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

Julie D. White1\*, Alejandra Ortega-Castrillón3,4, Harry Matthews5,6,7, Arslan A. Zaidi1,2, 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, genomics5, morphometrics6, 3D7, facial variation8.

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

The phenotypic complement to genomics is *phenomics*, which aims to obtain high-throughput and high-dimensional phenotyping in line with our ability to characterize genomes (Houle et al., 2010). The paradigm shift is simple and similar to the one made in the Human Genome Project: instead of ‘phenotyping as usual’ or measuring a limited set of simplified features that seem relevant, why not measure it all? In contrast to genomic technologies, which successfully measure and characterize complete genomes, the scientific development of phenomics lags behind. However, with the advent of new technologies, hardware exists for extensively and intensively collecting quantitative phenotypic data. For example, 3D image surface and/or medical scanners provide the optimal means to capture information of biological morphology and appearance at the phenomic level. Today, the challenge lies in the ability to provide semantic interpretations from large scale image data that capture the phenome in the context of genetic variations (Walter et al., 2010), which is a challenge that we address with the development of the MeshMonk software.

Of interest to anthropologists, geneticists, biologists, and medical clinicians is the ability to accurately and reproducibly characterize anatomical structures, like a femur, skull, or face, such that underlying qualities about the structure can be understood. The study of variation and covariation in anatomy can provide insights into the genetic causes and evolution of the anatomical structure. In addition, comparing the anatomy of an individual patient to a control population can indicate pathology to a medical practitioner. Traditionally, this has been achieved using visual clinical assessment or by taking measurements between manually placed anatomical ‘landmarks’. Landmarks are traditionally defined as precise locations on biological forms that hold some developmental, functional, structural or evolutionary significance (Richtsmeier et al., 2002) that are unambiguously defined and reliably locatable (Aldridge et al., 2005; Corner et al., 1992; Richtsmeier et al., 1995). Some examples include the endo- and exocanthi (the inner and outer corners of the eyes, respectively) and the pronasale (the tip of the nose).

However, manual landmarking is tedious to perform, difficult to standardize in practice, and prone to intra and inter-operator error (Cramon-Taubadel et al., 2007; Fagertun et al., 2014; Toma et al., 2009; Weinberg et al., 2004; Wong et al., 2008). Furthermore, sparse landmark configurations can only quantify the form at defined landmark indications that can be reliably identified and indicated by a human. An alternative is to automatically indicate quasi-landmarks across the entire surface of the structure. This is achieved by gradually warping a generic template image (anthropometric mask) composed of thousands of points into the shape of each target image through a non-rigid registration algorithm (Andresen and Nielsen, 2001; Claes, 2007; Claes et al., 2012; Hutton et al., 2003b; Snyders et al., 2014). The coordinates of these warped templates, now in the shape of each target, can then be treated by geometric morphometric analysis. An automatic approach like this is preferable for the analysis of large datasets, avoiding the problems of manual landmarking at different sites by multiple operators. They are also more suitable for applications that require synthesis of a recognizable instance of the actual structure, such as predicting a complete shape from DNA (Claes et al., 2014a), synthetic growth and ageing of a face (Imaizumi et al., 2015; Matthews et al., 2018a), constructing 3D facial composites for forensic applications (Blanz and Vetter, 1999), and characterization of dysmorphology for clinical diagnostics (Baynam et al., 2014; Hammond et al., 2005). Here, we report the MeshMonk toolbox for fast and reproducible high-throughput phenotyping of 3D images, or quasi-landmark indication, which can be applied to 3D facial images as well as 3D scans of other complex morphological structures, such as the human brain and skeletal bones.

