



Three-dimensional facial capture using a custom-built photogrammetry setup: Design, performance, and cost

Hans L. L. Wellens,^a Hanne Hoskens,^{b,c} Peter Claes,^{b,c,d,e,f} Anne Marie Kuijpers-Jagtman,^{g,h,i} and Alejandra Ortega-Castrillón^{b,c}

Sint-Michiels and Leuven, Belgium, and Melbourne, Australia, and Oxford, United Kingdom, and Groningen, The Netherlands, and Bern, Switzerland, and Jakarta, Indonesia

Introduction: Although stereophotogrammetry is increasingly popular for 3-dimensional face scanning, commercial solutions remain quite expensive, limiting its accessibility. We propose a more affordable, custom-built photogrammetry setup (Stereo-Face 3D, SF3D) and evaluate its variability within and between systems. **Methods:** Twenty-nine subjects and a mannequin head were imaged 3 times using SF3D and a commercially available system. An anthropometric mask was mapped viscoelastically onto the reconstructed meshes using MeshMonk (<https://github.com/TheWebMonks/meshmonk>). Within systems, shape variability was determined by calculating the root-mean-square error (RMSE) of the Procrustes distance between each of the subject's 3 scans and the subject's ground truth (calculated by averaging the mappings after a nonscaled generalized Procrustes superimposition). Intersystem variability was determined by similarly comparing the ground truth mappings of both systems. Two-factor Procrustes analysis of variance was used to partition the intersystem shape variability to understand the source of the discrepancies between the facial shapes acquired by both systems. **Results:** The RMSEs of the within-system shape variability for 3dMDFace and SF3D were 0.52 ± 0.07 mm and 0.44 ± 0.16 mm, respectively. The corresponding values for the mannequin head were 0.42 ± 0.02 mm and 0.29 ± 0.03 mm, respectively. The between-systems RMSE was 1.6 ± 0.34 mm for the study group and 1.38 mm for the mannequin head. A 2-factor analysis indicated that variability attributable to the system was expressed mainly at the upper eyelids, nasal tip and alae, and chin areas. **Conclusions:** The variability values of the custom-built setup presented here were competitive to a state-of-the-art commercial system at a more affordable level of investment. (*Am J Orthod Dentofacial Orthop* 2020;158:286-99)

In orthodontics, facial esthetics has traditionally been scrutinized using the lateral cephalogram's profile

^aPrivate practice, Sint-Michiels, Belgium.

^bDepartment of Electrical Engineering, ESAT/PSI, Katholieke Universiteit Leuven, Leuven, Belgium.

^cMedical Imaging Research Center, Universitair Ziekenhuis Leuven, Leuven, Belgium.

^dDepartment of Human Genetics, Katholieke Universiteit Leuven, Leuven, Belgium.

^eMurdoch Children's Research Institute, Melbourne, Australia.

^fDepartment of Biomedical Engineering, University of Oxford, Oxford, United Kingdom.

^gDepartment of Orthodontics, University Medical Center Groningen, Groningen, The Netherlands.

^hDepartment of Orthodontics, University of Bern, Bern, Switzerland.

ⁱFaculty of Dentistry, University of Indonesia, Jakarta, Indonesia.

All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest, and none were reported.

Address correspondence to: Hans L. L. Wellens, Groene-Poortdreef 16, Sint-Michiels 8200, Belgium; e-mail, wellens.hans@telenet.be.

Submitted, September 2019; revised, December 2019; accepted, January 2020. 0889-5406/\$36.00

© 2020 by the American Association of Orthodontists. All rights reserved.

<https://doi.org/10.1016/j.ajodo.2020.01.016>

outline combined with classic 2-dimensional (2D) facial photographs. With the introduction of cutting edge 3-dimensional (3D) imaging techniques into the orthodontic and/or craniofacial diagnostic toolset, (structured light) photogrammetry setups and 3D facial images derived from full-size cone-beam computed tomography (CBCT) exposures¹ have increasingly been adopted for this purpose. Three-dimensional CBCT does carry an increased radiation burden compared with traditional 2D radiology, especially if image quality is of primary concern. Combined with the ALARA principle,^{2,3} this has so far precluded its use as a de facto imaging solution for orthodontic diagnosis, at least in Europe.⁴ In addition, the restraints required to immobilize the patient's head during image capture using CBCT may potentially obscure facial regions of interest, such as the forehead and/or chin area. Combined with ethical objections associated with repeatedly exposing patients to ionizing radiation for growth-monitoring purposes,

treatment follow-up, or outcome assessment, this entails that there might be a bright future for nonionizing methods for diagnosing facial esthetics, growth, or treatment change, such as photogrammetry.

Several studies report on the accuracy and reliability of various commercially available photogrammetry solutions applied in an orthodontic and/or craniofacial setting. These include both active stereophotogrammetry solutions (which illuminate the patient's face with invisible structured-light patterns to provide the features required for interpreting and reconstructing the face's 3D geometry) from manufacturers such as AxisThree,⁵ and passive ones (which reconstruct the scene directly from visual cues present in the acquired image) from Canfield Imaging Systems (eg, VECTRA; Fairfield, NJ)⁶⁻⁸ and Dimensional Imaging (DI3D).^{5,9,10} Hybrid stereophotogrammetry solutions, which combine both techniques to achieve an optimal result,⁵ have also been presented by 3dMD.¹¹⁻¹³ Even low-cost solutions, such as David's SLS-2,¹⁴ Fuel3D's Scanify,^{15,16} and Microsoft's Kinect,¹⁷ have been investigated.

Interestingly, some of the aforementioned studies use direct anthropometry (ie, caliper and measuring tape) as the "gold standard," notwithstanding the notable variability of the latter approach.^{8,9,12} Aside from the general sparsity of the human face in terms of clearly definable landmarks, direct anthropometry is additionally hampered by skin compressibility and slight changes in facial expression. Some studies attempt to minimize the effects of landmark identification error by prelabeling the facial surfaces.⁸⁻¹⁰ Both tissue compressibility and facial pose variation can be circumvented by performing measurements on a mannequin head,^{8,9,12-14} or on plaster casts.^{10,15} Other studies replace direct anthropometric measurements with (repeated) digital ones,¹³ make use electromagnetic digitizers, coordinate-measuring machines,¹⁰ or other stereophotogrammetry devices,^{12,14,15,17} which are the "gold standard" for comparison.

A problem not adequately addressed by studies is the very feature-sparse nature of the human face, which entails broad regions having few landmarks at which accuracy and reliability can be gauged. One relatively straightforward solution might be to use elastic deformation of a standard anthropometric mask to densely sample and model the entire facial surface with a very large number of landmarks which, by the elastic deformation, are effectively homologous.¹⁸

In general, cutting edge technologies such as 3D facial capture typically demand a significant premium over traditional 2D methods. Budgetary constraints often deprive orthodontic departments and private practices alike from access to these technologies. Aside from

the cost, 1 of the major impediments to the democratization of this technology has traditionally been the imposing complexity of the photogrammetric algorithms involved in reconstructing the 3D scene. This limitation has changed somewhat with the introduction of relatively affordable and high-quality multibase photogrammetry software. In the latter, image reconstruction proceeds from a relatively large set of images taken from multiple but slightly different viewpoints. This approach was made possible because of relatively recent, fast-paced innovations in the field combined with the ever-increasing availability of relatively low-cost computational power. Taken together, this prompts the question of whether it would be possible to design, build, and test a custom-built photogrammetry-based setup for 3D facial capture. The current study aimed to assess and report the accuracy of this technology.

