Clinical Oral Investigations
https://doi.org/10.1007/s00784-019-03150-1

ORIGINAL ARTICLE

# Validity and reproducibility of the 3D VECTRA photogrammetric surface imaging system for the maxillofacial anthropometric measurement on cleft patients

Siti A. Othman [1,2] · Lyddia Saffai [3] · Wan N. Wan Hassan [1,2]

Received: 27 February 2019 / Accepted: 11 November 2019


## Abstract

**Objectives** To validate the accuracy and reproducibility of linear measurements of three-dimensional (3D) images and to compare the measurements with the direct anthropometry method on cleft lip and palate (CLP) patients.

**Materials and methods** Nineteen linear facial measurements were derived from 16 standardized surface landmarks obtained from 37 cleft patients (20 males, 17 females; mean age 23.84 years, standard deviation ± 6.02). They were taken manually with calipers and were compared with the digitally calculated distance on the 3D images captured using the VECTRA-M5 360° Imaging System with pre-marked landmarks. Another pair of 19 linear measurements were computed on the 3D images 2 weeks apart for intra- and inter-observer agreements. Statistical analyses used were paired *t* test, the Bland-Altman analysis, and the intra-class correlation coefficient (ICC) index.

**Results** Most of the linear measurements showed no statistically significant differences between the proposed method and direct anthropometry linear measurements. Nevertheless, bias of the 3D imaging system is present in the linear measurements of the nose width and the upper vermillion height. The measurements' mean biases were within 2 mm, but the 95% limit of agreement was more than 2 mm. Intra- and inter-observer measurements generally showed good reproducibility. Four inter-observer measurements, the upper and lower face heights, nose width, and pronasale to left alar base were clinically significant.

**Conclusions** Measurements obtained from this 3D imaging system are valid and reproducible for evaluating CLP patients.

**Clinical relevance** The system is suitable to be used in a clinical setting for cleft patients. However, training of the operator is strictly advisable.

**Keywords** Orthodontics · Oral and maxillofacial surgery · Cleft lip and palate · Three-dimensional analysis · Accuracy · Reproducibility

## Introduction

Cleft lip and palate (CLP) patients possess certain characteristics, and the most clinically relevant affected craniofacial regions are the nose and orolabial region. Most CLP patients have larger faces than normal patients [1]. Krimmel et al. [2] observed that the highest degree of deformity was seen in the horizontal dimension of the nose. They explained that their findings might have resulted from failure to restore the loss of the medial insertion points of the facial muscle during primary operations, which flattened and elongated the alar and resulted in an increase in the horizontal nasal dimension. Othman et al. [1] compared the volumetric changes of the facial characteristics in normal and unilateral CLP patients using the VECTRA 3D Imaging System. They found that there was a significant difference in the nasolabial area between normal faces and those with unilateral CLP.

Most CLP patients that undergo early treatments are babies. They are consulted from the beginning of their life. Thus, taking records in young CLP patients may impose problems to the operator. They are unable to stay still for a long time for detailed

✉ Siti A. Othman
sitiadibah@um.edu.my

1 Department of Paediatric Dentistry & Orthodontics, Faculty of Dentistry, University of Malaya, 50603 Kuala Lumpur, Malaysia

2 Clinical Craniofacial Dentistry Research Group, Faculty of Dentistry, University of Malaya, 50603 Kuala Lumpur, Malaysia

3 Bangsar Dental Clinic, Jalan Bangsar, 59200 Kuala Lumpur, Federal Territory, Malaysia

Published online: 21 November 2019

examination. Recent technology, the fast and non-invasive three-dimensional (3D) stereophotogrammetry system, may be able to overcome this problem due to its many advantages, such as quick capture speed and the ability to produce high-resolution photorealistic images. Furthermore, it had been validated through various studies using different subjects or samples. Some used mannequin heads [3], normal patients [4], cadaver heads [5], and Down syndrome patients [6]. Two studies had evaluated the reproducibility of the 3D stereophotogrammetry system on cleft infants [7, 8], but they did not assess the accuracy of the system on these patients. Another study also evaluated the reliability of a 3D stereophotogrammetry system on cleft infants, but they only focused on the nasolabial area [9].