Surface registration, utilized in the MeshMonk software, defines a mapping of the vertices from one (template) image onto their corresponding locations on another (target) and allows us to quantify and visualize both subtle and acute variation in surface form across a sample by finding the geometrical relationship (one-to-one correspondences) between 3D shapes following a predefined transformation model (Andresen and Nielsen, 2001; Claes, 2007; Claes et al., 2012; Hutton et al., 2003a; Snyders et al., 2014). When the template is warped onto each target, the coordinates of any anatomical landmark, manually annotated on the template, is also defined on each target, thus the complete quasi-landmark indication can also be considered a method for automatic placement of sparse anatomical landmarks (Wei et al., 2011). As a validation of the MeshMonk toolbox we compare manual and automatic indications of sparse landmarks.

# Materials and Methods

## MeshMonk

MeshMonk is a free, open-source implementation of a modular surface registration framework developed in a partnership between researchers at the Medical Imaging Research Center (MIRC) at KU Leuven, Pennsylvania State University (PSU), and WebMonks ([www.webmonks.vision](http://www.webmonks.vision)), with MIRC and PSU delivering the research and IP behind the methods and algorithms and WebMonks being the implementation partner.

The C++ library takes a multi-scale, iterative ICP-based approach (Besl & McKay, 1992). Characteristic to its registration process are (1) a bi-directional, weighted K-Nearest Neighbor point matching algorithm, (2) an outlier classification step and (3) a Visco-Elastic transformation model (Snyders et al., 2014). With the library come wrappers to mex the library’s functions so that they can be used in Matlab.

## Explanation of process (Alejandra?)

## Parameters and tuning (Alejandra?)

## Validation

### Sample and data curation (Give supplementary figure of three different facial cameras)

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). Stereo photogrammetry was 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 generates a dense 3D point cloud representing the surface geometry of the face from multiple 2D images with overlapping fields of view. During photo capture, participants were asked to adopt a neutral facial expression with their mouth closed and to gaze forward, following standard facial image acquisition protocols (Heike et al., 2010). 3D surface images were visually checked to make sure that no major holes or artifacts existed.

### Manual placement of validation landmarks

Of the larger sample, N=48 surface images were chosen at random for validation. This number was then reduced by excluding surface images from participants that reported major facial injury or surgery. This resulted in N=41 surface images for validation, which were diverse with respect to sex (NFemale=29, NMale=12), age (range: 18-79, =32.7), height (range: 149.86-184.00 cm, *M* =167.13 cm), weight (range: 43.00-103.80 kg, *M* =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 surface (i.e. still containing hair and clothing) in wavefront.obj format (Fig. X; Table X). Two independent observers placed landmarks three times each, with at least 24 hours in-between landmarking sessions, resulting in 6 total landmark indications 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 | Abbr. | 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 validation landmark on the anthropometric mask, then indicate these landmarks on the left-out face. Specifically, for each surface image the manual landmark coordinates were averaged and aligned to the anthropometric mask using barycentric coordinates (Hille, 1982), giving a set of 41 total landmark placements on the anthropometric mask, which were then converted to cartesian coordinates. One by one, each face was left out while averaging the other 40 landmark placements to “train” the automatic landmarks. This average was then placed back onto the left-out (target) face, which resulted in the automatic placement of the validation landmarks using a “training” set that did not include the target face.

The placement of automatic landmarks was performed three times, changing the manual landmark data used as input: once using the average of observer AZ’s three manual landmark iterations, again using the average of observer JW’s three manual landmark iterations, and a final time using the average of all six manual landmark 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 indications. Each observer’s landmarking indications were averaged together, per landmark and per face, to produce the average landmark position for that observer. The root mean squared error between the x, y, and z coordinates of each observer was calculated for the inter-observer error. 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.** Standard deviation between observer AZ’s three landmark iterations and observer JW’s three landmark iterations (intra-observer error) as well as inter-observer error, or the root mean squared error between the average of AZ’s three landmark iterations and the average of JW’s three landmark iterations. Values have been averaged across each face as well as x, y, and z axes to give an estimate of the error per landmark.