MATERIAL AND METHODS

The Stereo-face3D (SF3D) system, custom-built by the first author (H.L.L.W.), consists of 14 Canon EOS 1200D digital single-lens reflex cameras with Canon 18-55 mm EF-S lenses (Canon, Tokyo, Japan) mounted on a square aluminum frame (measuring 1 × 1 m), assembled from industry-standard system profiles (45 × 45 mm cross-section with a 10 mm slot) (Motedis, Ensdorf, Germany) (Fig 1). The frame is attached to a similarly constructed, wheel-mounted support assembly (dimensions, 1.12 m [width] × 1.7 m [height] × 0.7 m [length]), which provides a working surface for the control switches and laptop, and houses the power supplies, universal serial bus (USB) hubs and electronics (Figs 1 and 2). Two height-adjustable Bosch-Rexroth lifts (Bosch Rexroth, Lohr am Main, Germany) allow the frame's height to be adjusted over a distance of 0.4 m.

The cameras are mounted in a hemispherical arrangement to accommodate the human facial form better (Fig 1, A). This arrangement is accomplished both by applying a mild inset to the 4 central cameras (ie, positioning them slightly out of the plane and further away from the patient) (Fig 2; blue arrows), as well as by positioning the outermost cameras in a more forward manner as a result of their inward rotation around the 28-mm round aluminum struts they are connected to (Fig 1, A). Adjustable camera mounts allow for precise control over the camera positions and angles (Multi-mount 6, Vanguard, Guangdong, China).

Aside from the cameras, the frame also supports 3 remote-controlled, high CRI, 5600K light-emitting diode (LED) panels (Godox, Shenzhen, China), which provide uniform, shadow-free illumination. These are located on the frame's upper left and right corners

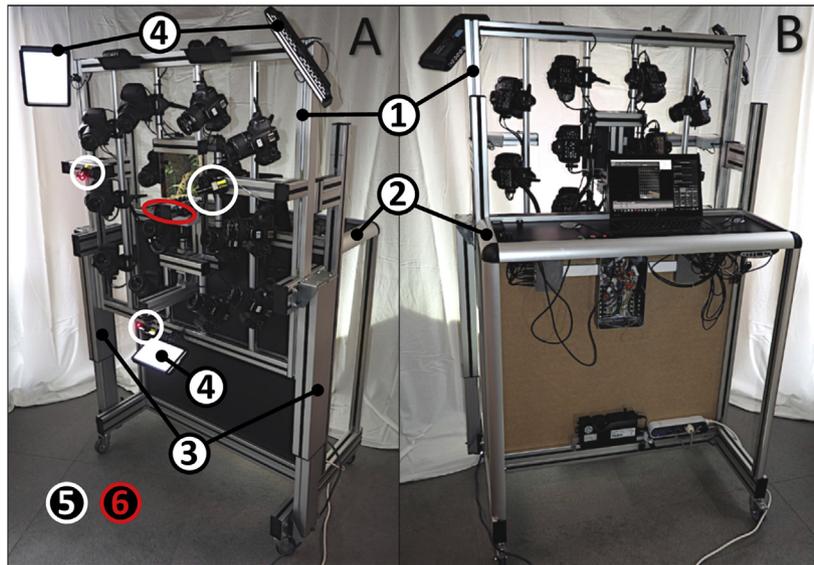


Fig 1. A, Frontal-oblique (patient side); B, rear view (operator side) of the custom-built StereoFace 3D system: 1, camera mounting frame; 2, support frame; 3, Bosch-Rexroth lift modules; 4, LED panels; 5, random pattern lasers; 6, positioning lasers.

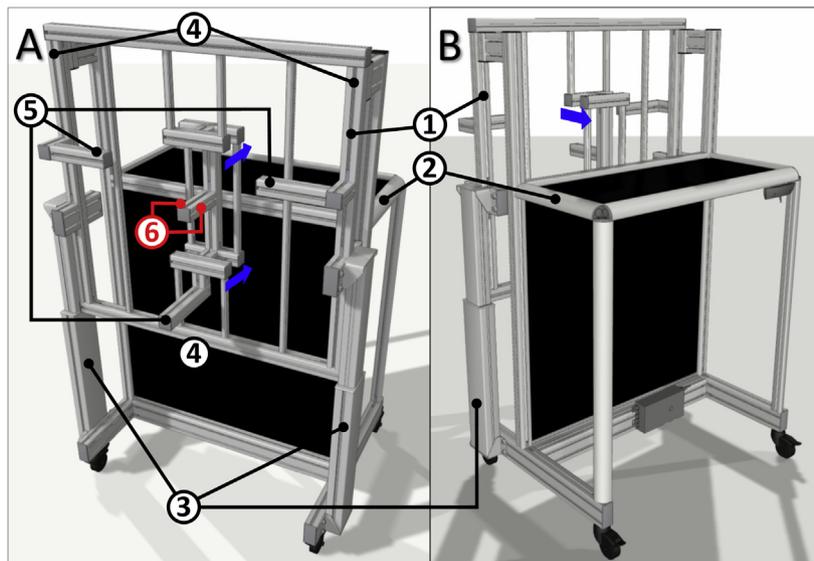


Fig 2. Computer aided design renderings of the StereoFace 3D system (orientated similarly to the views in Fig 1) showing the camera mounting frame as well as the wheel-mounted, height-adjustable support structure. 1, camera mounting frame; 2, support frame; 3, Bosch-Rexroth lift modules; 4, LED panels; 5, random pattern lasers; 6, positioning lasers. The blue arrows indicate the slight inset applied to the 4 central cameras.

(2 × Godox LED500LW) and on its lower-middle section (1 × Godox LED308W) (Figs 1, A and 2). These panels are low-weight, dimmable, flicker-free, and do not generate heat. In addition, the panels are powered by a separate Meanwell HRPG-150-15 enclosed power

supply unit (15 V, 10 A) (MeanWell, Guangzhou, China). Patient positioning is facilitated by 2 diode line-lasers (Picotronic, Koblenz, Germany) on either side of a 20 × 20 cm mirror, positioned approximately in the center of the frame (Figs 1, A and 2). Furthermore, 3 eye-

safe 660-nm random pattern lasers (dot) provide additional texture to the relatively feature-sparse human face (SL-660-S-C; Osela, Lachine, Canada), mounted slightly obliquely from the upper left and right corners, and centrally from below (Figs 1, A and 2).

