

## Reproducibility and reliability of three-dimensional soft tissue landmark identification using three-dimensional stereophotogrammetry

Asli Baysal<sup>a</sup>; Ahmet Oguz Sahan<sup>b</sup>; Mehmet Ali Ozturk<sup>c</sup>; Tancan Uysal<sup>d</sup>

### ABSTRACT

**Objective:** To evaluate the intraexaminer repeatability and interexaminer reproducibility of soft tissue landmarks on three-dimensional (3-D) stereophotogrammetric images.

**Materials and Methods:** Thirty-four stereophotogrammetric images were taken and 19 soft tissue points were identified. The images were obtained using the 3-DMD Face (3-DMD TM Ltd, Atlanta, Ga) system. Two examiners marked 34 images manually with a mouse-driven cursor 4 weeks apart. Intraexaminer marking differences were calculated and classified as <0.5 mm, 0.5–1 mm, and >1 mm. Intraclass correlation coefficients were calculated for intraexaminer reliability. A paired-samples *t*-test was used to evaluate the difference between the examiners. Interexaminer reproducibility was evaluated by kappa analysis. Statistical significance was set at  $P < .05$ .

**Results:** Only one landmark (labiale superior) had an intraexaminer marking difference less than 0.5 mm. Existing landmarks had an intraexaminer difference less than 1 mm, but higher than 0.5 mm. The intraclass correlation coefficients (ICCs) indicated good intraexaminer repeatability for both observers. The ICC range for examiners 1 and 2 was 0.986–1.000 and 0.990–1.000, respectively. Kappa scores showed good interexaminer agreement, especially on the z-axis.

**Conclusions:** Except labiale superior, the soft tissue landmarks used in this study were shown to have moderate reproducibility, but the difference between the landmarks was less than 1 mm, and they had clinically acceptable reproducibility. (*Angle Orthod.* 2016;86:1004–1009)

**KEY WORDS:** Cephalometry; 3-D stereophotogrammetry; Orthodontics; Reproducibility; Computer-assisted three-dimensional imaging

### INTRODUCTION

Improvement in the facial appearance is one of the most important factors for patients seeking orthodontic treatment. With the paradigm shift from hard tissue to soft tissue, the key determinant in orthodontic diagnosis and treatment success now lies behind soft tissue positions. This paradigm shift necessitates using three-

dimensional (3-D) imaging, because traditional two-dimensional cephalometry places emphasis on hard tissue landmarks, because their reproducibility is better than soft tissue landmarks.<sup>1</sup> The rapid development of 3-D imaging systems has enabled us to diagnose facial soft tissues in detail and open a new window for 3-D soft tissue evaluation in orthodontic planning and treatment results. These 3-D imaging techniques are CBCT,<sup>2–4</sup> facial surface laser scanning,<sup>5,6</sup> and 3-D stereophotogrammetry.<sup>7</sup>

It is now possible to obtain viable and objective results in clinical trials using such advanced 3-D stereophotogrammetry, being one of the current 3-D imaging methods.<sup>8</sup> Especially in cases where 3-D cone beam computerized tomography is contraindicated, 3-D facial scanning seems a good alternative to record and analyze facial soft tissues.<sup>9</sup>

One use of stereophotogrammetry is the analysis of proportions and distances between soft tissue points of the face. An examiner marks the soft tissue landmarks manually using a mouse-driven cursor. Since it is not possible to palpate soft tissue points on 3-D photo-

<sup>a</sup> Associate Professor, Department of Orthodontics, Faculty of Dentistry, Izmir Katip Celebi University, Izmir, Turkey.

<sup>b</sup> Research Assistant, Department of Orthodontics, Faculty of Dentistry, Izmir Katip Celebi University, Izmir, Turkey.

<sup>c</sup> Assistant Professor, Department of Orthodontics, Faculty of Dentistry, Izmir Katip Celebi University, Izmir, Turkey.