|  |  |  |  |
| --- | --- | --- | --- |
| *Landmark* | *Standard deviation (mm)* | | *RMSE (mm)* |
| *Observer AZ* | *Observer JW* | *Inter-observer* |
| *Alar curvature left* | 0.6020 | 0.4339 | 0.4337 |
| *Alar curvature right* | 0.6304 | 0.3773 | 0.4648 |
| *Chelion left* | 0.6080 | 0.4472 | 0.5914 |
| *Chelion right* | 0.6002 | 0.4934 | 0.4220 |
| *Crista philtri left* | 0.5016 | 0.3041 | 0.5488 |
| *Crista philtri right* | 0.5358 | 0.2949 | 0.6699 |
| *Endocanthion left* | 0.7447 | 0.4372 | 0.6168 |
| *Endocanthion right* | 0.7697 | 0.4462 | 0.5102 |
| *Exocanthion left* | 0.5863 | 0.4380 | 0.4166 |
| *Exocanthion right* | 0.6543 | 0.3579 | 0.4038 |
| *Glabella* | 0.5761 | 0.6881 | 0.6423 |
| *Labiale inferius* | 0.5032 | 0.3175 | 0.8283 |
| *Labiale superius* | 0.4254 | 0.2666 | 0.4504 |
| *Nasion* | 0.5365 | 0.5402 | 0.6983 |
| *Pogonion* | 0.8208 | 0.7593 | 0.8466 |
| *Pronasale* | 0.4593 | 0.3157 | 0.4700 |
| *Subalare left* | 0.5018 | 0.4262 | 0.5664 |
| *Subalare right* | 0.4883 | 0.4848 | 0.6057 |
| *Subnasale* | 0.4504 | 0.4695 | 0.4921 |
| ***Mean*** | **0.5787** | **0.4367** | **0.5620** |

#### 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 intraclass correlation coefficient between the average of all six manual landmarking iterations and the automatic landmark placements that were trained using this average. We also calculated the root mean squared error 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. Bland-Altman plot for similarity between manual and automatic landmark placements.** For x, y, and z, Bland-Altman plot showing the differences between the manual and automatic landmark placements against the averages of the two techniques. Blue lines represent the mean difference value (solid) and 95% confidence limits (dashed). Red lines represent the upper and lower limits (sold) and the 95% confidence limits (dashed). Also given are the intra-class correlation coefficient, and 95% confidence interval for the ICC, for the manual and automatic comparison.

**Table X. Root mean squared error between manual and automatic landmarks**. Root mean squared error between the manual and automatic landmarks for the x, y, and z coordinates was calculated using the mean of all manual landmark indications and the automatic data trained using this mean. Values are presented for each axis, averaged across all faces, as well as averaged across the axes (mean).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Landmark* | *Root mean squared error (mm)* | | | |
| *X* | *Y* | *Z* | *Mean* |
| *Alar curvature left* | 0.1605 | 0.5233 | 0.6085 | **0.4308** |
| *Alar curvature right* | 0.1653 | 0.5221 | 0.5661 | **0.4178** |
| *Chelion left* | 1.1061 | 0.7131 | 0.6077 | **0.8089** |
| *Chelion right* | 0.9822 | 0.6600 | 0.5539 | **0.7320** |
| *Crista philtri left* | 0.7537 | 0.8927 | 0.4515 | **0.6993** |
| *Crista philtri right* | 0.7556 | 1.0005 | 0.4395 | **0.7319** |
| *Endocanthion left* | 0.7751 | 0.5437 | 0.4002 | **0.5730** |
| *Endocanthion right* | 1.0360 | 0.6517 | 0.5024 | **0.7300** |
| *Exocanthion left* | 0.9081 | 0.7362 | 0.8761 | **0.8401** |
| *Exocanthion right* | 0.9421 | 0.6537 | 0.9457 | **0.8472** |
| *Glabella* | 0.4806 | 1.3053 | 0.5583 | **0.7814** |
| *Labiale inferius* | 0.4560 | 0.7216 | 0.4756 | **0.5511** |
| *Labiale superius* | 0.5887 | 0.8055 | 0.3319 | **0.5754** |
| *Nasion* | 0.3543 | 0.9732 | 0.4748 | **0.6008** |
| *Pogonion* | 0.4313 | 1.0009 | 0.3791 | **0.6038** |
| *Pronasale* | 0.3987 | 0.5606 | 0.2827 | **0.4140** |
| *Subalare left* | 0.7271 | 0.4349 | 0.5570 | **0.5730** |
| *Subalare right* | 0.6526 | 0.4329 | 0.6008 | **0.5621** |
| *Subnasale* | 0.3239 | 0.4752 | 0.2620 | **0.3537** |
| ***Mean*** | **0.1605** | **0.5233** | **0.6085** | **0.4308** |