Removing cameras from the setup to replace depleted batteries is highly undesirable in a carefully calibrated setup. This issue is avoided by using Canon DC-10 DC couplers (Canon, Tokyo, Japan), which, in turn, are fed directly from a power supply (SP-320-7.5 V power supply unit [40 A]; Mean Well, Guangzhou, China). Apart from the power cable, each camera requires 1 USB cable for image transfer and one 2.5-mm jack cable for camera focusing and triggering. The USB cables are connected to two 7-port industrial USB hubs, which are powered by the same power supply unit. The same power supply also feeds the electronics (after appropriate downregulation of the voltage), which consists of an Arduino Uno microcontroller (<https://www.arduino.cc/>) along with pushbuttons and relays for controlling the positioning of cameras and random pattern lasers, and 8-bit shift registers combined with optocouplers for focusing and triggering the cameras (Fig 1, B). The electronic components were soldered on 3 Adafruit Perma-Proto full-size printed circuit boards (Adafruit, New York).

Camera settings and the image retrieval process are controlled using Smart Shooter 3 GRID software (<https://kuvacode.com/smart-shooter>), whereas the imported images are reconstructed using 3DFlow's 3DF Zephyr PRO (multibase) stereophotogrammetry software (<https://www.3dflow.net/3df-zephyr-pro-3d-models-from-photos/>).

Preparing for image capture typically involves removing the lens covers from the cameras, powering up the setup and laptop, flipping the camera reset switch on the working surface (thus, providing power to the cameras too). The LED panels light up when the setup is powered on, after which their brightness can be adjusted using a remote control. The whole startup takes between 1 and 2 minutes and does not have to be repeated when capturing multiple subjects sequentially.

After seating the subject centered in front of the setup, a headband displaying 4 machine-vision markers is loosely fitted. This headband, which serves to scale the reconstructed facial mesh to life-size, is positioned such that as much of the forehead as possible remains exposed, while also ensuring that the markers are visible to a sufficient number of cameras (at least 3, but preferably more). Any loose hair is tucked away behind it in the process. The positioning lasers are then activated, and the subject is instructed to look into the centrally placed mirror with the nose tilted slightly upwards. The frame's



Fig 3. Mannequin head showing the 2 eye-safe positioning lasers projecting 1 cross each on the patient's face. The optimal working distance is obtained when both crosses coincide on the patients' midline, just below the nose.

height and anteroposterior position are subsequently adjusted to align the projections of both laser crosses on the subject's facial midline subnasally, which provides an easy visual cue for both patient and practitioner to confirm proper positioning (Fig 3). Upon instructing the subject to maintain a neutral (ie, relaxed) facial expression, the image acquisition button is pressed, after which all cameras automatically and simultaneously focus and trigger. An example of the images acquired by each camera is presented in Figure 4. The process is then continued on a computer fitted with a sufficiently powerful Nvidia graphics card (Nvidia, Santa Clara, Calif). After loading the images into 3DF Zephyr Pro, the reconstruction into a 3D mesh takes about 5–15 minutes to complete, depending on the desired mesh resolution and system specifications (at which point the presence of the imaging subject is no longer required).

Assessing the accuracy of the SF3D setup ideally required both intra- and intersystem evaluations, for which we had access to both the frequently used 3dMDFace and Vectra H1 systems. Because the latter requires 3 acquisitions from different angles to perform 1 facial reconstruction, whereas both SF3D and 3dMDFace



Fig 4. Example of the 14 camera views, as acquired from the mannequin head. The red dots are generated by the random pattern lasers, which provide additional texture to the relatively feature-sparse human facial surface.

are used as stationary, single-shot systems, we opted to use the 3dMDFace system for comparison. In brief, the 3dMD system is a hybrid structured-light stereophotogrammetry system consisting of 3 (stereo) pairs of 2 cameras each, with 1 pair positioned centrally in front of the patient, and the other 2 placed on either side.

A study group of 30 volunteers of diverse ethnicity was recruited from the Medical Imaging Research Lab at the University Hospital Gasthuisberg in Leuven,

Belgium, using the exclusion criterion of having undergone any facial surgical interventions and having dense facial hair such as mustache and/or beard. The age and sex distribution of the sample, calculated using Microsoft Office Excel (Microsoft, Redmond, Wash), are shown in [Table 1](#). To account for the highly variable nature of human facial expression, we repeated image acquisitions 3 times in a row for each individual, using the methodology presented earlier. Furthermore, the technical

Table 1. Age and sex distribution of the study group

Ser	n	Age, years			
		Mean	SD	Min	Max
Total	29	28.3	3.8	21	38
Male	13	30	3.9	24	38
Female	16	27	3.3	21	34

SD, standard deviation; Min, minimum; Max, maximum.

baseline performance of both systems, defined as the performance in the absence of biologic variability (ie, facial pose), was assessed and compared by scanning a mannequin head 3 times consecutively with each system.

All acquisitions were mapped using MeshMonk,¹⁹ an open-source software toolbox that allows for spatially dense (ie, high-resolution) registration of 3D surfaces. MeshMonk nonrigidly (ie, viscoelastically) maps an anthropometric mask (ie, landmark template) onto the previously generated 3D facial surfaces; thus, it establishes high-resolution configurations of quasi-landmarks, which are homologous across subjects.^{18,20,21} Homologous, in this context, refers to the position of each quasi-landmark relative to all other quasi-landmarks being identical in all individuals.²¹ In contrast to similar studies focusing on smaller sets of manually placed landmarks, information is provided on the entire facial surface, including regions that are rarely evaluated, like the eyelids and nostrils.

The variation both within and between systems under investigation was calculated by superimposing the MeshMonk-generated, mapped facial meshes using a generalized Procrustes analysis without scaling.²² Therefore, the meshes were translated to the origin and rotated to minimize the squared distance between the corresponding (quasi) landmarks, but not scaled to centroid size, effectively superimposing them in size-and-shape space. Furthermore, each subject's true facial shape (ie, the ground truth) was established by calculating the mean of the 3 generalized Procrustes analysis-superimposed acquisitions. The Procrustes distance to the ground truth (ie, the Euclidean distance between 2 corresponding landmark configurations of Procrustes coordinates) was then used to quantify shape variation, reported as the root-mean-square error (RMSE). This approach was applied both to the test subjects' mapped reconstructions, as well as those of the mannequin head. A schematic representation of the acquisitions is shown in Figure 5.

System validation then proceeded by calculating and comparing the precision of 3dMD and SF3D, defined here as the mean difference in RMSE between repeated

measures of the same subject, comparable with the approach reported in the article by Aldridge et al.²³ For intrasystem validation, the 3 acquisitions of each subject, as well as the mannequin head were compared with the corresponding averaged shape by calculating the RMSE values. Intersystem validation involved comparing the average shape of each set of 3 acquisitions per subject across both systems by calculating their RMSE values, as well as the normal distances between these average shapes. The 2 average shapes of the mannequin head generated using both systems were similarly compared. Furthermore, a 2-factor Procrustes analysis of variance (ANOVA)²⁴ was used to partition the intersystem shape variation to understand better the factors of the discrepancies between the facial shapes acquired with both photogrammetry systems. This approach is comparable with the assessment of system error in the study by Aldridge et al.²³ (defined here as the proportion of total variance attributable to a particular factor). For a detailed explanation of the use and interpretation of 2-factor ANOVA, the reader is referred to the pertaining studies by Claes et al.^{25,26}

RESULTS

Figure 6 provides an example of surface reconstructions of the mannequin head, as acquired by the 3dMD system (Fig 6, left) and the SF3D setup (Fig 6, right). Facial meshes from the latter system on average contained approximately 350,000 vertices; those originating from the 3dMDFace system had approximately 34,000. Because a headband with machine-vision markers was used in the SF3D system, the amount of forehead area included in the analysis had to be reduced. To effectively compare the facial surfaces generated by both systems, we used the facial area depicted in red in Figure 7 for all mappings of the reconstructed meshes. One subject's set of SF3D generated facial meshes exhibited very strong artifacts in the forehead region upon mapping. On closer inspection of the accompanying photographs, this was attributed to the headband being positioned too low in combination with hair sticking out from under it. Therefore, that subject was removed from further analysis.

