Precision and accuracy of the 3dMD photogrammetric system in craniomaxillofacial application

Lübbers, H T; Medinger, L; Kruse, A; Grätz, K W; Matthews, F
Precision and accuracy of the 3dMDface™ photogrammetric system in cranio-maxillo-facial application

Lübbers, Heinz-Theo MD, DMD*†; Medinger, Laurent MD, DMD*; Kruse, Astrid MD, DMD*; Grätz, Klaus Wilhelm MD, DMD*; Matthews, Felix MD, MBA†
Address correspondence and reprint requests to Heinz-Theo Lübbers, MD, DMD, Clinic for Cranio-Maxillofacial Surgery, University Hospital of Zurich, Frauenklinikstrasse 24, CH-8091 Zurich, Switzerland; E-mail: heinz-theo.luebbers@usz.ch

Abstract

Background

In modern anthropometry of such complex structures as the face, 3D scanning techniques have become more and more common. Before establishing them as a golden standard, however, meticulous evaluation of their precision and accuracy under both ideal and clinical circumstances is essential. Potential sources of error need to be identified and addressed.

Materials and methods

Under ideal circumstances, a phantom is used to examine the precision and accuracy of the 3dMDface™ system. A clinical setting is simulated by varying different parameters like angle, distance, and system re-registration, as well as data evaluation under different levels of magnification.

Results

The handling of the system was unproblematic in matters of data acquisition and data analysis. It was very reliable, with a mean global error of 0.2mm (range 0.1 – 0.5mm) for mannequin head measurements. Neither the position of the head nor of the camera influenced these parameters. New referencing of the system did not influence precision and accuracy.
**Conclusions**

The precision and accuracy of the tested system is more than sufficient for clinical needs and greater than that of other methods, such as direct anthropometry and 2d-photography. The evaluated system can be recommended for evaluation and documentation of the facial surface and could offer new opportunities in reconstructive, orthognatic, and craniofacial surgery.

**Keywords**

plastic surgery; maxillofacial surgery; craniofacial surgery; comparative study; phantom; imaging; photogrammetric
INTRODUCTION

Anthropometry, which was developed in the late 19th century, is the biological science of measuring the human body and its characteristics. (1) Though its applications are usually medical, today it also plays an important role in commercial settings, like clothing design, ergonomics, and architecture.

In cranio-maxillofacial and plastic surgery, anthropometry is especially challenging due to the complex structures of the face, which do not allow an assessment with simple measurements. For the underlying bony structures, the development of computer tomography (CT) by Hounsfield (2) and Ambrose (3) solved the difficulties. An objective, accurate, and reliable system for quantifying the soft tissues of the face in dimension and color is still needed.

Today, direct measurements and 2d photography are still state of the art for craniofacial anthropometry (4, 5), even though the pitfalls are well known and discussed. (1, 6-11) However, interest in overcoming the limitations of these techniques has led to the development of numerous 3D scanning devices which have an obvious appeal over the “old-fashioned” techniques. (12-20) Kau et al. give an overview of the 3d scanning device types available. (21)

Despite the huge amount of literature about the new 3D-systems, a clear and objective evaluation of accuracy and reliability under different circumstances is missing for many of them. Obviously, before any of these new techniques is applied clinically, it is crucial to evaluate their reliability. However, 3D representation of skull and soft tissue is a promising tool in orthognatic, craniofacial, and reconstructive surgery. In complicated cases, 3D stereolithography models are nowadays often necessary and their production is time- and cost-consuming.

Aim of the study
The aim of the study is to evaluate the precision (repeatability and reproducibility) and accuracy of the 3dMDface™ system. Therefore, a phantom was used, and various sources of error were examined. Furthermore, besides operator and capture errors, accuracy and bias in relation to direct anthropometry are evaluated.

Material and Methods

Model

For the experiments, a mannequin head was chosen as an ideal object because it does not move or perform facial expression. To minimize errors resulting from landmark identification, the mannequin head was prepared with 41 artificial landmarks as shown in figure 1. The labels were positioned to cover all face regions, with emphasis on the oral-nasal region.

Data acquisition

The direct line distances between the pre-labeled landmarks on the mannequin head were measured with standard clinical sliding and spreading calipers and measuring tape. Measurement was performed by three observers in one session at the same time and place that the images were captured by the 3dMDface™ system (Table 1, Study No. 1). Each observer measured the distances 5 times. The median of all 15 measurements for each distance was accepted as the real distance between the two labels.

The 3D data was acquired under clinical lighting using a 3dMDface™ System (3dMD Inc., Atlanta, GA, USA). The system is based on the combination of stereophotogrammetry and structured light, and is connected to a personal desktop computer where the captured data set is saved and calculated into a 3D VRML file (45,000 to 65,000 polygons) ready for evaluation. Data acquisition was performed in
natural head posture (NHP), with the Frankfurt horizontal line parallel to the floor and
with variations following the protocol given in table 1. If shown in the table, new
system calibration was performed before image capturing.