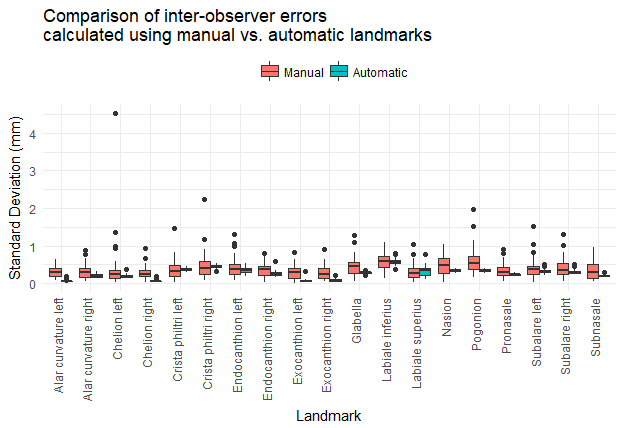
#### Comparison of Arslan ML – Arslan Auto, Julie ML – Julie Auto, etc. 2 way comparison to make sure that the automatic landmarking doesn’t add error

#### Comparison of inter-observer errors (not sure if I want to keep this)

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).

Ability to have up to 50 indications on a single face and then average them is what makes this good. Even though they were done on different faces. Not available on manual . Mapping allowed us to merge

#### Centroid size comparison

Citation for geomorph: Adams and Otarola-Castillo 2013

#### Manovas

#### Using the average of six iterations vs. the average of three iterations? Probably supplementary material.

# Results (not updated)

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 root mean squared error 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; 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 (In the discussion, make sure to talk about how this is an expected result because of the averaging of many landmarks during the training process).

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).

## Comparison of Arslan ML – Arslan Auto, Julie ML – Julie Auto, etc. 2 way comparison to make sure that the automatic landmarking doesn’t add error

## Centroid size comparison

1. **Discussion**

Many authors have supplemented or replaced traditional analysis of inter-landmark measurements with geometric morphometric analysis (Adams et al., 2004; Rohlf and Marcus, 1993), which uses representations of the complete geometry of a configuration of landmarks, rather than simple distances. This can take the form of all pairwise distances between landmarks (Lele and Richtsmeier, 1991), but is more commonly the coordinates of the landmarks themselves. These techniques allow multivariate analysis, comparison, and prediction of the entire form (up to the resolution and density of landmark points) and have become a mainstay of taxonomic analyses (Baab, 2008; Frost et al., 2003; Havarti et al., 2004; Terhune et al., 2007), morphological evolution (Bastir et al., 2006; Klingenberg, 2010; Klingenberg and Marugan-Lobon, 2013; O’Higgins, 2000), the examination of morphological ontogeny and growth (Kesterke et al., 2016; Martinez-Abadias et al., 2012; Matthews et al., 2018b; Mitteroecker et al., 2004; Smith, 2006), studies of population admixture (Martinez-Abadias et al., 2006; Quinto-Sánchez et al., 2015; Schlager and Alexandra, 2015), and genotype-phenotype mapping studies (Claes et al., 2014b, 2018; Liu et al., 2012; Paternoster et al., 2012; Shaffer et al., 2016), and studies of dysmorphology (Hammond et al., 2014; Klingenberg et al., 2010; Richtsmeier et al., 2000; Shaner et al., 2000; Starbuck et al., 2011).