The intrasystem variation, quantified as the RMSE between each repeated acquisition and reconstruction and the corresponding average shape per subject, is shown in Figure 8. The unfilled boxplots and compact boxplots represent the 3dMDFace and SF3D system, respectively. The study group's mean RMSE value was 0.52 mm (standard deviation [SD] = 0.07) when imaged by 3dMDFace and 0.44 mm (SD = 0.16) for SF3D. The intrasystem RMSE over the 3 acquisitions was

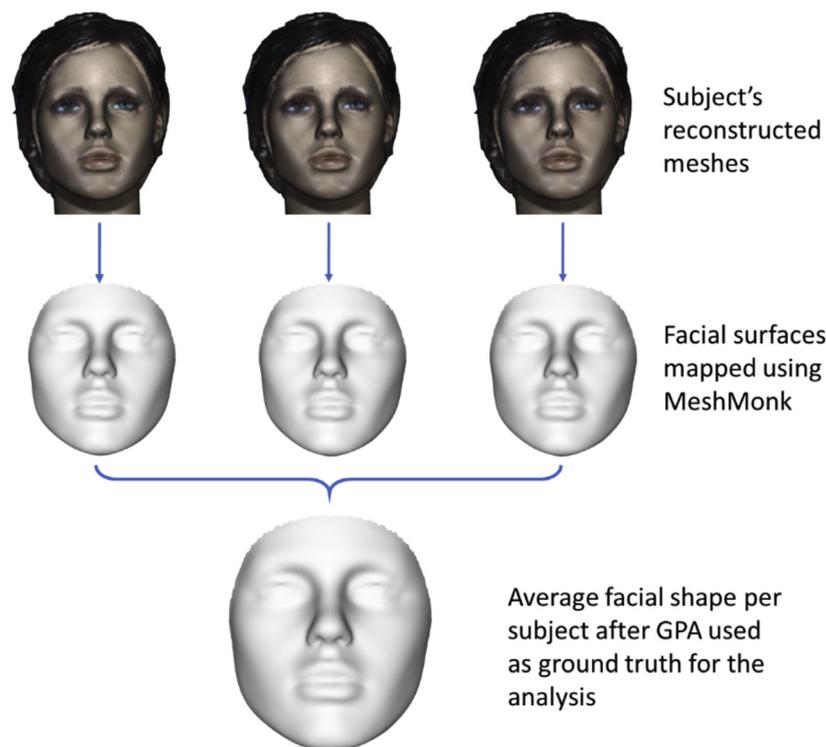


Fig 5. Schematic representation of the acquisitions performed by each system. Each subject was scanned 3 times, after which the images were registered using MeshMonk. The mapped results were then superimposed using a generalized Procrustes superimposition without scaling. The resulting average shape was considered to be the ground truth facial shape per subject. GPA, generalized Procrustes analysis.

nonsignificant ($P > 0.05$; 1-way ANOVA). However, the mean difference in RMSE between both systems was highly significant ($P < 0.001$; Kruskal-Wallis rank sum test). When imaging the mannequin head, 3dMDFace's RMSE was found to be 0.42 mm (SD = 0.02), whereas the corresponding value for SF3D was 0.29 mm (SD = 0.03).

Intersystem variation, computed as the RMSE between the average shape of each subject across both systems, is presented in Figure 9. The average RMSE of the study group was 1.6 mm (SD = 0.34), whereas the corresponding value was 1.38 mm for the mannequin head. To understand better the location and direction of these differences, a heatmap of the normal distances between the 2 systems' overall mean shape (ie, that of the entire study group) was plotted in Figure 10, A. The corresponding results for the mannequin head are depicted in Figure 10, B. Further partitioning of the shape differences between the study group's mean shapes as generated by 3dMDFace and SF3D was performed with a 2-factor ANOVA, as shown in Figure 11. All values are normalized per column.

All the RMSE means and standard deviations of the inter- and intrasystem variation for subjects and the mannequin head are listed in Table II.

When partitioning the facial shape differences between both systems into individual variability vs the system using the 2-factor ANOVA, the intersubject variations were mainly located at the nose, the lower-posterior part of the cheeks, and to a lesser extent, the chin (Fig 11, upper row). The choice of system was mainly expressed at the level of the upper eyelids, nasal tip and alae, and chin area (Fig 11, second row). The very low random error found (Fig 11, fourth row) confirmed that the main sources of variation indeed were individual differences in facial morphology, as well as the system used.

DISCUSSION

In this study, we presented a custom-built stereophotogrammetry solution for 3D facial capture and assessed its performance in vitro (on a mannequin head) and in vivo (in the study group) both within and between systems (comparing to the commercially available 3dMDFace system).

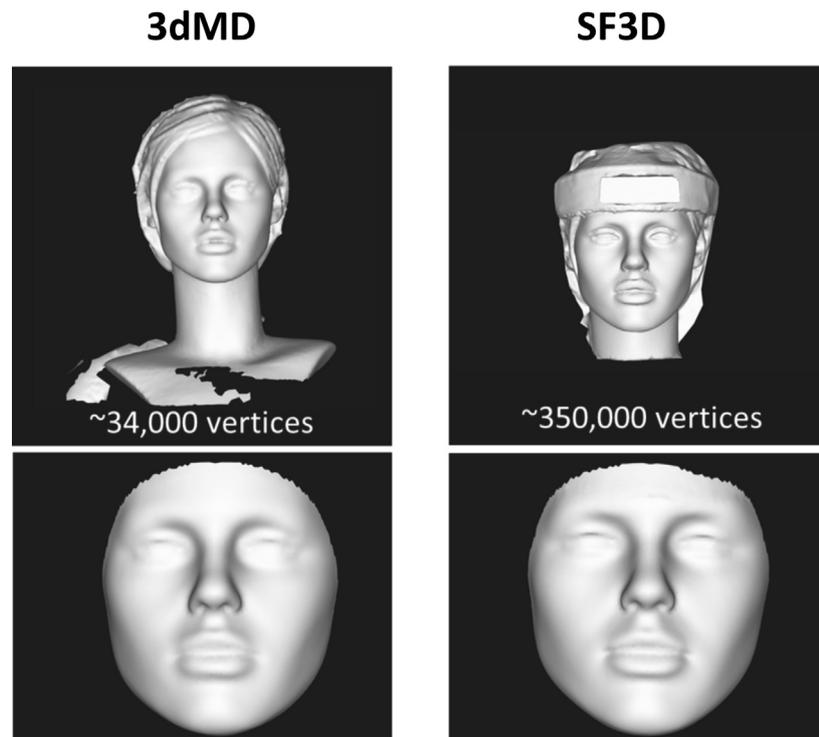


Fig 6. The top panel depicts the original, reconstructed meshes of the mannequin head, as acquired by the 3dMDFace system (*left*) and the SF3D system (*right*). The bottom pane displays the corresponding registered results using MeshMonk.