**Data processing**

Further data processing was performed on a standard desktop computer using
the 3dMD-Patient-Software (3dMD Inc., Atlanta, GA, USA) belonging to the capture
device. The labels were digitized on the surface of the 3D model and the x-, y- and z-
coordinates of this markings were exported to an Excel 2003 file (Microsoft
Corporation, Redmond, WA, USA) for further calculations. A zoom tool could be used
for magnification on the screen. Single coordinates were excluded when not being
captured because they were out of the field of vision due to rotation of the head.

**Operational definitions**

As the aim of this study was to validate how accurate the 3dMDface™ system is
compared to the “gold standard” of direct measurements, this standard is
operationally defined by accuracy, bias, and precision.

1. Accuracy is the agreement between a measurement and the “true” value of a
   parameter (22, 23)—in our case, the 3D model and the results of direct
   anthropometry.

2. Bias measures whether 3dMD tends to over- or underestimate direct values
   systematically.

3. Precision is divided into the following sub elements:

   a. Repeatability is the degree of similarity of multiple measurements of the
      same part using the same technique. This aspect has 3 subdivisions:
      i. “Operator error,” which results from inaccuracies during
         repeated digital measurements of the same 3D model
         derived once out of one dataset;
ii. “Capture error,” which results from a systems error when capturing the same object multiple times.

iii. Registration error, which is that added by new calibration of the system in between two captures.

b. Reproducibility is the magnitude of the differences between repeated measurements by different operators who are using the same technique. (22, 24)

The discrepancy, which is the distance between 2 landmark coordinates and is calculated as the square root of the sum of squared deviation in all 3 spatial directions is $\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$, an analog to the target registration error (TRE) described in different articles. (25-27)

Data analysis

To assess the above-mentioned parameters, two kinds of measurements were performed:

1. Point error measurement

   By means of a fusion analysis, the 3D coordinates of each landmark were aligned via translation and rotation to match the coordinates of the corresponding landmark on another model. The null hypothesis was that there is one translation matrix for all corresponding landmarks, leading to a perfect fit.

2. Distance error measurements

   Though 3D coordinates weren’t available for direct anthropometry, the distances directly measured on the mannequin head were compared to the corresponding distances calculated from the 3D coordinates of the VRLM models by the above-mentioned formula caliper distance, which equals
\[\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}\]. The null hypothesis was that corresponding distances are identical.

**Statistic tools**

The acquired data was analyzed using descriptive statistics as well as parametric student t-tests. The tests were performed with SPSS 11.5 (SPSS Inc, Chicago, IL, USA) and were considered significant if \(p<0.05\).

**Results**

**Operator error**

Operator error as an error resulting from inaccuracies in placing the landmarks was assessed by multiple redigitizing of one dataset. The dataset was analyzed 20 times without the zoom feature of the Software and 20 times with zoom (factor 10). The TRE between corresponding landmarks was calculated.

The measurements without zoom (figure 2) show an operator error of an average TRE of 0.10mm with a minimum of 0.001 and a maximum of 0.419mm.

The measurements with zoom (figure 3) show a significantly \((p<0.01)\) reduced operator error with an average TRE of 0.04mm.

**Capture error, recalibration**

Comparing landmark configurations from different image datasets of the same object quantifies the instability of the system. This was examined by studies No. 2 and 3 as outlined in table 1.

Figure 4 shows the results for a situation with re-calibration of the capture system before the acquisition of each dataset. The average TRE was 0.11mm with a range of 0.01 to 0.57mm. Taking into account the operator error from figure 2, the result is an instability of about 0.01mm \((p=0.15, \text{difference not significant})\) when images are taken with a re-calibration of the system in between.
Capture error, object positions

The influence of different object positions was evaluated through studies no. 4, 5 and 6 according to table 1. Representatively, the results of study row 5 are shown in figure 5. For evaluation the datasets were fused onto each other by rotation and translation until maximum superposition was achieved. Zero degree rotation was defined as the reference dataset. The mean TRE was 0.195mm with a range of 0.01 to 0.59mm. Taking into account the operator error from figure 2, the results are an instability of about 0.095mm when images are taken of the same object in different positions. These differences were not statistically significant for all measurements, even if figure 5 shows a little increase of the mean TRE with a greater deviation from the neutral position.

Accuracy and bias to direct anthropometry

To evaluate any differences between the 3D photo and reality, 201 distances between landmarks were measured by caliper and compared to the corresponding distances derived out of the 3D data.

A Pearson’s product-moment correlation coefficient (r) was calculated to compare direct and digital values for the measurements. Correlations were all statistically significant, with a grand mean of the r values calculated across all valid distances as 1.00. Paired t-tests comparing means between digital and direct measurements for all distances demonstrated a statistically insignificant difference. Additionally, linear regression analysis was performed. The allocation of the differences between direct measurements and distances derived out of the 3D data is shown in figure 6.