Due to the volumetric changes in the facial region of these patients, it might have caused some image distortion and unclear facial landmarks. Furthermore, it has been reported that 3D imaging systems are affected by changes in muscle tone, facial expression, and head posture [10]. This could lead to inaccuracy of the 3D system, and eventually, evaluation, treatment plan, and outcome could also be compromised. To our knowledge, there are no studies that have been carried out to determine the accuracy of the measurements taken with a 3D imaging system on adult CLP patients. Previous studies were done on infant cleft patients using the 3dMDcranial System [7–9]. One study was found utilizing adult CLP patients as subjects. However, their study was on the degree of 3D asymmetry measurements of facial soft tissues [11]. Thus, the aim of this prospective experimental study was to validate the accuracy and the reproducibility of the VECTRA-M5 360° Imaging System. The null hypothesis was that there is no significant difference between measurements taken by the 3D imaging system and direct anthropometry measurements on adult CLP patients.

## Materials and methods

### Study design

This was a prospective experimental study to validate the accuracy and the reproducibility of linear measurements taken by the 3D VECTRA-M5 360° Imaging System and to compare the measurements with the direct anthropometry method on adult CLP patients. Ethical approval was obtained from the Medical Ethics Committee, Faculty of Dentistry, University of Malaya, Kuala Lumpur. The Ethics Committee/IRB Reference Number is DF CD1410/0086(P). The methods were carried out in accordance with the approved guidelines.

### Sampling and sample

The study sample consisted of 37 cleft lip (CL) or CLP individuals with a mean age of 23.84 years (standard deviation ± 6.02) and more males (54.1%) than females (45.9%). Subjects were recruited from the Combined Cleft Lip and Palate Clinic, Faculty of Dentistry, University of Malaya. Recruitment took place from September 2015 to September 2016. The inclusion criteria were as follows: (1) repaired CL or repaired CLP patients; (2) aged 18 years old and above; (3) not wearing any facial prosthesis; (4) non-syndromic CL or CLP patients; and (6) no previous facial surgery. Exclusion criteria were as follows: (1) patients aged below 18 years old and (2) patients with cleft palate, alveolus, or soft palate only.

Sample size was calculated using G*Power Software Version 3.1.9.2 [12]. It was based on a previous study [13] on the assessment of the accuracy of the stereophotogrammetry system and considered the most affected facial region of CLP patients [1, 2, 14]. Sample size was chosen with consideration for a mean difference between two methods of 0.75 mm and a standard deviation of 1.43 mm [13]. Therefore, with a sample size of 37, there will be 0.87 power ($\alpha = 0.05$) to detect at least 1 mm difference in measurement.

For analyzing reproducibility of the system, the sample size was calculated using an equation proposed by Walter et al. [15] to determine the total number of samples needed to test the intra-class correlation coefficient. Ten images were needed taking into account the acceptable correlation was set at 0.40 [16] and the range of intra- and inter-class correlation reported by previous studies was 0.73 to 0.99 for nose width [3, 4, 6]. Randomization software was used to randomly select the 10 unlabeled 3D images [17]. Measurements on these 10 selected images were taken once and again after 2 weeks to reduce memory bias by the same observer for intra-observer reproducibility and another set by a different observer for inter-observer reproducibility.

All suitable subjects were given verbal and written explanations regarding the purpose as well as the nature of the study procedure, and invited to join the research. Verbal and written consent was obtained from the subjects who agreed to participate.

### VECTRA 3D imaging system

The VECTRA-M5 360° Imaging System, developed by Canfield Scientific Inc., Fairfield, NJ, USA, was used in this study (Fig. 1). It is a photogrammetry system that consists of five pods of cameras placed at different angulations from the subject. The camera pods were fixed to an interconnected rigid frame, and the subject sat in the middle of the frame. Each pod contained one color camera and one monochrome camera. These two-dimensional (2D) digital cameras captured the image simultaneously, and the outcome was seen in a 3D image. The VECTRA-M5 360° Imaging System is able to capture a full 360° picture of the head in less than 2 ms. The images were then processed within a few minutes by the Mirror® software (Canfield, Fairfield, NJ, USA) to produce a colored and textured 3D image of the subject, which the operator was