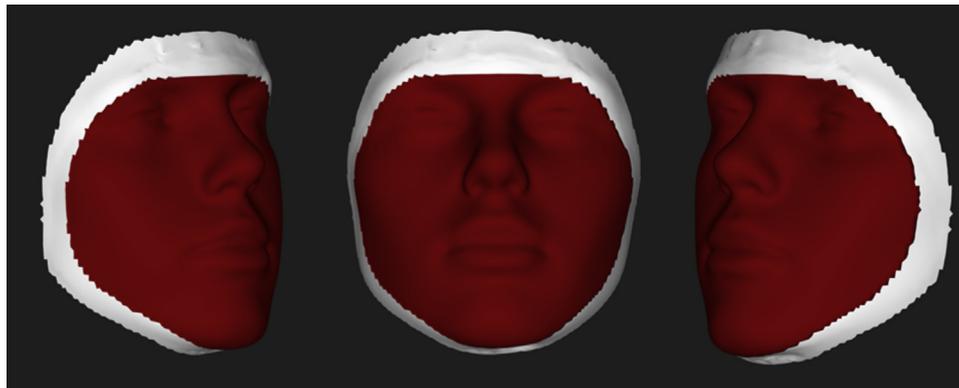


Fig 7. Superimposition area (in red) overlaid on the mapped mannequin head. Only the region colored red was considered in the current investigation.

When visually comparing the custom-built SF3D system to the 3dMDFace system (Fig 6, *top left*, and *right*), it is evident that the higher vertex count provided by SF3D 350,000 on average compared with 34,000 by 3dMDFace—resulted in a more detailed facial mesh. Given that the 3dMDFace system covers not only the face but also the shoulder region, its actual facial vertex count is even lower. The resulting difference in detail is

most obvious around the mouth, nose, and eyes. In contrast, the forehead was much better represented in the 3dMDFace meshes. This finding was a consequence of difficulties in recovering the position of the machine-vision markers in the SF3D generated images using the 3DZephyr software package. Although this problem has since been resolved, for this study, the headband required for scaling the SF3D generated meshes had to

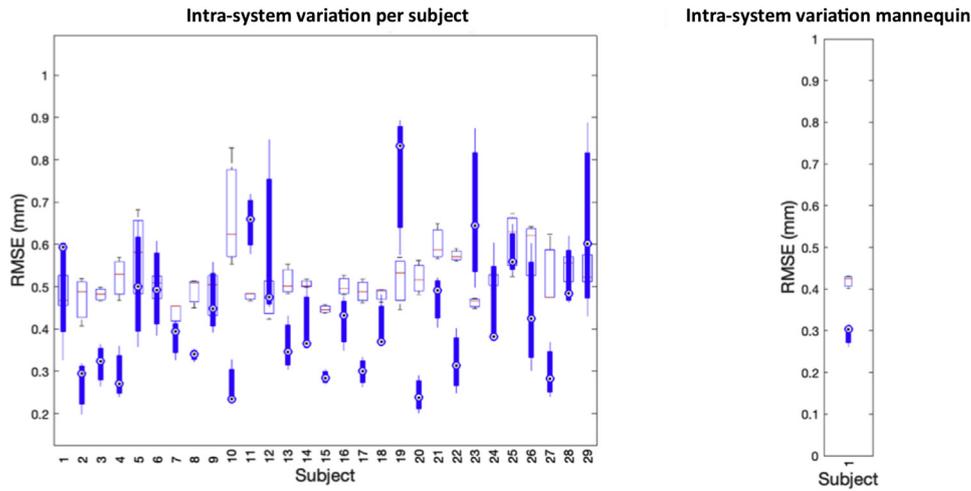


Fig 8. Intrasystem variation expressed in RMSE. The unfilled boxplots represent the 3dMDFace system, whereas the compact boxplots represent the SF3D setup. *Left*, RMSE of the study group with a mean RMSE of 0.41 (SD = 0.15) mm for SF3D and a mean of 0.51 (SD = 0.07) mm for 3dMDFace. *Right*, RMSE values of the mannequin head, with a mean of 0.29 (SD 0.03) mm for SF3D and 0.42 (SD 0.01) mm for 3dMDFace. The difference between means of the study group and the mannequin can be attributed to biologic error (ie, facial pose instability) and/or registration error.

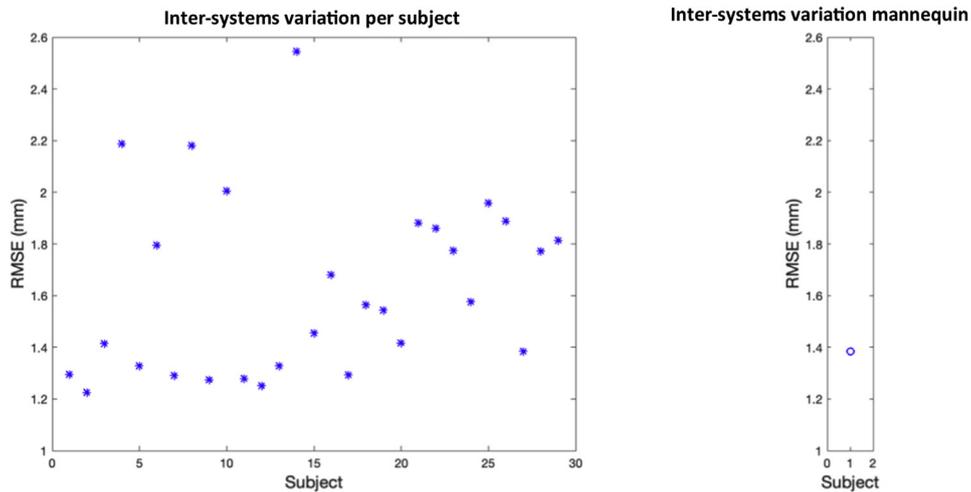


Fig 9. Intersystem variation expressed in RMSE, with a mean of 1.63 ± 0.34 mm for the study group (*left*) and 1.38 ± 0.25 mm for the mannequin head (*right*).

be positioned fairly low, partly obscuring the forehead (Fig 6, upper right image).

On closer scrutiny of the images mapped by Mesh-Monk, the differences in detail between both systems become less obvious (Fig 6, lower panes). This finding is attributable to the anthropometric mask’s resolution, which currently contains about 7,160 vertices. Although this vertex count makes perfect sense based upon the tradeoff between resolution and computational load, the resulting mappings do present a severe down-

sampling of the high-resolution meshes generated by the SF3D system. As a result, overall facial shape is well preserved by the mapped representations. The details of the surfaces acquired with both facial scanners, and the SF3D system, in particular, look significantly smoothed out after mapping.