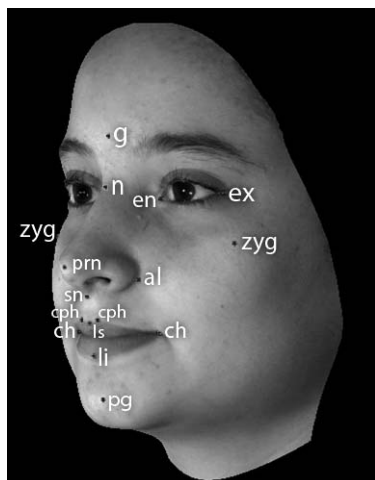
<sup>d</sup> Professor, Department of Orthodontics, Faculty of Dentistry, Izmir Katip Celebi University, Izmir, Turkey.

Corresponding author: Dr Asli Baysal, Izmir Katip Celebi Üniversitesi, Dişhekimliği Fakültesi, Ortodonti A.D., İzmir, 35640 Turkey (e-mail: baysalasli@hotmail.com)

Accepted: February 2016. Submitted: December 2015.

Published Online: March 29, 2016

© 2016 by The EH Angle Education and Research Foundation, Inc.



**Figure 1.** Soft tissue points used in this study.

graphs, the repeatability of the measurements of such soft tissue points—especially pogonion and zygion—is open to discussion.<sup>10</sup>

There are several studies on repeatability and reliability of 3-D soft tissue facial landmarks,<sup>8–13</sup> but conflicting results have been reported in these studies and no consensus has been reached. The aim of this study was to evaluate the intraexaminer repeatability and interexaminer reproducibility of soft tissue landmarks on 3-D stereophotogrammetric images. For the present study, the null hypothesis assumes that there is no difference in the intraexaminer reliability (no statistically meaningful inter- or intraexaminer consistency) or interexaminer reproducibility (absolute agreement) of soft tissue landmarks on 3-D stereophotogrammetric images.

## MATERIALS AND METHODS

Thirty-four stereophotogrammetric images taken from the patients (17 male, 17 female) of the same race who applied to postgraduate orthodontics clinics were used. Of the 34 patients, 9 had Class I, 20 had Class II, and 5 had Class III malocclusion. The mean age of the patients was  $13.1 \pm 1.3$  years. The ethics committee of Izmir Katip Celebi University approved the study (No. 54). Each patient's parent(s) signed an informed consent before inclusion in the study.

Inclusion criteria were as follows:

- no history of previous orthodontic treatment,
- no craniofacial or neuromuscular deformities,
- no history of facial trauma or plastic surgery intervention,
- being in the healthy percentage of the child mass index.

Stereophotogrammetric images were obtained using the 3DMD Face (3DMD TM Ltd, Atlanta, Ga) system. This system consisted of six cameras in total: two infrared cameras and one color camera in 2 sets. The cameras were positioned at predetermined angles and distances. 3DMDvultus software creates an image from the records it obtains through the six cameras from two separate points on the  $x$ ,  $y$ , and  $z$  coordinates. The software is also used for image processing, mapping, and measurements and is calibrated with the camera system to obtain flawless 3-D images for processing.

Definitions of the 19 soft tissue points used in the study (Figure 1) are given in Table 1. Both examiners had undergone special training for image capture, processing, and measurement. Prior to the study, the examiners had annotated more than 100 scans. To

**Table 1.** Soft Tissue Landmarks Used in This Study.

Abbreviation	Landmark	Definition
n	Nasion	Deepest point of the nasal bridge
g	Glabella	Most prominent midline point between the eyebrows
EN-R	Endocanthion-R	Inner commissure of the right eye fissure
EN-L	Endocanthion-L	Inner commissure of the left eye fissure
PRN	Pronasale	Most protruded point of apex nasi
SN	Subnasale	Midpoint of the angle in the columella base
Phil R	Philtrum-R	Midline groove in the upper lip that runs from the top of the lip to the nose
Phil L	Philtrum-L	Midline groove in the upper lip that runs from the top of the lip to the nose
PG	Pogonion	Most anterior midpoint of the chin
EX-R	Exocanthion-R	Outer commissure of the right eye fissure
AL-R	Alare-R	Outermost point of right alar contour
CH-R	Celion-R	Point located at the left labial commissure
EX-L	Exocanthion-L	Outer commissure of the left eye fissure
AL-L	Alare-L	Outermost point of left alar contour
CH-L	Cheilion-L	Point located at the left labial commissure
Lab-Sup	Labiale Superior	Midpoint of upper vermillion line
Lab-Inf	Labiale Inferior	Midpoint of lower vermillion line
Zyg-R	Zygion-R	The point where zygomatic dentation is the most apparent in soft tissue
Zyg-L	Zygion-L	The point where zygomatic dentation is the most apparent in soft tissue