**Fig. 1** VECTRA-M5 360° Imaging System

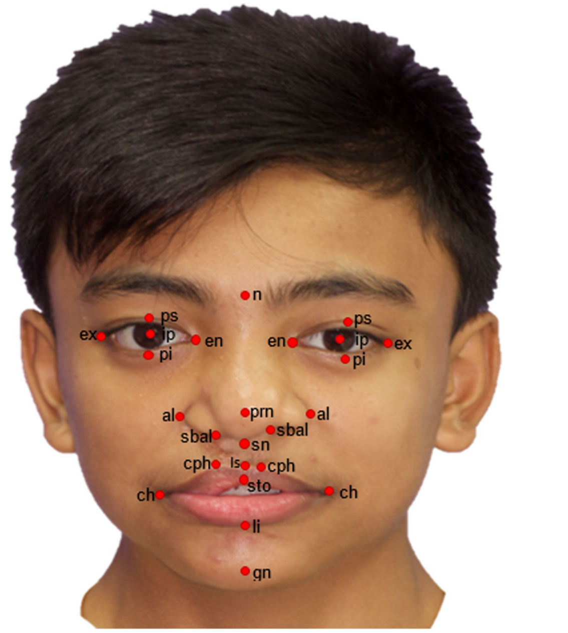


**Fig. 2** Anthropometric landmarks. *n* soft tissue nasion, *prn* pronasale, *sn* subnasale, *Is* labrale superius, *sto* stomion , *li* labrale inferius, *gn* soft tissue gnathion, *en* endocanthion, *ex* exocanthion, *ps* palpebrale superius, *pi* palpebrale inferius, *pi* pupil, *al* alare, *sbal* subalare, *cph* christa philtri, *ch* cheilion

able to explore in *x*, *y*, and *z* dimensions. The image-capturing procedure was carried out at the Faculty of Dentistry, University of Malaya, Kuala Lumpur.

Calibration of the VECTRA-M5 360° Imaging System was made according to the manufacturer's instructions to ensure detailed geometric configuration of all cameras. It is important to run this process so that the relationship of the cameras and other components is understood and recorded by the system. The camera was calibrated daily or when the system had been moved or altered.

## Facial landmarks and linear measurements

Nineteen linear facial measurements derived from 16 standardized surface landmarks were used in this study [14, 18, 19]. The landmarks and linear measurement descriptions (Fig. 2; Table 1) were selected for their clinical relevance in describing the facial analysis of the cleft patients.

## Direct anthropometry

Each subject sat on a stool with their head in a natural position to ensure reproducibility [20]. The subjects were instructed to relax and minimize their facial expression, which can influence measurements. Fifteen out of the 25 landmarks were marked on each subject's face with 0.5-mm-diameter dots using a liquid eyeliner [21].

A digital sliding caliper (Pro-Max, Mitutoyo, Japan) was used for direct facial linear measurements, calculated to the nearest 0.01 mm. Two sets of measurements were taken with at least 15 min between measurements. These measurements were averaged and were accepted as the maximum accuracy achievable. As manual measurements served as the gold standard, it was assumed that these measurements represented the actual distances between landmarks [22]. Measurements were performed with care and avoided placing pressure on the soft tissue, which may lead to soft tissue deformation by the

**Table 1** Linear measurement descriptions

| Linear measurements | Description |
|---|---|
| n–gn | Face height |
| n–sto | Upper face height |
| sn–gn | Lower face height |
| n–sn | Nose height |
| al–al | Nose width |
| sn–prn | Protrusion of nasal tip |
| n–prn | Nose dorsum length |
| ch–ch | Mouth width |
| sn–sto | Upper lip height |
| sn–Is | Cutaneous upper lip height |
| Is–sto | Upper vermillion height |
| sto–li | Lower vermillion height |
| sbal–sbal | Subalare width |
| al (lt)–prn | Pronasale to left alar base |
| al (rt)–prn | Pronasale to right alar base |
| sbal (lt)–sn | Subnasale to left alar base |
| sbal (rt)–sn | Subnasale to right alar base |
| sbal–cph (lt) | Left upper lateral lip length |
| sbal–cph (rt) | Right upper lateral lip length |

pointed arms of the calipers [18]. All measurements were made by a single operator (LS) and took an average of 45 min to complete per subject.

### Digitalized facial measurements

Subjects wore a head cap prior to image capture as the imaging system was not able to record hair, which can lead to image distortion. Subjects sat in the center of the imaging system frame with their head in a natural position [20]. The computer monitor displayed five paneled windows to preview the subject's position prior to image capture. The central image preview had a grid to guide the operator to adjust the position of the subject to get the maximum field of view. Similar to direct anthropometry, the subjects were asked to relax with a neutral facial expression. Two image captures were taken, one without and another with landmarks labeled on the subject's face. After each procedure, the data was immediately processed to form a 3D image in natural color (Fig. 2) to assess whether it was acceptable. Otherwise, the image capture process was repeated.

Landmarks on the unlabeled faces were identified by their natural references and definitions (Fig. 2; Table 1) while those of the labeled faces were identified by the marks made on the faces, which were taken prior to image capture using the liquid eyeliner. The mouse was used to maneuver the pointer and manually mark the identified landmarks on the processed 3D image. The linear measurements between the landmarks that represented the facial characteristics were calculated by the software. As with direct caliper measurements, each subject's images were measured twice, 15 min apart. The values were then averaged.

