6002 | 0.4934 | 0.2984 |
| *Crista philtri left* | 0.5016 | 0.3041 | 0.3881 |
| *Crista philtri right* | 0.5358 | 0.2949 | 0.4737 |
| *Endocanthion left* | 0.7447 | 0.4372 | 0.4362 |
| *Endocanthion right* | 0.7697 | 0.4462 | 0.3608 |
| *Exocanthion left* | 0.5863 | 0.4380 | 0.2946 |
| *Exocanthion right* | 0.6543 | 0.3579 | 0.2855 |
| *Glabella* | 0.5761 | 0.6881 | 0.4542 |
| *Labiale inferius* | 0.5032 | 0.3175 | 0.5857 |
| *Labiale superius* | 0.4254 | 0.2666 | 0.3185 |
| *Nasion* | 0.5365 | 0.5402 | 0.4938 |
| *Pogonion* | 0.8208 | 0.7593 | 0.5987 |
| *Pronasale* | 0.4593 | 0.3157 | 0.3323 |
| *Subalare left* | 0.5018 | 0.4262 | 0.4005 |
| *Subalare right* | 0.4883 | 0.4848 | 0.4283 |
| *Subnasale* | 0.4504 | 0.4695 | 0.3480 |
| ***Mean*** | **0.5787** | **0.4367** | **0.3974** |

#### Direct comparison of manual and automatic landmark placements

As one measure of validation of the automatic landmark placements, we compared the raw coordinate values of the manual landmarks with the raw coordinate values of the automatic landmarks. Because of the leave-one-out nature of our approach, we can compare the manual and automatic landmark coordinates directly without fear of training bias. We calculated the Pearson’s correlation between the average of all manual landmarking iterations and the automatic landmark placements that were trained using this average. We also calculated the standard deviation between the x, y, and z coordinates of the average of all manual landmarking iterations and the automatic landmarks trained using this average.



**Figure X. Correlation between raw x, y, and z coordinates.** The Pearson’s correlation between the average x, y, and z coordinates from the two observers and the automatic landmarks trained using all manual landmarking data.

**Table X. Standard deviation between manual and automatic landmarks**. We calculated the standard deviation of the x, y, and z coordinates for the manual and automatic landmarks, using the average of all manual landmarking iterations as the training set.

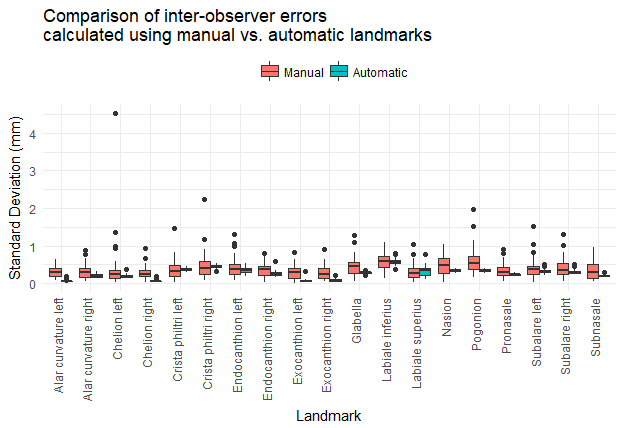
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Landmark* | *Standard deviation (mm)* | | | |
| *X* | *Y* | *Z* | *Mean* |
| *Alar curvature left* | 0.1135 | 0.3700 | 0.4303 | **0.3046** |
| *Alar curvature right* | 0.1169 | 0.3692 | 0.4003 | **0.2954** |
| *Chelion left* | 0.7821 | 0.5042 | 0.4297 | **0.5720** |
| *Chelion right* | 0.6945 | 0.4667 | 0.3917 | **0.5176** |
| *Crista philtri left* | 0.5330 | 0.6313 | 0.3192 | **0.4945** |
| *Crista philtri right* | 0.5343 | 0.7074 | 0.3108 | **0.5175** |
| *Endocanthion left* | 0.5481 | 0.3844 | 0.2830 | **0.4051** |
| *Endocanthion right* | 0.7326 | 0.4608 | 0.3553 | **0.5162** |
| *Exocanthion left* | 0.6421 | 0.5205 | 0.6195 | **0.5941** |
| *Exocanthion right* | 0.6662 | 0.4622 | 0.6687 | **0.5990** |
| *Glabella* | 0.3398 | 0.9230 | 0.3948 | **0.5525** |
| *Labiale inferius* | 0.3225 | 0.5103 | 0.3363 | **0.3897** |
| *Labiale superius* | 0.4163 | 0.5696 | 0.2347 | **0.4069** |
| *Nasion* | 0.2505 | 0.6882 | 0.3357 | **0.4248** |
| *Pogonion* | 0.3050 | 0.7078 | 0.2681 | **0.4269** |
| *Pronasale* | 0.2819 | 0.3964 | 0.1999 | **0.2928** |
| *Subalare left* | 0.5141 | 0.3076 | 0.3938 | **0.4052** |
| *Subalare right* | 0.4615 | 0.3061 | 0.4248 | **0.3975** |
| *Subnasale* | 0.2291 | 0.3360 | 0.1852 | **0.2501** |
| ***Mean*** | **0.4465** | **0.5064** | **0.3675** | **0.4401** |