**Table 2.** Intraexaminer Correlation Coefficients

Landmarks	Examiner 1			Examiner 2		
	x	y	z	x	y	z
Nasion	0.999	0.999	1.000	0.998	0.999	0.998
Glabella	1.000	0.999	0.998	1.000	0.999	1.000
En-R	1.000	0.999	1.000	1.000	0.999	1.000
En-L	1.000	0.999	1.000	1.000	0.999	1.000
PRN	1.000	0.999	1.000	1.000	0.999	0.999
SN	1.000	1.000	1.000	1.000	1.000	1.000
Phil-R	0.999	0.999	1.000	0.999	0.999	1.000
Phil-L	1.000	0.999	1.000	1.000	0.999	1.000
PG	1.000	0.999	1.000	1.000	0.997	0.996
Ex-R	1.000	0.998	1.000	1.000	0.999	1.000
Ex-L	1.000	1.000	1.000	1.000	0.998	0.999
AL-R	1.000	1.000	1.000	1.000	0.996	0.997
AL-L	0.999	1.000	1.000	1.000	0.997	0.998
CH-R	1.000	0.998	1.000	0.996	0.998	1.000
CH-L	1.000	1.000	1.000	0.998	0.997	1.000
Lab-Sup	1.000	1.000	1.000	1.000	0.999	1.000
Lab-Inf	1.000	0.999	1.000	1.000	0.999	1.000
Zyg-R	1.000	1.000	1.000	1.000	0.992	0.994
Zyg-L	1.000	0.986	0.996	0.998	0.990	0.993

prevent memory bias, two examiners independent of each other marked all soft tissue landmarks on 34 images 4 weeks apart.

### Marking Procedure

The 3-D photographs were adjusted according to the 3-D coordinate system suggested by Plooji et al.<sup>8</sup> To evaluate the landmark coordinate data, both examiners separately marked the soft tissue points in the frontal view, and then these markings were verified in profile and top-down views. The coordinates of the soft tissue points on the x, y, and z planes were extracted and recorded. The marking procedure was repeated after 4 weeks by the same examiners and the second coordinates were recorded. The software was designed to automatically calculate the Euclidean distance of the landmark positions; the difference in landmark positions was also recorded separately for both examiners. The 19 landmarks were classified according to the reading differences.<sup>12</sup> The classifications were

- highly reproducible landmarks wherein differences between readings were less than 0.5 mm,
- moderately reproducible landmarks wherein differences between readings were more than 0.5 mm, but less than 1 mm,
- poorly reproducible landmarks wherein differences between readings showed values greater than 1 mm.

### Statistical Analysis

IBM SPSS version 22.0 (SPSS Inc, Chicago, Ill) was used for statistical analysis. Descriptive statistics

included the mean, standard deviation, and minimum and maximum values. Intraexaminer reliability was calculated using intraclass correlation coefficients. A paired-samples *t*-test was used to assess the difference between examiners. Interexaminer reproducibility was evaluated by kappa analysis. Statistical significance was set at  $P < .05$ .

### RESULTS

Intraclass correlation coefficients (ICCs) for the examiners are given in Table 2. The ICC values indicate good intraexaminer repeatability for both observers. The ICC range for examiners 1 and 2 was 0.986–1.000 and 0.990–1.000, respectively.

Descriptive statistics of the data and interexaminer reproducibility assessments for the samples are shown in Table 3. All the landmarks were reproducible to less than 1 mm. The lowest reproducibility value was recorded for zygion (Zyg) point for both examiners, but this point was also classified as a moderately reproducible landmark (<1 mm). The landmark labiale superior (Lab-Sup) was the most reproducible landmark for both examiners.