Through studies utilizing manually placed sparse landmarks, we have begun to understand the biological basis and evolution of complex phenotypes, both normative and clinical. However, there is still much to be learned. One avenue for improvement is to expand and speed up the production and analysis of data using methods derived from engineering and computer vision, which allow for the description of shapes as “big data” structures instead of sparse sets of landmarks or linear distances, thus matching our ability to describe phenotypes with our ability to describe genomes.

## Validation

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.

## Usefulness of MeshMonk (previous and future uses)

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?)

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

One of the possible applications of MeshMonk is spatially dense landmarking of 3D facial scans. This process involves the cleaning of 3D surface image to remove hair, ears, and any dissociated polygons. Five crude positioning landmarks are then 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 surface images and their reflections, constructed by changing the sign of the *x* coordinate (Claes et al., 2011), using the MeshMonk software and parameters described in the methods. This establishes homologous spatially dense (~10,000) quasi-landmark (QL) configuration for all 3D surface 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. Mahalanobis distance for each face to the overall average face in the symmetrized shape space can be used to detect registration outliers.

### Brains

### Skulls

### Femur

## Future improvements/issues with the algorithm

# Conclusion

# 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, AOC, and HM wrote the first draft of the manuscript under supervision of PC. PC and JW conceptualized the design of the study. 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 and Claes Labs, 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

Adams, D. C., Rohlf, F. J., and Slice, D. E. (2004). Geometric morphometrics: Ten years of progress following the ‘revolution.’ *Ital. J. Zool.* 71, 5–16. doi:10.1080/11250000409356545.

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.

Andresen, P. R., and Nielsen, M. (2001). Non-rigid registration by geometry-constrained diffusion. *Med. Image Anal.* 5, 81–88. doi:10.1016/S1361-8415(00)00036-0.

Baab, K. L. (2008). The taxonomic implications of cranial shape variation in Homo erectus. *J. Hum. Evol.* 54, 827–847. doi:10.1016/j.jhevol.2007.11.003.

Bastir, M., Rosas, A., and O’Higgins, P. (2006). Craniofacial levels and the morphological maturation of the human skull. *J. Anat.* 209, 637–654. doi:10.1111/j.1469-7580.2006.00644.x.

Baynam, G., Walters, M., Claes, P., Kung, S., LeSouef, P., Dawkins, H., et al. (2014). Phenotyping: Targeting genotype’s rich cousin for diagnosis. *J. Paediatr. Child Health* 51, 381–386. doi:10.1111/jpc.12705.

Blanz, V., and Vetter, T. (1999). A morphable model for the synthesis of 3D faces. in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques - SIGGRAPH ’99* (New York, NY: ACM Press/Addison-Wesley Publishing Co), 187–194. doi:10.1145/311535.311556.

Claes, P. (2007). A robust statistical surface registration framework using implicit function representations: application in craniofacial reconstruction. Available at: http://mic.uzleuven.be/download/public/MIC/publications/2967/PHD\_ pclaes.pdf.

Claes, P., Hill, H., and Shriver, M. (2014a). Towards DNA-based facial composites: preliminary results and validation. *Forensic Sci. Int.*

Claes, P., Liberton, D., Daniels, K., Matthes Rosana, K., Quillen, E., Pearson, L., et al. (2014b). Modeling 3D Facial Shape from DNA. *PLOS Genet.* 10, 1–14.

Claes, P., Roosenboom, J., White, J. D., Swigut, T., Sero, D., Li, J., et al. (2018). Genome-wide mapping of global-to-local genetic effects on human facial shape. *Nat. Genet.* doi:10.1038/s41588-018-0057-4.