For the intrasystem variation, the SF3D setup exhibited lower RMSE, but higher variability values for both the study group (0.6 mm lower RMSE and 0.1 mm higher SD; Table II) and the mannequin head

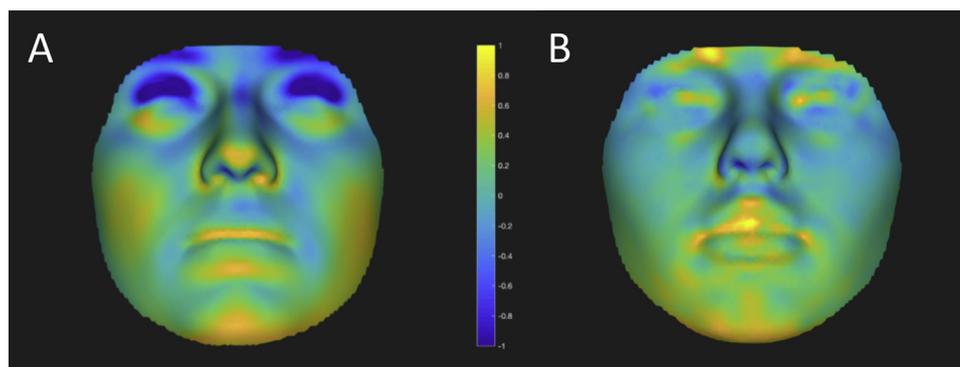


Fig 10. **A**, Heatmap depicting the normal distances between the average facial shapes of the study group (**A**) as generated by 3dMDFace and SF3D, showing the intersystem variation. Yellow colors indicate 3dMDFace positioned the mesh outwards compared with SF3D, whereas blue suggests 3dMDFace's mesh is positioned inward relative to that of SF3D. **B**, the corresponding heatmap depicting the normal distances between the average facial shapes of the mannequin head (**B**) as generated by 3dMDFace and SF3D (intersystem variation).

(0.12 mm lower RMSE and a 0.01 mm higher SD; [Table II](#)) compared with 3dMDFace. Both systems' lower RMSE values when acquiring the mannequin head were not unexpected because of the absence of biologic variability associated with changes in facial pose. Interestingly, when comparing the differences in RMSE means and SDs within systems, the influence of facial pose variability was found to be greater for the SF3D (mean RMSE difference of study subjects vs mannequin head: 0.160 [SD = 0.136] mm vs 0.098 [SD = 0.051] mm for SF3D and 3dMDFace, respectively; [Table II](#)). This finding might tentatively be attributed to the increased resolution of the SF3D system, allowing for smaller variations in facial morphology to be picked up, and thus, explaining why the intrasystem RMSE SD values between both systems differed by only 0.013 mm for the mannequin head; whereas there was an almost 8-fold higher value for the study group (0.1 mm; [Table II](#)).

Similarly, for the intersystem variability, a lower mean RMSE value was found for the mannequin head compared with the study group, corroborating the influence of the biologic error when comparing facial surface meshes. Interestingly, the study group's heatmap of normal distances indicated that the 3dMDFace system consistently positioned the upper eyelids slightly more posteriorly and the lower eyelids slightly anteriorly compared with the SF3D system ([Fig 10, A](#)). Similarly, the nasal tip and alae, lips, lower-posterior part of the cheeks, and chin were also positioned slightly more anteriorly by the 3dMDFace system compared with the SF3D ([Fig 10, A](#)).

The results of this study indicate that it is possible to design and build a multibase stereophotogrammetry setup, which rivals and, in some aspects, surpasses commercially available solutions for 3D facial capture, albeit at considerable effort. Overall, the errors reported for both systems under investigation (introduced by facial pose variability and/or system choice) were found to be quite small and well within the accepted limits of the geometric morphometric community. Therefore, the basic question amounts to whether the aim (high-quality 3D facial capture at reduced cost) justifies the means (the time and effort invested in acquiring all system components and putting them together). For the final build of the SF3D system, we spent about \$8900 on materials and \$4900 on software, totaling \$13,800 without value-added tax. This finding would appear to compare favorably to the 3dMDFace system, which, as configured in the present study, amounted to more than \$50,000 (in 2013), including delivery, installation, training, and software (but not including annual maintenance fees). Although, the amount mentioned above for SF3D does not include the cost of labor for assembly, anyone aiming to replicate this project would benefit from not having to reinvent the wheel, at considerable savings in both cost and time.

Furthermore, there is significant leeway in terms of how good is good enough concerning the resolution obtained by the SF3D system. It may very well be sufficient to reduce the number of cameras in the currently proposed system from 14 to 10, equally lowering the number of required direct current couplers, mounts, cables, and electronics. Although this would reconstruct the

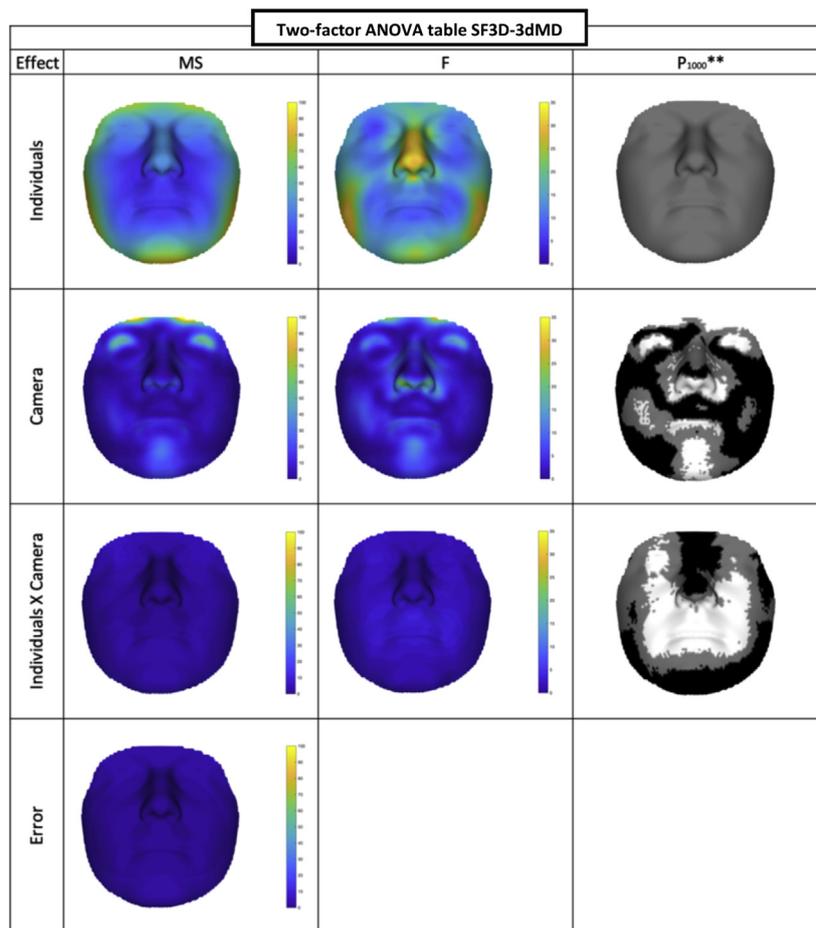


Fig 11. Two-factor ANOVA intersystem shape variation after an isotropic model. *P*-values using 100 permutations: white *P* < 0.001; gray *P* < 0.05; black *P* ≥ 0.05 ns. MS (mean square) is the sum of squares divided by the appropriate degrees of freedom, reflecting the magnitude value. F (F-ratio) is the MS divided by an appropriate error MS, reflecting the relative magnitude or strength of the effect. The interaction term is used as an error term for the main effects of individuals and sides, whereas the actual error term is used for the interaction term.