Considering the difference between the two examiners, right alare landmark (AL-R) showed the lowest (0.110 mm) and left endocanthion (EN-L) showed the highest reproducibility (0 mm). No significant statistical difference was found between the markings of the first and second examiner. Thus, all the landmarks used in this study have been found to show high intra- and interexaminer consistency according to the results of the paired-samples *t*-test.

Table 4 shows the kappa scores of each annotation point on the x, y, and z coordinates. Highest consistency was observed on the z-axis for all the annotation points.

The null hypotheses testing intraexaminer consistency and interexaminer agreement failed to be rejected.

### DISCUSSION

Because of the paradigm shift from hard to soft tissue, 3-D facial soft tissue evaluations have become necessary. These facial scanning systems are continuously being developed, and the major concern about these new systems is reproducibility and reliability of the measurements obtained from the facial scans. The aim of this study was to assess the reproducibility and reliability of 19 anthropometric soft tissue landmarks marked on 3-D stereophotogrammetric images.

The most error-prone landmarks are those that are more leniently defined, such the points defining the jaw line.<sup>13</sup> These landmarks are relatively harder to determine visually, so require other methods such as

**Table 3.** Differences Between First and Second Examiner

Landmarks	Examiner 1				Examiner 2				Difference	Sig.
	Mean	SD	Minimum	Maximum	Mean	SD	Minimum	Maximum		
Nasion	0.788	0.314	0.250	1.310	0.824	0.285	0.250	1.420	0.036	0.649
Glabella	0.936	0.458	0.150	1.780	0.894	0.463	0.150	1.580	0.042	0.726
EN-R	0.740	0.358	0.130	1.490	0.734	0.299	0.190	1.290	0.006	0.937
EN-L	0.717	0.400	0.200	1.560	0.715	0.327	0.200	1.290	0.001	0.986
PRN	0.704	0.370	0.190	1.890	0.724	0.418	0.200	2.400	0.020	0.830
SN	0.759	0.310	0.130	1.290	0.703	0.400	0.100	1.600	0.056	0.529
Phil-R	0.794	0.376	0.150	1.680	0.736	0.297	0.280	1.270	0.059	0.485
Phil-L	0.692	0.284	0.170	1.400	0.766	0.353	0.220	1.380	0.073	0.255
PG	0.966	0.424	0.190	1.830	0.932	0.446	0.230	2.000	0.034	0.762
EX-R	0.706	0.408	0.040	1.700	0.789	0.345	0.190	1.640	0.083	0.372
EX-L	0.0668	0.271	0.270	1.280	0.754	0.395	0.080	1.940	0.086	0.308
AL-R	0.688	0.263	0.160	1.250	0.799	0.274	0.260	1.370	0.111	0.080
AL-L	0.674	0.251	0.160	1.280	0.755	0.262	0.170	1.220	0.081	0.208
CH-R	0.659	0.373	0.070	1.640	0.712	0.287	0.170	1.140	0.053	0.444
CH-L	0.679	0.314	0.250	1.550	0.612	0.352	0.060	1.420	0.067	0.441
Lab-Sup	0.482	0.297	0.110	1.260	0.468	0.208	0.100	1.090	0.014	0.839
Lab-Inf	0.573	0.296	0.100	1.130	0.625	0.398	0.110	2.000	0.052	0.565
Zyg-R	0.937	0.498	0.290	2.610	0.992	0.525	0.120	2.160	0.055	0.683
Zyg-L	0.976	0.612	0.220	2.510	0.912	0.569	0.150	2.250	0.063	0.696

palpation.<sup>9</sup> They are located on convex surfaces, and the operator should control their localization on the x-, y-, z-axes separately. Examples such as zygion, gonion, and tragion have been shown to have unacceptable reproducibility.<sup>8,10</sup> In this study, zygion, pogonion, nasion, and glabella had the highest intra-examiner variability (less than 1 mm), and the landmarks were classified as moderately reproducible.

Landmarks with well-defined borders or edges were shown to have higher degrees of reproducibility than the ones located on gently curving slopes.<sup>12</sup> The reproducibility and reliability of landmarks that are near orbits (endocanthion and exocanthion) and the mouth (labiale superior, labiale inferior, and cheilion) are

considered moderate. These landmarks are easy to determine visually. Improvements in 3-D imaging systems' resolution have made determination of these landmarks easier.