#### Comparison of inter-observer errors

As an illustration of the low errors involved in the automatic landmark placements, we calculated the inter-observer error between automatic landmark iterations trained using the average of observer AZ’s three landmark iterations and the average of observer JW’s three landmark iterations (Sup Table X). These values can then be compared to the inter-observer error calculated using just the manual landmarks, described in section 2.3.4.1. We additionally performed Levene’s test (Levene, 1960) to determine if the variances of the inter-observer errors calculated using the manual and automatic landmarks were equal (the null hypothesis) or unequal (the alternative hypothesis; Table X).

**Table X. Comparison of inter-observer errors.** The standard deviation between average landmark configurations for the manual and automatic landmarks averaged across scans as well as the F value and P value from performing a Levene’s test per landmark.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Landmark* | *Manual (mm)* | *Auto (mm)* | *F value* | *P value* |
| *Alar curvature left* | 0.3067 | 0.0728 | 59.6244 | **2.83 x 10-11** |
| *Alar curvature right* | 0.3287 | 0.2133 | 22.2346 | **1.01 x 10-5** |
| *Chelion left* | 0.4182 | 0.1998 | 4.6453 | **0.0341** |
| *Chelion right* | 0.2984 | 0.0637 | 24.5101 | **4.03 x 10-6** |
| *Crista philtri left* | 0.3881 | 0.3811 | 29.1832 | **6.60 x 10-7** |
| *Crista philtri right* | 0.4737 | 0.4472 | 18.1685 | **5.49 x 10-5** |
| *Endocanthion left* | 0.4362 | 0.3504 | 14.2000 | **0.0003** |
| *Endocanthion right* | 0.3608 | 0.2669 | 28.4103 | **8.85 x 10-7** |
| *Exocanthion left* | 0.2946 | 0.0808 | 47.7334 | **1.06 x 10-9** |
| *Exocanthion right* | 0.2855 | 0.0961 | 28.0100 | **1.03 x 10-6** |
| *Glabella* | 0.4542 | 0.2938 | 41.5866 | **7.95 x 10-9** |
| *Labiale inferius* | 0.5857 | 0.5773 | 26.3847 | **1.93 x 10-6** |
| *Labiale superius* | 0.3185 | 0.3289 | 2.4213 | 0.1236 |
| *Nasion* | 0.4938 | 0.3511 | 87.7550 | **1.67 x 10-14** |
| *Pogonion* | 0.5987 | 0.3478 | 23.9927 | **4.95 x 10-6** |
| *Pronasale* | 0.3323 | 0.2376 | 38.2428 | **2.49 x 10-8** |
| *Subalare left* | 0.4005 | 0.3239 | 16.4805 | **0.0001** |
| *Subalare right* | 0.4283 | 0.3113 | 25.6819 | **2.54 x 10-6** |
| *Subnasale* | 0.3480 | 0.2072 | 42.6476 | **5.57 x 10-9** |
| *Mean* | 0.3974 | 0.2711 |  |  |