Table II. Intrasystem and intersystem shape variability expressed as RMSE

Intrasystem variability	Study group				Mannequin head			
	Mean	SD	Min	Max	Mean	SD	Min	Max
RMSE, mm								
3dMD	0.517	0.068	0.407	0.828	0.419	0.016	0.400	0.432
SF3D	0.437	0.164	0.198	0.895	0.295	0.029	0.262	0.317
Diff	0.08	-0.096			0.124	-0.013		
Intersystem shape variability	Mean	SD	Min	Max				
RMSE, mm								
Study group	1.630	0.342	1.225	2.546				
Mannequin head	1.384							
Diff	0.246							

SD, standard deviation; *Min*, minimum; *Max*, maximum; *Diff*, difference.

face to a somewhat lower resolution compared with SF3D, it would still be considerably higher than what can be obtained using commercially available alternatives. Further cost-reductions might be obtained by omitting the random pattern lasers, although we found this to be quite detrimental to the received mesh quality, mostly in prepubertal children with very feature-sparse faces (eg, in the absence of freckles, pimples, wrinkles, etc.).

From a practical point of view, it might be useful to consider between-system differences in image capture and processing speed, storage, and calibration requirements as well as operating modes. The capture speed of 3DMDFace is 1.5 milliseconds⁵ compared with 6.3 milliseconds for SF3D, both of which are adequate to avoid movement artifacts. Comparing processing times (ie, the time required to generate a mesh from the acquired images) is not straightforward because this depends heavily on the selected software settings and file types. It took 3D zephyr anywhere from 5–15 minutes to generate a mesh, compared with <10 seconds for 3DMDFace.⁵ However, algorithmic optimizations in more recent updates of 3DF Zephyr have dramatically improved processing speeds. Using the software settings we employed, OBJ files of 54 MB on average were generated (range, 45–60 MB) compared with 4–26 MB for 3DMDFace.⁵ Our reported files sizes pertain to the unprocessed meshes (ie, without removing superfluous information such as the headband, hair and/or whiskers, part of the ears, ...). Although an increased file size is not unexpected because of the higher vertex count we obtained (see below), storage requirements can be halved by saving the processed meshes instead.

Calibrating 3DMDFace takes approximately 20 seconds. No separate calibration was performed for SF3D, as the system is recalibrated for each acquisition as part of the reconstruction (ie, targetless calibration).^{27,28} Each camera's internal parameters (ie, principal distance, principal point offset, radial, and decentering distortion) as well as the external ones (the camera's position and orientation relative to the subject) are (re)calculated based upon the principles of structure from motion combined with bundle adjustments.^{27,28} This was required because we opted to enable autofocus to maximize image quality (while keeping focal length constant). Intriguingly, targetless calibration seems to be able to yield camera calibration parameters of greater precision and equal accuracy compared with targeted self-calibration, probably because of the high number of image correspondences used for the calibration in the structure from motion algorithm.²⁸

Whereas 3DMDFace is capable of recording 3D movies, the SF3D system currently only supports static

image capture. Although modern digital single-lens reflex or mirrorless cameras allow for relatively fast sequential image capture (burst mode), it is highly doubtful all cameras would remain sufficiently well synchronized using this functionality. Synchronizing high-speed image capture over multiple cameras is specific to (expensive) machine-vision cameras, but also comes at a cost in terms of image capture resolution; the speed of image capture is usually inversely correlated with sensor resolution. Because of the higher number of relatively high-resolution cameras in SF3D (14 cameras of 8 megapixels each, totaling 112 megapixels), the average sample density was approximately 500 vertices per cm, compared with 62 vertices/cm² for 3DMDFace.⁵

A possibly more fundamental question is what resolution we really need for 3D facial capture in an orthodontic/craniofacial surgical context. The results of the current study suggest the SF3D system is slightly more precise while also slightly more variable compared with the 3DMDFace system (Table II), both of which are probably attributable to the increased resolution of the former compared with the latter system. We would argue the increased resolution makes sense if the focus of research or diagnosis is aimed toward (the effects of treatment on) the chin, lips, and paranasal areas because these regions exhibited the largest between-systems differences (apart from the eyelids). This finding might particularly be the case for orthodontic treatment involving extractions or functional appliances, craniofacial surgical patients, and cleft lip and palate patients if the number of patients is limited such that computational time is less important. Studies involving higher resolution meshes such as those generated by the SF3D system should probably use an anthropometric mesh with a higher vertex count, to be able to maintain most of the original meshes' detail in the mapped versions thereof. Furthermore, the increased variability associated with the use of higher resolution mesh representations of the human face may be countered by averaging multiple 3D captures of the same patients, thus preserving detail while improving reliability.

Aside from overall cost and algorithmic complexity, another major impediment to the democratization of stereophotogrammetry has traditionally been the absence of (freely) available software tools for comprehensive facial mesh analysis. Although the reconstruction software provided with the various 3D scanners provides some analytical tools, their utility is usually limited to performing iterative closest point algorithms and deriving distance heatmaps. Because of the introduction of sliding semilandmarks^{29,30} and viscoelastic mapping^{18,20,21} to the geometric morphometric toolbox, multiple software packages such as Geomorph,³¹

Morpho,³² mesher (<https://www.github.com/zarquon42b/mesher>), and Meshmonk¹⁹ have been made available free of charge. Their application to the generated datasets does require some (mild) programming effort. Alternatively, Halazonetis' Viewbox software (<http://www.dhal.com/>) also has some built-in functions for sliding semilandmark analysis of facial meshes.

With regard to the limitations of this study, it needs to be pointed out that only adults were included in the study group. The variability in facial expression we report is therefore expected to be smaller compared with studies including babies and young children.³³

CONCLUSIONS

- (1) The StereoFace3D system was slightly more precise, while also slightly more variable, compared with the 3dMDFace system, both of which are probably attributable to the higher resolution of the StereoFace3D system.
- (2) The errors reported for both systems under investigation (introduced by facial pose variability and/or system choice) were found to be small and well within the accepted limits of the geometric morphometric community.
- (3) The performance of the custom-build stereophotogrammetry setup equaled and in some aspects even surpassed, commercially available solutions for 3-dimensional facial capture.

SUPPLEMENTARY DATA

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.ajodo.2020.01.016>.