Hajeer et al.<sup>8</sup> identified a landmark as highly reproducible if the standard deviation (SD) from centroid was 0.5 mm or less. They evaluated intra-examiner reproducibility and found that 20 out of 30 landmarks showed high reproducibility. Gwilliam et al.<sup>10</sup> found 12 out of 24 landmarks that were reproducible (intraexaminer data) assuming that the SD is less than 1 mm. They also evaluated the interexaminer data and reported that only cheilion and labiale superior were reproducible. In the current study, there was 1 landmark classified as highly reproducible and 18 landmarks moderately reproducible for each examiner.<sup>10</sup> For the interexaminer data, all landmarks were found to be reproducible, and no statistically significant difference was found between two examiners.

Glabella, nasion, and pogonion showed poor reproducibility on the y-axis.<sup>12</sup> The authors related this finding to good manipulation skills and clinical knowledge of the natural head position. In the present study, all facial scans were taken in the natural head position, which has been shown to be highly reproducible.<sup>14</sup> Another standardization criterion for the scanning procedure was the selection of participants from the subjects within the healthy percentage according to body mass index. This may have led the examiners to encounter more standardized soft tissue contours and not deformations such as soft tissue sag.

At the beginning of the study, to prevent possible errors that might arise from differing landmark definitions by the examiners, 19 facial landmarks were

**Table 4.** Kappa Scores

Landmark	x	y	z
Nasion	0.0838	0.832	1.000
Glabella	0.0840	1.000	1.000
EN-R	0.885	1.000	1.000
EN-L	0.887	1.000	1.000
PRN	1.000	0.943	1.000
SN	0.889	1.000	1.000
Phil-R	0.939	0.928	1.000
Phil-L	0.933	0.928	1.000
PG	0.833	1.000	1.000
EX-R	1.000	1.000	1.000
AL-R	0.892	0.872	1.000
CH-R	0.785	0.913	1.000
EX-L	1.000	1.000	1.000
AL-L	0.767	0.812	1.000
CH-L	0.788	1.000	1.000
Lab-Sup	0.889	1.000	1.000
Lab-Inf	0.788	1.000	1.000
Zyg-R	1.000	0.714	1.000
Zyg-L	1.000	0.726	1.000



clearly defined and, if there was a conflict, the examiners discussed the localization until a consensus was reached. These measures may have resulted in high interexaminer reproducibility. Also, we found high correlation coefficients for both examiners in all three planes of space. All landmarks showed a kappa score of 1.000 for the z-axis. In general, relatively poor kappa scores were found for the x-axes, so we suggest that clinicians be more cautious when placing landmarks on the x-axis.

Plooji et al.<sup>9</sup> asserted that the actual 3-D distance or difference between two landmarks remains unclear if the landmark is defined separately on the x, y, and z coordinates, thus these landmarks should accurately be identified even in the 3-D mode. They evaluated the actual distance between two similar landmarks and calculated the mean differences, finding that this distance was less than 0.5 mm for 76% of the first examiner's recordings, 92% of the second examiner's recordings, and 80% of the interexaminer recordings. Contrarily, we found that one landmark (labiale superior) had an intraexaminer marking difference less than 0.5 mm. The other 18 landmarks had an intraexaminer difference less than 1 mm but higher than 0.5 mm. The different results should be addressed, especially since the same scanning system was used in both studies. The only valid explanation may be the experience of the examiners. If this assumption is correct, there is a learning curve for 3-D tracings, and experience is very important for correct 3-D landmark positioning.

Toma et al.<sup>12</sup> found lower intra- and interexaminer reproducibility comparable with present results, but they assumed that landmarks have clinically acceptable reproducibility if the difference is less than 1 mm. Similar to our study, they found that labiale superior was the most reproducible landmark.

Familiarity with the software program is also likely to play a part in the examiners' ability to place landmarks.<sup>10</sup> In addition to training in the use of the program, the examiners were experienced in using it and able to manipulate the images with full 3-D control. This skill is also thought to improve the reproducibility of landmark placement in this study.

Toma et al.<sup>12</sup> reported that poorly reproducible landmarks were related to the eyes, and this was linked to acquisition of the image using a laser-based system. We did not encounter such a finding using the stereophotogrammetric scanning system, which may be account for the negative side of the laser-based systems.