**Figure X. Comparison of inter-observer errors calculated using manual and automatic landmarks.** The interobserver error was calculated as described in section 2.3.4.1 and averaged across x, y, and z dimensions to give an average error value per image. We also calculated the inter-observer error of automatic landmarks trained using the three iterations of each observer separately and averaged these values across x, y, and z dimensions to give an average error value per image. For each landmark, Levene’s test was performed to determine if the variances were identical (Table X).

#### Centroid size comparison

# Results

To validate the placement of automatic landmarks resulting from the MeshMonk anthropometric mask registration, we compared the placement of 19 automatically placed landmarks to those placed manually by two independent observers, while considering the manually placed landmarks to be the “gold standard.” Measurement errors were calculated as the standard deviation between landmarking it the automatic and manual *x*, *y*, and *z* coordinates.

## Intra- and inter-observer error of manual landmarks

The quantitative study of morphology using 3D coordinates requires specific attention to measurement error and has a robust presence in the literature. For each independent observer, we calculated the intra-observer error of the manual landmarks as the standard deviation between the *x*, *y*, and *z* coordinates of each landmark iteration. Table X reports the per-landmark standard deviation, averaged across dimensions and images. The average standard deviation of observer AZ across all landmarks was 0.5787 mm while the average standard deviation of observer JW across all landmarks was 0.4367 mm. The average inter-observer error, measured as the standard deviation between the *x*, *y*, and *z* coordinates of each observer’s centroid configuration was 0.3974 mm. This range of deviation is considered highly precise and is similar to previously reported measures of landmark error (Aldridge et al., 2005; von Cramon-Taubadel et al., 2007).

## Direct comparison of manual and automatic landmark placements

The correlation between the manual and automatic landmarks was calculated based upon the average of all six iterations of manual landmarks and the automatic landmarking iteration based on this average. The Pearson’s correlation coefficients were high: 0.9995226 for the x-dimension, 0.9997573 for the y-dimension, and 0.9999215 for the z-dimension (Figure X). We also calculated the standard deviation between the average manual landmarks and the automatic landmarks, reported in Table X. The standard deviation averaged across dimensions and landmarks was 0.4401 (0.4465 along the x-axis, 0.5064 along the y-axis, and 0.3675 along the z-axis). Per-landmark values are given in Table X.

## Comparison of inter-observer errors

We calculated the inter-observer error using the automatic landmark placements trained using each observer’s manual landmark averages (i.e. AutoAZ vs. AutoJW) and compared this to the inter-observer error calculated using the manual landmark placements (i.e. MLAZ vs. MLJW) using Levene’s test, which was chosen to compare variances while being robust to departures from normality. The inter-observer errors and the Levene test statistics are provided in Table X and correspond to those in Figure X. In all but one case, the variance of the inter-observer error was significantly smaller when calculated using the automatic landmarks. The only case in which the two variances were not significantly different was the labiale superius landmark (F statistic = 2.4213, p-value = 0.1236).

## Centroid sizes

1. **Discussion**

Manual landmarks were considered the gold standard and have long been used and validated in morphological studies (Aldridge paper).

The standard deviations are all considered highly precise, even when calculated as the difference between the ML and auto landmarks

The correlation between the ML and auto landmarks is extremely high

The variance of the Auto landmarks is on a whole MUCH smaller than the ML landmarks. This speaks well of the repeatability of the auto landmarking.