REFERENCES

1. Fourie Z, Damstra J, Gerrits PO, Ren Y. Accuracy and repeatability of anthropometric facial measurements using cone beam computed tomography. *Cleft Palate Craniofac J* 2011;48:623-30.
2. Radiation Protection No 172, Cone Beam CT for Dental and Maxillofacial Radiology, European Commission 2012. Available at: https://ec.europa.eu/energy/sites/ener/files/documents/172_0.pdf. Accessed April 29 2019.
3. Nederlandse vereniging van orthodontisten [Concept for Radiological Guidelines in Orthodontics]. Richtlijn Orthod Röntgen; 2018. Available at: https://www.knmt.nl/sites/default/files/conceptrichtlijn_orthodontie_rontgen.pdf. Accessed April 29 2019.
4. Halazonetis DJ. Cone-beam computed tomography is not the imaging technique of choice for comprehensive orthodontic assessment. *Am J Orthod Dentofacial Orthop* 2012;141:407.
5. Tzou CH, Artner NM, Pona I, Hold A, Placheta E, Kropatsch WG, et al. Comparison of three-dimensional surface-imaging systems. *J Plast Reconstr Aesthet Surg* 2014;67:489-97.
6. de Menezes M, Rosati R, Ferrario VF, Sforza C. Accuracy and reproducibility of a 3-dimensional stereophotogrammetric imaging system. *J Oral Maxillofac Surg* 2010;68:2129-35.
7. Othman SA, Ahmad R, Mericant AF, Jamaludin M. Reproducibility of facial soft tissue landmarks on facial images captured on a 3D camera. *Aust Orthod J* 2013;29:58-65.
8. Metzler P, Sun Y, Zemann W, Bartella A, Lehner M, Obwegeser JA, et al. Validity of the 3D VECTRA photogrammetric surface imaging system for cranio-maxillofacial anthropometric measurements. *Oral Maxillofac Surg* 2014;18:297-304.
9. Winder RJ, Darvann TA, McKnight W, Magee JDM, Ramsay-Baggs P. Technical validation of the Di3D stereophotogrammetry surface imaging system. *Br J Oral Maxillofac Surg* 2008;46:33-7.
10. Khambay B, Nair N, Bell A, Miller J, Bowman A, Ayoub AF. Validation and reproducibility of a high-resolution three-dimensional facial imaging system. *Br J Oral Maxillofac Surg* 2008;46:27-32.
11. Weinberg SM, Scott NM, Neiswanger K, Brandon CA, Marazita ML. Digital three-dimensional photogrammetry: evaluation of anthropometric precision and accuracy using a Genex 3D camera system. *Cleft Palate Craniofac J* 2004;41:507-18.
12. Weinberg SM, Naidoo S, Govier DP, Martin RA, Kane AA, Marazita ML. Anthropometric precision and accuracy of digital three-dimensional photogrammetry: comparing the Genex and 3dMD imaging systems with one another and with direct anthropometry. *J Craniofac Surg* 2006;17:477-83.
13. Lübbers HT, Medinger L, Kruse A, Grätz KW, Matthews F. Precision and accuracy of the 3dMD photogrammetric system in craniomaxillofacial application. *J Craniofac Surg* 2010;21:763-7.
14. Secher JJ, Darvann TA, Pinholt EM. Accuracy and reproducibility of the David SLS-2 scanner in three-dimensional facial imaging. *J Craniomaxillofac Surg* 2017;45:1662-70.
15. Ritschl LM, Roth M, Fichter AM, Mittermeier F, Kuschel B, Wolff KD, et al. The possibilities of a portable low-budget three-dimensional stereophotogrammetry system in neonates: a prospective growth analysis and analysis of accuracy. *Head Face Med* 2018;14:11.
16. Liu C, Artopoulos A. Validation of a low-cost portable 3-dimensional face scanner. *Imaging Sci Dent* 2019;49:35-43.
17. Maués CPR, Casagrande MVS, Almeida RCC, Almeida MAO, Carvalho FAR. Three-dimensional surface models of the facial soft tissues acquired with a low-cost scanner. *Int J Oral Maxillofac Surg* 2018;47:1219-25.
18. Claes P, Walters M, Clement J. Improved facial outcome assessment using a 3D anthropometric mask. *Int J Oral Maxillofac Surg* 2012;41:324-30.
19. White JD, Ortega-Castrillón A, Matthews H, Zaidi AA, Ekrami O, Snyders J, et al. MeshMonk: open-source large-scale intensive 3D phenotyping. *Sci Rep* 2019;9:6085.
20. Claes P. A robust statistical surface registration framework using implicit function representations-application in craniofacial reconstruction: Available at: <https://lirias.kuleuven.be/1670551?limo=0>; 2007. Accessed March 30 2019.
21. Claes P, Walters M, Vandermeylen D, Clement JG. Spatially dense 3D facial asymmetry assessment in both typical and disordered growth. *J Anat* 2011;219:444-55.
22. Dryden IL, Mardia KV. *Statistical shape analysis*. Chichester: Wiley; 1998.
23. Aldridge K, Boyadjiev SA, Capone GT, DeLeon VB, Richtsmeier JT. Precision and error of three-dimensional phenotypic measures acquired from 3dMD photogrammetric images. *Am J Med Genet A* 2005;138A:247-53.

24. Klingenberg CP, Barluenga M, Meyer A. Shape analysis of symmetric structures: quantifying variation among individuals and asymmetry. *Evolution* 2002;56:1909–20.
25. Claes P, Walters M, Shriver MD, Puts D, Gibson G, Clement J, et al. Sexual dimorphism in multiple aspects of 3D facial symmetry and asymmetry defined by spatially dense geometric morphometrics. *J Anat* 2012;221:97–114.
26. Claes P, Reijnen J, Shriver MD, Snyders J, Suetens P, Nielandt J, et al. An investigation of matching symmetry in the human pinnae with possible implications for 3D ear recognition and sound localization. *J Anat* 2015;226:60–72.
27. Barazzetti L, Mussio L, Remondino F, Scaioni M. Targetless camera calibration. *Int Arch Photogramm Remote Sens Spatial Inf Sci* 2012;XXXVIII-5/W16:335–42.
28. Fraser CS, Stamatopoulos C. Automated target-free camera calibration. *Int Arch Photogramm Remote Sens Spatial Inf Sci* 2014; II-5:339–46.
29. Gunz P, Semilandmarks MP. A method for quantifying curves and surfaces. *Hystrix It J mamm* 2013;24:103–9.
30. Pomidor BJ, Makedonska J, Slice DE. A landmark-free method for three-dimensional shape analysis. *PLoS One* 2016;11: Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4783062/:e0150368>. Accessed March 30 2019.
31. Adams DC, Otárola-Castillo E. Geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods Ecol Evol* 2013;4:393–9.
32. Schlager S. Morpho and Rvcg—shape analysis in R: R-packages for geometric morphometrics, shape analysis and surface manipulations. In: *Statistical shape and deformation analysis*. Elsevier; 2017. p. 217–56.
33. TJJ Maal, Verhamme LM, Loon B van, Plooij JM, Rangel FA, Kho A, et al. Variation of the face in rest using 3D stereophotogrammetry. *Int J Oral Maxillofac Surg* 2011;40:1252–7.