### Statistical analysis

Data collected were analyzed using the Statistical Package for the Social Sciences (SPSS) Version 23 (SPSS for Windows, SPSS Inc., Chicago, IL, USA). The Shapiro-Wilk test was carried out on all data of the direct and 3D digitized measurements. The result confirmed data normality. To assess systematic bias between the methods for a specific variable, differences in linear measurements were assessed using the paired $t$ test. Bland analyses were performed to assess agreement between direct and 3D digitized measurements. For this analysis, a total deviation of ± 2.0 mm was defined to be clinically acceptable. Therefore, a 95% limit of agreement beyond 2 mm was considered as not clinically acceptable [4, 18, 23]. Intra-class correlation coefficient (ICC) was used to evaluate the consistency between measurements.

To assess if repeated measurements were significantly different at specific anatomic landmarks per rater and between raters (after a 2-week interval), for the 3D stereophotogrammetry images, the paired $t$ test (normally distributed data) and non-parametric Wilcoxon signed rank test (non-normally distributed data) were performed. To evaluate the reliability of the measurements per observer (LS) and between two different observers (LS and RA), the ICC was assessed. For this analysis, high reliability is indicated when the value is close to 1 and low reliability is indicated when the value is near to 0.

The level of significance for the paired $t$ test was set at $p < 0.002$ after Bonferroni correction in order to reduce type 1 error.

## Results

The sample group comprised 17 left CLP (45.9%), 15 bilateral CLP (40.5%), 3 right cleft lip (8.1%), and 2 right CLP (5.4%) with a mean age of 23.8 (SD 6.0) years. The means and SD for direct and 3D digital measurements for each variable are shown in Table 2. The average differences were generally negative, comprising 14 linear measurements out of 19 measurements (73.7%) that were taken, which indicated that 3D digital measurements had slightly higher measurement values.

### Accuracy of the 3D stereophotogrammetry

#### Facial region

Table 3 displays the statistical analysis comparing the direct and 3D digital method of measurements for the facial region. Paired $t$ test showed that the differences were not statistically significant ($p > 0.002$ and 95% CI included zero). The highest mean difference was the upper face height (− 0.37 mm ± 2.04) followed by the lower face height (− 0.11 mm ± 1.79), and the least difference was face height (− 0.06 mm ± 0.99). There was good correlation between both methods with the ICC values of 0.95 to 0.99. However, Bland-Altman 95% limits of agreement showed clinically acceptable differences (< 2 mm) only for the face height. The upper and lower face heights showed clinically unacceptable differences between the two methods with 95% limits of agreement that exceeded ± 2.00 mm (Fig. 3).

#### Nose region

The results for the nose regions showed that all the measurements between direct and 3D digital methods had no statistically significant differences except for nose width ($p < 0.002$ and 95% CI exclude zero) (Table 4). The consistency of the measurements showed good correlation with the strength range of 0.86 to 0.98. The least agreeable measurement between the two methods was the nose width (ICC = 0.86). Nevertheless, it was still considered of good agreement. However, 95% limits of agreements were clinically unacceptable for the nose dorsum length (− 2.43 to 3.15) and nose width (− 5.09 to 2.39) (Fig. 4).

**Table 2** Means, mean differences, standard deviations, and minimum and maximum values for manual and 3D digital measurements