Claes, P., Walters, M., and Clement, J. (2012). Improved facial outcome assessment using a 3D anthropometric mask. *Int. J. Oral Maxillofac. Surg.* 41, 324–330. 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.

Corner, B. D., Lele, S., and Richtsmeier, J. T. (1992). Measuring Precision of Three-Dimensional Landmark Data. *J. Quantative Anthropol.* 3, 347–359.

Cramon-Taubadel, N. Von, 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.

Fagertun, J., Harder, S., Rosengren, A., Moeller, C., Werge, T., Paulsen, R. R., et al. (2014). 3D facial landmarks: Inter-operator variability of manual annotation. *BMC Med. Imaging* 14, 35. doi:10.1186/1471-2342-14-35.

Frost, S., Marcus, L., Bookstein, F. L., Reddy, D., and Delson, E. (2003). Cranial allometry, phylogeography, and systematics of large-bodied papionins (Primates: Cercopithecinae) inferred from geometric morphometric analysis of landmark data. *Anat. Rec. Part A* 275A, 1048–1072. doi:10.1002/ar.a.10112.

Hammond, P., Hutton, T. J., Allanson, J. E., Buxton, B. F., Campbell, L. E., Clayton-Smith, J., et al. (2005). Discriminating Power of Localized Three-Dimensional Facial Morphology. *Am. J. Hum. Genet.* 77, 999–1010.

Hammond, P., McKee, S., Suttie, M., Allanson, J. E., Cobben, J.-M., Maas, S. M., et al. (2014). Opposite effects on facial morphology due to gene dosage sensitivity. *Hum. Genet.* 133, 1117–25. doi:10.1007/s00439-014-1455-z.

Havarti, K., Frost, S., and McNulty, K. (2004). Neanderthal taxonomy reconsidered: implications of 3D primate models of intra- and interspecific differences. *Proc. Natl. Acad. Sci. USA* 101, 1147–1152. doi:10.1046/j.1469-7580.2000.19710103.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.

Houle, D., Govindaraju, D. R., and Omholt, S. (2010). Phenomics: The next challenge. *Nat. Rev. Genet.* 11, 855–866. doi:10.1038/nrg2897.

Hutton, T., Buxton, B., Hammond, P., and Unit, B. (2003a). Automated registration of 3D faces using dense surface models. in *British Machine Vision Conference*, eds. R. Harvey and A. Bangham (Norwich: Citeseer), 1–10. doi:10.5244/C.17.45.

Hutton, T. J., Buxton, B. F., Hammond, P., and Potts, H. W. W. (2003b). Estimating Average Growth Trajectories in Shape-Space Using Kernel Smoothing. *IEEE Trans. Med. Imaging* 22, 747–753.

Imaizumi, K., Taniguchi, K., Ogawa, Y., Matsuzaki, K., Nagata, T., Mochimaru, M., et al. (2015). Three-dimensional analyses of aging-induced alterations in facial shape: a longitudinal study of 171 Japanese males. *Int. J. Legal Med.* 129, 385–393. doi:10.1007/s00414-014-1114-x.

Kesterke, M. J., Raffensperger, Z. D., Heike, C. L., Cunningham, M. L., Hecht, J. T., Kau, C. H., et al. (2016). Using the 3D Facial Norms Database to investigate craniofacial sexual dimorphism in healthy children, adolescents, and adults. *Biol. Sex Differ.* 7. doi:10.1186/s13293-016-0076-8.

Klingenberg, C. P. (2010). Evolution and development of shape: integrating quantitative approaches. *Nat. Rev. Genet.* 11, 623–635.

Klingenberg, C. P., and Marugan-Lobon, J. (2013). Evolutionary Covariation in Geometric Morphometric Data: Analyzing Integration, Modularity, and Allometry in a Phylogenetic Context. *Syst. Biol.* 62, 591–610. doi:10.1093/sysbio/syt025.