Don’t necessarily have accuracy on the rest of the face (i.e. the cheeks), but neither do manual landmarks.

MeshMonk gives us much more data than the automatic landmarking methods that have the purpose of estimating a sparse set of landmarks. Cite recent successes in GWAS of facial shapes, both clinical and non-clinical (Plos Genetics 2014, Nature Genetics 2018, Karlijne’s paper in this issue).

Opportunities for using MeshMonk on other surfaces besides faces (Harry?)

# Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

# Author Contributions

JW performed validation analyses and landmarked the 3D scans used for validation with AZ. JW and AOC and wrote the first draft of the manuscript under supervision of PC. PC conceptualized the design of the study. HM, OE, SVD, and MS provided input throughout the analyses and writing process. JS developed the MeshMonk code.

# Funding

# Acknowledgments

We thank the participants for providing the data necessary to carry out this validation. We are also grateful to all members of the Shriver Lab and Claes Lab, as well as all the members of WebMonks.

# Ethics statement

Institutional review board (IRB) approval was obtained at all locations and all participants signed a written consent form before participation. The Pennsylvania State University IRB board approved the collection of the participants recruited at the following locations: State College, PA (IRB 44929 and 4320); New York, NY (IRB 45727); Urbana-Champaign, IL (IRB 13103); Dublin, Ireland; Rome, Italy; Warsaw, Poland; and Porto, Portugal (IRB 32341).

# References

Aldridge, K., Boyadjiev, S. A., Capone, G. T., DeLeon, V. B., and Richtsmeier, J. T. (2005). Precision and error of three-dimensional phenotypic measures acquired from 3dMD photogrammetric images. *Am. J. Med. Genet.* 138 A, 247–253. doi:10.1002/ajmg.a.30959.

Claes, P., Walters, M., and Clement, J. (2012). Improved facial outcome assessment using a 3D anthropometric mask. *Int. J. Oral Maxillofac. Surg.* doi:10.1016/j.ijom.2011.10.019.

Claes, P., Walters, M., Vandermeulen, D., and Clement, J. G. (2011). Spatially-dense 3D facial asymmetry assessment in both typical and disordered growth. *J. Anat.* 219, 444–55. doi:10.1111/j.1469-7580.2011.01411.x.

Heike, C. L., Upson, K., Stuhaug, E., and Weinberg, S. M. (2010). 3D digital stereophotogrammetry: a practical guide to facial image acquisition. *Head Face Med.* 6, 18. doi:10.1186/1746-160X-6-18.

Hille, E. (1982). *Analytic Function Theory, Volume I*. Second edi. New York: Chelsea Publishing Company.

Levene, H. (1960). “Robust tests for equality of variances,” in *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, eds. I. Olkin and H. Hotelling (Stanford: Stanford University Press), 278–292.

Rohlf, F. J., and Slice, D. (1990). Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks. *Syst. Zool.* 39, 40–50. doi:10.2307/2992207.

Snyders, J., Claes, P., Vandermeulen, D., and Suetens, P. (2014). Development and comparison of non-rigid surface registraion and extensions (technical report KUL/ESAT/PSI/1401). Leuven, Belgium.

von Cramon-Taubadel, N., Frazier, B. C., and Mirazon-Lahr, M. (2007). The problem of assessing landmark error in geometric morphometrics: Theory, methods, and modifications. *Am. J. Phys. Anthropol.* 134, 24–35. doi:10.1002/ajpa.

# Data Availability Statement

The informed consent with which the data were collected does not allow for dissemination of identifiable data to persons not listed as researchers on the IRB protocol. Thus, the full surface 3D facial images used for validation cannot be made publicly available. In the interest of reproducibility, we have provided the 19 manual and automatic landmarks used for validation as well as the code used to analyze them. These data are available in the following GitHub repository: https://github.com/juliedwhite/RemappingValidation/. The MeshMonk code and tutorials are available at https://github.com/TheWebMonks/meshmonk.