| Variables | Manual measurements (mm) | | | | 3D digital measurements (mm) | | | | Mean differences |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Min | Max | Mean | SD | Min | Max | |
| Face height | 118.18 | 8.72 | 101.00 | 134.58 | 118.24 | 8.62 | 102.21 | 133.31 | − 0.06 |
| Upper face height | 77.32 | 6.77 | 67.04 | 94.00 | 77.69 | 6.27 | 66.26 | 88.96 | − 0.37 |
| Lower face height | 60.63 | 5.76 | 48.72 | 72.34 | 60.74 | 5.80 | 48.37 | 72.76 | − 0.11 |
| Nose height | 58.00 | 5.43 | 42.12 | 66.44 | 58.01 | 5.16 | 44.25 | 67.04 | − 0.01 |
| Nose width | 41.67 | 4.35 | 33.25 | 50.65 | 43.02 | 4.15 | 35.09 | 51.25 | − 1.35 |
| Protrusion of nasal tip | 17.74 | 2.15 | 13.03 | 21.50 | 17.49 | 2.09 | 13.15 | 21.65 | 0.28 |
| Nose dorsum length | 49.16 | 4.57 | 39.63 | 61.53 | 48.80 | 4.59 | 40.08 | 60.03 | 0.36 |
| Mouth width | 51.97 | 4.27 | 45.03 | 62.62 | 52.24 | 4.40 | 45.47 | 61.48 | − 0.28 |
| Upper lip height | 19.58 | 3.51 | 10.53 | 26.29 | 19.83 | 3.52 | 11.93 | 27.21 | − 0.25 |
| Cutaneous upper lip height | 11.66 | 3.23 | 5.66 | 20.02 | 11.70 | 3.23 | 4.12 | 19.12 | − 0.05 |
| Upper vermillion height | 8.12 | 2.86 | 2.10 | 15.26 | 9.15 | 3.02 | 2.26 | 16.06 | − 1.02 |
| Lower vermillion height | 11.20 | 2.41 | 3.00 | 15.12 | 10.57 | 2.15 | 3.33 | 14.66 | 0.63 |
| Subalare width | 20.09 | 3.76 | 12.19 | 33.67 | 20.15 | 3.55 | 12.95 | 33.66 | − 0.06 |
| Pronasale to left alar base | 30.75 | 3.18 | 20.44 | 36.51 | 30.44 | 2.82 | 24.75 | 35.46 | 0.31 |
| Pronasale to right alar base | 30.35 | 3.05 | 24.32 | 36.87 | 30.10 | 2.66 | 25.93 | 37.73 | 0.25 |
| Subnasale to left alar base | 10.21 | 2.09 | 7.01 | 15.68 | 10.30 | 2.03 | 5.76 | 15.35 | − 0.09 |
| Subnasale to right alar base | 11.37 | 1.83 | 7.61 | 16.65 | 11.60 | 1.88 | 8.43 | 17.40 | − 0.23 |
| Left upper lateral lip length | 14.22 | 2.68 | 8.67 | 19.67 | 14.37 | 2.90 | 9.00 | 20.15 | − 0.16 |
| Right upper lateral lip length | 15.14 | 3.14 | 8.67 | 21.57 | 15.33 | 3.14 | 8.72 | 22.35 | − 0.19 |

### Orolabial region

There were no statistically significant differences in all orofacial measurements between both methods except for upper vermillion height ($p < 0.002$ and 95% CI exclude zero) (Table 5). Generally, the ICC values showed good correlations for all measurements between the two methods (range 0.87 to 0.97). There is only one measurement which was the lower vermillion height that fell slightly in the range of acceptable correlation (0.61). The systematic differences between the two methods were small, with the largest mean bias found on measurement of the upper vermillion height showing that the 3D method was 1.02 mm longer than the manual method. Nonetheless, the 95% limit of agreement was not clinically acceptable for eight out of 12 linear measurements of the orolabial region (Fig. 5). Only measurements of the cutaneous upper lip height, subalare width, and subnasale to the left and right alar bases between the two methods were clinically acceptable (< 2.0 mm).

## Reproducibility of 3D landmark localization

### Intra-observer reproducibility

Table 6 presents the results for intra-observer reliability on ten selected samples. Shapiro-Wilk test found that three of the measurements were not normally distributed. Therefore, the Wilcoxon signed rank test was performed on these non-normally distributed measurements.

**Table 3** Mean difference, SD, Bland-Altman 95% limit of agreement, $p$ values, and the ICC between the two methods in the face region.

| Measurements | Paired $t$ test (manual–3D) | | | | | Bland-Altman | | | ICC | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean difference | SD | 95% CI | | $p$ value | Mean bias | 95% limits of agreement | | $r$ | 95% CI | | $p$ value |
| | | | Lower | Upper | | | Lower | Upper | | Lower | Upper | |
| Face height | − 0.06 | 0.99 | − 0.39 | 0.27 | 0.70 | − 0.06 | − 2.00 | 1.88 | 0.99 | 0.99 | 1.00 | 0.00* |
| Upper face height | − 0.37 | 2.04 | − 1.05 | 0.31 | 0.28 | − 0.37 | − 4.37 | 3.64 | 0.95 | 0.91 | 0.97 | 0.00* |
| Lower face height | − 0.11 | 1.79 | − 0.70 | 0.49 | 0.72 | − 0.11 | − 3.61 | 3.40 | 0.95 | 0.91 | 0.98 | 0.00* |