Klingenberg, C. P., Wetherill, L., Rogers, J., Moore, E., Ward, R., Autti-Rämö, I., et al. (2010). Prenatal alcohol exposure alters the patterns of facial asymmetry. *Alcohol* 44, 649–657. doi:10.1016/j.alcohol.2009.10.016.

Lele, S., and Richtsmeier, J. T. (1991). Euclidean distance matrix analysis: A coordinate‐free approach for comparing biological shapes using landmark data. *Am. J. Phys. Anthropol.* 86, 415–427. doi:10.1002/ajpa.1330860307.

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.

Liu, F., van der Lijn, F., Schurmann, C., Zhu, G., Chakravarty, M. M., Hysi, P. G., et al. (2012). A Genome-Wide Association Study Identifies Five Loci Influencing Facial Morphology in Europeans. *PLoS Genet.* 8. doi:10.1371/journal.pgen.1002932.

Martinez-Abadias, N., Gonzalez-Jose, R., Gonzalez-Martin, A., Van der Molen, S., Talavera, A., Hernandez, P., et al. (2006). Phenotypic evolution of human craniofacial morphology after admixture: a geometric morphometrics approach. *Am. J. Phys. Anthropol.* 129, 387–98. doi:10.1002/ajpa.20291.

Martinez-Abadias, N., Mitteroecker, P., Parsons, T. E., Esparza, M., Sjovold, T., Rolian, C., et al. (2012). The Developmental Basis of Quantitative Craniofacial Variation in Humans and Mice. *Evol Biol* 39, 554–567.

Matthews, H., Penington, A., Clement, J., Kilpatrick, N., Fan, Y., and Claes, P. (2018a). Estimating age and synthesising growth in children and adolescents using 3D facial prototypes. *Forensic Sci. Int.* 286, 61–69. doi:10.1016/j.forsciint.2018.02.024.

Matthews, H. S., Penington, A. J., Hardiman, R., Fan, Y., Clement, J. G., Kilpatrick, N. M., et al. (2018b). Modelling 3D craniofacial growth trajectories for population comparison and classification illustrated using sex-differences. *Sci. Rep.* 8. doi:10.1038/s41598-018-22752-5.

Mitteroecker, P., Gunz, P., Bernhard, M., Schaefer, K., and Bookstein, F. L. (2004). Comparison of cranial ontogenetic trajectories among great apes and humans. *J. Hum. Evol.* 46, 679–698. doi:10.1016/j.jhevol.2004.03.006.

O’Higgins, P. (2000). The study of morphological variation in the hominid fossil record: biology, landmarks and geometry. *J. Anat.* 197, 103–120.

Paternoster, L., Zhurov, A. I., Toma, A. M., Kemp, J. P., St. Pourcain, B., Timpson, N. J., et al. (2012). Genome-wide association study of three-dimensional facial morphology identifies a variant in PAX3 associated with nasion position. *Am. J. Hum. Genet.* 90, 478–485. doi:10.1016/j.ajhg.2011.12.021.

Quinto-Sánchez, M., Adhikari, K., Acuña-Alonzo, V., Cintas, C., Silva de Cerqueira, C. C., Ramallo, V., et al. (2015). Facial asymmetry and genetic ancestry in Latin American admixed populations. *Am. J. Phys. Anthropol.* 157, 58–70. doi:10.1002/ajpa.22688.

Richtsmeier, J. T., Baxter, L. L., and Reeves, R. H. (2000). Parallels of craniofacial maldevelopment in Down syndrome and Ts65Dn mice. *Dev. Dyn.* 217, 137–45. doi:10.1002/(SICI)1097-0177(200002)217:2<137::AID-DVDY1>3.0.CO;2-N.