Plooji et al.<sup>9</sup> stated that the quality of 3-D photographs has improved, especially at the levels of the nasal base, ears, and hairline. With the advances in 3-D imaging and the software we used, we can say that

landmark identification is becoming more accurate and that those used in this study showed high levels of reproducibility and accuracy in analyzing facial soft tissue on stereophotogrammetric images. However, clinicians should remember that some points, especially those located on gently curved slopes are more prone to error in landmark localization.

## CONCLUSIONS

- All soft tissue landmarks were shown to have high reproducibility and reliability, both in inter- and intraexaminer comparison.
- Soft tissue landmarks used in this study were shown to differ less than 1 mm between repeated markings.

## ACKNOWLEDGMENTS

This work was supported by research grants from the Scientific and Technological Research Council of Turkey (project 112R033) and Izmir Katip Celebi University, Scientific Research Projects Unit (project 2013-3-TSBP-32). The authors thank Dr. Bulent Ozkan for his assistance with the statistical analysis.

## REFERENCES

1. Sarver DM, Ackerman JL. Orthodontics about face: the re-emergence of the esthetic paradigm. *Am J Orthod Dentofacial Orthod.* 2000;117:575–576.
2. Groeve PD, Schutyser F, Cleynenbreugel JV, Suetens P. Registration of 3D photographs with spiral CT images for soft tissue simulation in maxillofacial surgery. *Lecture Notes In Computer Science; Vol. 2208. Proceedings of the 4th International Conference on Medical Image Computing and Computer-assisted Intervention.* 2001;991–996.
3. Mozzo P, Procacci C, Tacconi A, Martini PT, Andreis IA. A new volumetric CT machine for dental imaging based on the cone-beam technique: preliminary results. *Eur Radiol* 1998; 8:1558–1564.
4. Swennen GRJ, Schutyser FAC, Hausamen J-E. *Three-dimensional Cephalometry: A Color Atlas and Manual.* Berlin: Springer GmbH 2005;320.
5. Kau CH, Richmond S, Zhurov AI, et al. Reliability of measuring facial morphology with a 3-dimensional laser scanning system. *Am J Orthod Dentofacial Orthop* 2005;128:424–430.
6. McCance AM, Moss JP, Wright WR, Linney AD, James DR. A three-dimensional soft tissue analysis of 16 skeletal class III patients following bimaxillary surgery. *Br J Oral Maxillofac Surg.* 1992;30:221–232.
7. Ras F, Habets LL, van Ginkel FC, Prah-Andersen B. Quantification of facial morphology using stereophotogrammetry—demonstration of a new concept. *J Dent.* 1996; 24:369–374.
8. Hajeer MY, Ayoub AF, Millett DT, Bock M, Siebert JP. Three-dimensional imaging in orthognathic surgery: the clinical application of a new method. *Int J Adult Orthodon Orthognath Surg.* 2002;17:318–330.
9. Plooji JM, Swennen GRJ, Rangel FA, et al. Evaluation of reproducibility and reliability of 3D soft tissue analysis using

- 3D stereophotogrammetry. *Int J Oral Maxillofac Surg.* 2009;38:267–273.
10. Gwilliam JR, Cunningham SJ, Hutton T. Reproducibility of soft tissue landmarks on three-dimensional facial scans. *Eur J Orthod.* 2006; 28:408–415.
  11. Kau CH, Richmond S, Zhurov AI, et al. Reliability of measuring facial morphology with a 3-dimensional laser scanning system. *Am J Orthod Dentofacial Orthop* 2005;128: 424–430.
  12. Toma AM, Zhurov A, Playle R, Ong E, Richmond S. Reproducibility of facial soft tissue landmarks on 3D laser-scanned facial images. *Orthod Craniofac Res.* 2009;12:33–42.
  13. Fagertun J, Harder S, Rosengren A, et al. 3D facial landmarks: inter-operator variability of manual annotation. *BMC Medical Imaging.* 2014;14:35.
  14. Chiu CS, Clark RK. Reproducibility of natural head position. *J Dent.* 1991;19:130–131.