*$p < 0.001$

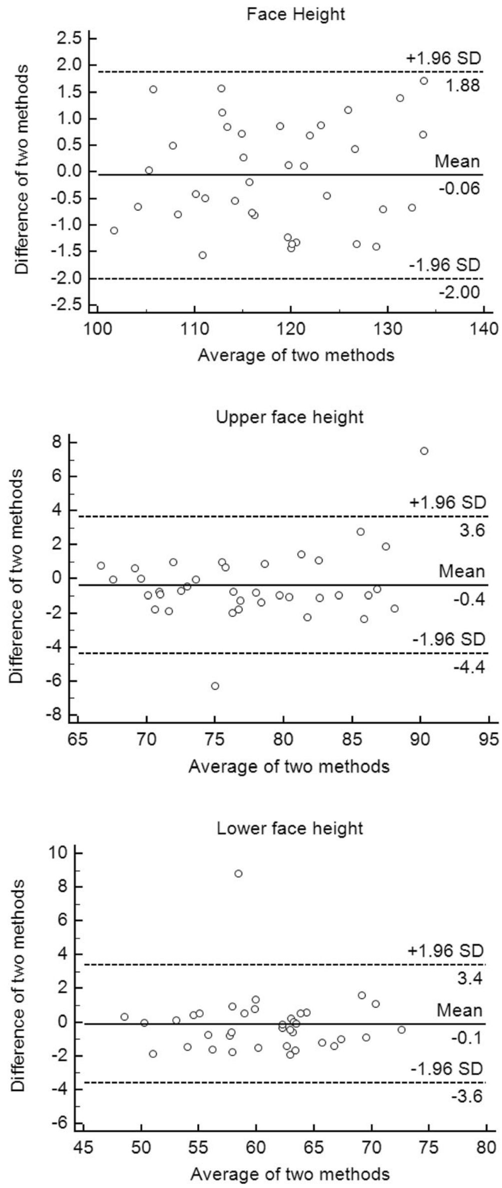


**Fig. 3** Bland-Altman plots of the face region

Paired *t* tests showed that the intra-observer's measurements were not statistically significant ($p > 0.002$) for all linear measurements. Fourteen out of 19 linear measurements were in good agreement (ICC scores were between 0.77 and 0.96; $p < 0.05$). The ICC of four measurements fell in the range of acceptable agreement (0.61 to 0.72; $p < 0.05$). Only one measurement, subnasale to left alar base, was in poor agreement (0.38; $p > 0.05$).

### Inter-observer reproducibility

The same statistical tests were prepared to compare the reproducibility of the calculated distances between two observers. The results are displayed in Table 7. One of the measurements, subalare width, was not normally distributed; thus, Wilcoxon signed rank test was conducted.

Paired *t* test showed that all linear measurements between the two observers were not statistically significant. However, upper face height (2.13 mm ± 3.07), lower face height (− 3.31 mm ± 4.46), nose width (2.62 mm ± 2.93), and pronasale to left alar base (2.47 mm ± 2.75) showed mean differences beyond 2 mm.

Nine linear measurements showed good agreement between the two observers (ICC between 0.77 and 0.89; $p < 0.05$), and seven other linear measurements were in the range of acceptable correlation (0.51–0.72; $p < 0.05$). The other three, protrusion of nasal tip, pronasale to right alar base, and subnasale to left alar base, showed poor agreement (0.34–0.42; $p > 0.05$).

## Discussion

Any newly developed system needs to be validated prior to application in any clinical use. Validation measures the extent to which the diagnostic test, in this instant the 3D stereophotogrammetry system, is able to perform and accomplish what it is designed for. It is determined by measuring how well a test performs against the gold or criterion standard, which in this case is the direct measurement [24]. Past studies have reported the validity of the 3D stereophotogrammetry systems. Validation tests have been performed on the mannequin head [3, 21, 25], cadaver heads [5], normal facial morphology patients [4, 25], non-synostotic cranial deformity [26], and also static objects [22, 27, 28]. The system has been reported as valid to be used in a clinical setting. The clinically acceptable limit of the difference between the diagnostic and the gold standard has been set at 2 mm [4, 28]. The Bland-Altman analysis was performed to analyze the limit of agreement [4, 29, 30]. From various studies that have been conducted, the accepted mean difference that is considered to be clinically acceptable is below 2 mm [4, 18, 23]. We therefore propose that a difference of 2 mm is an appropriate threshold to indicate clinical relevance distance between the diagnostic and the gold standard. This means that a significant change can be observed and detected if the difference is 2 mm or more.

The current study design of implementation of the 3D surface analysis in the craniofacial region had been executed in previous studies. However, most of the studies on CLP patients were on infants or adolescents [7–9]. The reason might be that these patients are still in the treatment phase; hence, subjects are easily recruited from the clinic. In contrast to adult patients, they usually have completed their treatment. In consequence, subjects' recruitment is difficult and as a result not many studies have been done on adult CLP patients.

The linear measurement method was chosen in this study to measure validity as it reflects the clinical setting. The alternative method using landmark coordinates does not reflect this.