Richtsmeier, J. T., Burke Deleon, V., and Lele, S. R. (2002). The promise of geometric morphometrics. *Am. J. Phys. Anthropol.* 119, 63–91. doi:10.1002/ajpa.10174.

Richtsmeier, J. T., Paik, C. H., Elfert, P. C., Cole III, T. M., and Dahlman, H. R. (1995). Precision, Repeatibility, and Validation of the Localization of Cranial Landmarks Using Computed Tomography Scans. *Cleft Palate-Craniofacial J.* 32, 217–227.

Rohlf, F. J., and Marcus, L. F. (1993). A revolution morphometrics. *Trends Ecol. Evol.* 8, 129–132. doi:10.1016/0169-5347(93)90024-J.

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.

Schlager, S., and Alexandra, R. (2015). Analysis of the Human Osseous Nasal Shape — Population Differences and Sexual Dimorphism. *Am. J. Phys. Anthropol.* 00. doi:10.1002/ajpa.22749.

Shaffer, J. R., Orlova, E., Lee, M. K., Leslie, E. J., Raffensperger, Z. D., Heike, C. L., et al. (2016). Genome-Wide Association Study Reveals Multiple Loci Influencing Normal Human Facial Morphology. *PLoS Genet.* 12, 1–21. doi:10.1371/journal.pgen.1006149.

Shaner, D. J., Peterson, A. E., Beattie, O. B., and Bamforth, J. S. (2000). Assessment of soft tissue facial asymmetry in medically normal and syndrome-affected individuals by analysis of landmarks and measurements. *Am. J. Med. Genet.* 93, 143–154. doi:10.1002/1096-8628(20000717)93:2<143::AID-AJMG12>3.0.CO;2-Q.

Smith, K. K. (2006). Craniofacial development in marsupial mammals: developmental origins of evolutionary change. *Dev. Dyn.* 235, 1181–93. doi:10.1002/dvdy.20676.

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.

Starbuck, J. M., Reeves, R. H., and Richtsmeier, J. T. (2011). Morphological integration of soft-tissue facial morphology in Down Syndrome and siblings. *Am. J. Phys. Anthropol.* 146, 560–8. doi:10.1002/ajpa.21583.

Terhune, C. E., Kimbel, W. H., and Lockwood, C. A. (2007). Variation and diversity in Homo erectus: a 3D geometric morphometric analysis of the temporal bone. *J. Hum. Evol.* 53, 41–60.

Toma, A. M., Zhurov, A., Playle, R., Ong, E., and Richmond, S. (2009). Reproducibility of facial soft tissue landmarks on 3D laser-scanned facial images. *Orthod. Craniofacial Res.* 12, 33–42. doi:10.1111/j.1601-6343.2008.01435.x.

Walter, T., Shattuck, D. W., Baldock, R., Bastin, M. E., Carpenter, A. E., Duce, S., et al. (2010). Visualization of image data from cells to organisms. *Nat. Methods* 7, S26–S41. doi:10.1038/nmeth.1431.

Wei, R., Claes, P., Walters, M., Wholley, C., and Clement, J. G. (2011). Augmentation of linear facial anthropometrics through modern morphometrics: A facial convexity example. *Aust. Dent. J.* 56, 141–147. doi:10.1111/j.1834-7819.2011.01315.x.

Weinberg, S. M., Scott, N. M., Neiswanger, K., Brandon, C. A., and Marazita, M. L. (2004). Digital three-dimensional photogrammetry: Evaluation of anthropometric precision and accuracy using a Genex 3D camera system. *Cleft Palate-Craniofacial J.* 41, 507–518. doi:10.1597/03-066.1.

Wong, J. Y., Oh, A. K., Ohta, E., Hunt, A. T., Rogers, G. F., Mulliken, J. B., et al. (2008). Validity and reliability of craniofacial anthropometric measurement of 3D digital photogrammetric images. *Cleft Palate-Craniofacial J.* 45, 232–239. doi:10.1597/06-175.

# 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.