A comparative analysis of all error classes is given in figure 7.

Observation during application
The observation during testing showed some downsides of the technology. First, there are difficulties in capturing data if hair compromises any of the camera’s view of the area. Prominent areas can compromise the view of less prominent areas, resulting in poor 3D representation.

**DISCUSSION**

Conventional methods for studying facial symmetry have limitations. Radiography measures skeletal landmarks, but ignores the aesthetic aspects of soft tissue. The 3dMD face system allows the collection of images stored in digital format. In the present study, data acquisition was performed in NHP, because Kau et al. were able to show that this position is clinically reproducible. (28)

The evaluated parameters of accuracy, bias, and precision are outlined above and basically represent the quality of the produced 3D model when it is matched to reality. (29)

Concerning operator error, we recommend using the zoom tool whenever the operator is in any doubt about a landmark, but not as a routine. Even if the operator error can be reduced by a factor of 2 (from 0.1mm to 0.04mm), through using the zoom facility of the software, this strategy seems unnecessary due to the negligible error even without zoom.

The error of recalibration of the system (about 0.01mm with consideration of operator error), as outlined is figure 4, is negligible for itself and especially in comparison to the operator error, which is about 10 times higher.

The error resulting from the object’s position, as shown in figure 5 (about 0.095mm with consideration of operator error), is about the size of the operator error and in itself is negligible for clinical application.
Another point of discussion is the accuracy and bias compared to direct anthropometry. The distances directly measured and the distances calculated out of the datasets revealed no relevant difference and therefore no imminent error that would lead to systematically wrong results.

Overall, the system error of the 3dMDface™ system is comparable to that of other 3D imaging systems.(12, 30-34)

The problems with areas covered by hair (obvious in figure 1) are not very significant for the facial application. However, there is a cranial extension for the system which is meant to capture a 3D dataset of the whole head. In this, the hair is expected to be a major problem for correct detection of the skull.

Prominent and less prominent areas like the nose and the edges of a not-yet-treated cleft sometimes make it impossible to get a good 3D representation of these important regions.

**CONCLUSIONS**

The 3dMDface™ provides a good digital representation of reality under clinical circumstances. All occurring errors are negligible in themselves as well as in aggregation. Further development is necessary to reduce the influence of impaired camera vision in the cleft and nose areas. This can probably be addressed by additional cameras with different view angles.

Of course the user interface of the camera system and the software platform for further investigation can always be improved in matters of user friendliness and performance.

Further investigation has to be done in matters of the influence of facial expression on the results. A main goal might be the identification of reliable landmarks that are not affected by facial expression, but are still clinically relevant.
The possibility of reproducible identification of these landmarks has to be taken into account. As a last step the whole concept should be transferred into 4D, which means the 3D capturing not only of a still image, but also of a moving object.

ACKNOWLEDGEMENTS

The authors would like to thank the Surgical Planning Lab, Harvard Medical School, Boston, for hosting the planning phase of the study. We also thank Hildegard Eschle, senior librarian of the Dental School at the University in Zurich for helping with the literature research.

Conflict of interests

The authors declare that they have no conflict of interest.
Figure 1. Mannequin head as captured by the system
Figure 2. Operator error without zoom tool
Figure 3. Operator error with zoom tool
Figure 4. Error with re-calibration (including operator error)
Figure 5. Error through rotation of the object (including operator error)
Figure 6. Allocation of the differences between direct measurements and distances derived out of the 3D dataset.
Figure 7. Investigated error classes and results
<table>
<thead>
<tr>
<th>Study No.</th>
<th>A-P-position</th>
<th>Vertical inclination</th>
<th>Horizontal rotation</th>
<th>No. of acquired datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Direct caliper measurements</td>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>15cm posterior</td>
<td>neutral</td>
<td>none</td>
<td>18 (new calibration for each)</td>
</tr>
<tr>
<td>3</td>
<td>neutral</td>
<td>10 degrees down</td>
<td>none</td>
<td>17 (new calibration for each)</td>
</tr>
<tr>
<td>4</td>
<td>neutral</td>
<td>neutral</td>
<td>-5 to 5 degrees</td>
<td>11 (in 1 degree steps)</td>
</tr>
<tr>
<td>5</td>
<td>neutral</td>
<td>neutral</td>
<td>-30 to 30 degrees</td>
<td>21 (in 3 degree steps)</td>
</tr>
<tr>
<td>6</td>
<td>5cm posterior</td>
<td>5cm posterior</td>
<td>-30 to 30 degrees</td>
<td>21 (in 3 degree steps)</td>
</tr>
</tbody>
</table>

Table 1. Protocol of data acquisition
References