**Table 4** Mean difference, SD, Bland-Altman 95% limit of agreement, *p* values, and the ICC between the two methods in the nose region

| Measurements | Paired *t* test (manual–3D) | | | | | Bland-Altman | | | ICC | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean difference | SD | 95% CI | | | Mean bias | 95% limits of agreement | | *r* | 95% CI | | *p* value |
| | | | Lower | Upper | *p* value | | Lower | Upper | | Lower | Upper | |
| Nose height | − 0.01 | 0.99 | − 0.34 | 0.32 | 0.93 | − 0.01 | − 1.95 | 1.93 | 0.98 | 0.97 | 0.99 | 0.00* |
| Nose width | − 1.35 | 1.91 | − 1.99 | − 0.72 | 0.00# | − 1.35 | − 5.09 | 2.39 | 0.86 | 0.58 | 0.94 | 0.00* |
| Protrusion of nasal tip | 0.28 | 0.82 | − 0.23 | 0.52 | 0.08 | 0.25 | − 1.36 | 1.86 | 0.92 | 0.85 | 0.96 | 0.00* |
| Nose dorsum length | 0.36 | 1.42 | − 0.12 | 0.83 | 0.14 | 0.36 | − 2.43 | 3.15 | 0.95 | 0.91 | 0.97 | 0.00* |

#$p < 0.002$; *$p < 0.001$

Furthermore, linear measurements help to avoid the errors that can be made by landmark markings [31, 32]. In this study, landmarks were labeled by a liquid eyeliner, which produces a fine dot with a dimension of 0.5 mm as recommended by Kook et al. [3]. This will also reduce the error of landmark localization [6, 24].

Aynechi et al. [18] investigated the influence of landmark labeling on the accuracy and precision of an indirect facial anthropometric technique. Ten stereophotogrammetry images were obtained, one group with labeled landmarks and another without prior to image capturing. They found statistically significant differences in seven of the labeled landmarks and six in the unlabeled landmarks. The error was below 2 mm; hence, they concluded that the measurements obtained via the 3D imaging system have good agreement with the direct method of measurement with or without landmarks. Landmarks such as nasion and gnathion require palpation for identification. Thus, to reduce any bias in landmark placement on the 3D images, landmark labeling on the patients prior to image capturing was advocated.

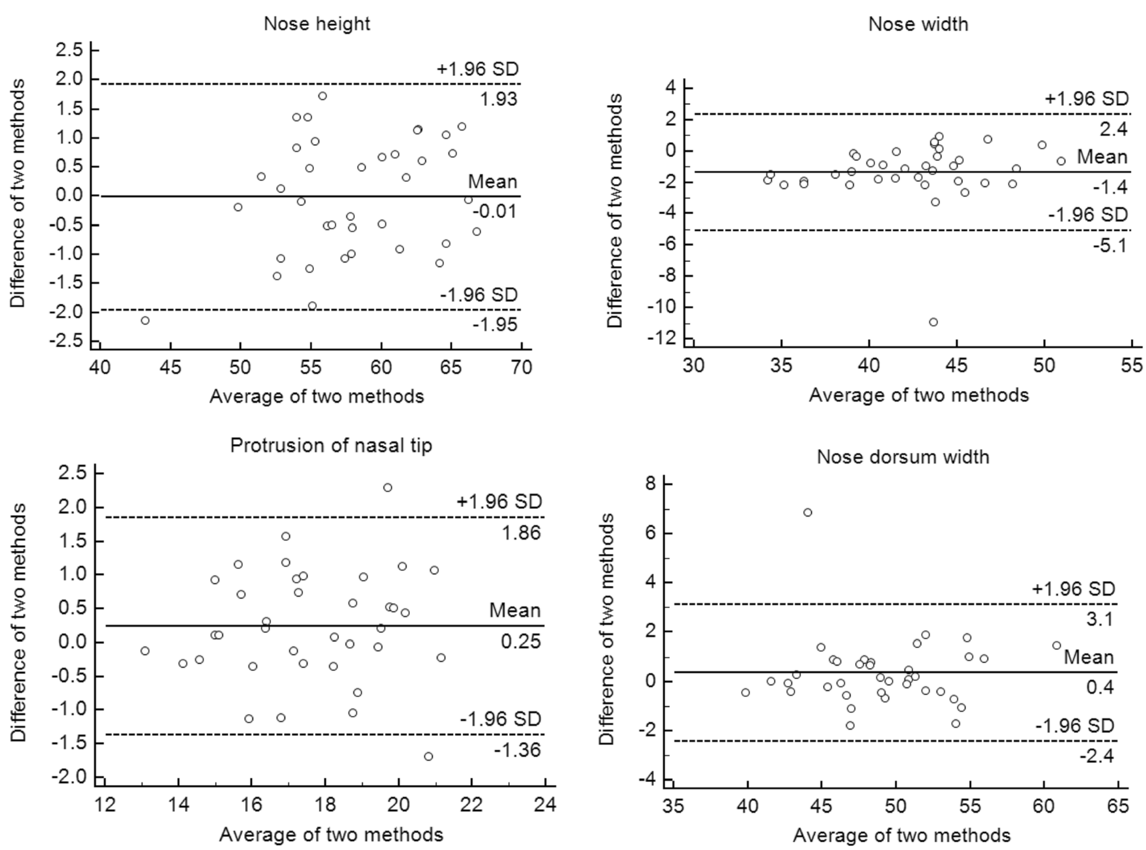


**Fig. 4** Bland-Altman plots of the nose region

**Table 5** Mean difference, SD, Bland-Altman 95% Limit of agreement, $p$ values, and the ICC between the two methods in the orolabial region

| Measurements | Paired $t$ test (manual–3D) | | | | | Bland-Altman | | | ICC | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean difference | SD | 95% CI | | $p$ value | Mean bias | 95% limits of agreement | | $r$ | 95% CI | | $p$ value |
| | | | Lower | Upper | | | Lower | Upper | | Lower | Upper | |
| Mouth width | − 0.28 | 1.06 | − 0.63 | 0.08 | 0.12 | − 0.28 | − 2.35 | 1.80 | 0.97 | 0.94 | 0.98 | 0.00* |
| Upper lip height | − 0.25 | 1.42 | − 0.72 | 0.22 | 0.29 | − 0.25 | − 3.03 | 2.54 | 0.91 | 0.85 | 0.96 | 0.00* |
| Cutaneous upper lip height | − 0.05 | 0.92 | − 0.35 | 0.26 | 0.76 | − 0.05 | − 1.85 | 1.76 | 0.93 | 0.92 | 0.98 | 0.00* |
| Upper vermillion height | − 1.02 | 1.08 | − 1.38 | − 0.66 | 0.00# | − 1.02 | − 3.14 | 1.09 | 0.87 | 0.45 | 0.96 | 0.00* |
| Lower vermillion height | 0.63 | 1.98 | − 0.03 | 1.29 | 0.06 | 0.63 | − 3.24 | 4.50 | 0.61 | 0.36 | 0.78 | 0.00* |
| Subalare width | − 0.06 | 0.95 | − 0.38 | 0.25 | 0.69 | − 0.06 | − 1.92 | 1.80 | 0.97 | 0.94 | 0.98 | 0.00* |
| Pronasale to left alar base | 0.31 | 1.29 | − 0.12 | 0.74 | 0.15 | 0.32 | − 2.21 | 2.84 | 0.91 | 0.82 | 0.95 | 0.00* |
| Pronasale to right alar base | 0.25 | 1.25 | − 0.17 | 0.66 | 0.24 | 0.25 | − 2.20 | 2.69 | 0.90 | 0.82 | 0.95 | 0.00* |
| Subnasale to left alar base | − 0.09 | 0.89 | − 0.39 | 0.21 | 0.54 | − 0.09 | − 1.82 | 1.65 | 0.91 | 0.83 | 0.95 | 0.00* |
| Subnasale to right alar base | − 0.23 | 0.86 | − 0.52 | 0.06 | 0.12 | − 0.23 | − 1.92 | 1.47 | 0.89 | 0.79 | 0.94 | 0.00* |
| Left upper lateral lip length | − 0.16 | 1.07 | − 0.52 | 0.20 | 0.38 | − 0.16 | − 2.26 | 1.94 | 0.93 | 0.86 | 0.96 | 0.00* |
| Right upper lateral lip length | − 0.19 | 1.04 | − 0.54 | 0.16 | 0.27 | − 0.19 | − 2.24 | 1.86 | 0.94 | 0.90 | 0.97 | 0.00* |

#$p < 0.002$; *$p < 0.001$

## Validation of the 3D VECTRA imaging system

The validity of the 3D VECTRA 360° imaging system using a mannequin head had been conducted by Metzler et al. [25]. The mannequin head was labeled with 52 landmarks, and 28 3D images were taken. They measured the distances between landmarks manually using a conventional caliper and compared it with the digitally calculated distances acquired from labeling by two independent observers. Non-significant differences were found between both groups, with a mean difference of 1.33 mm. In the present study, the descriptive data exhibited that 14 (74%) of the 3D measurements had larger values than the manual measurements, which is in accordance with Aynechi et al. [18] and Ghaddousi et al. [27].

Interestingly, there were no statistically significant differences in the linear measurements between methods in the facial regions which were presented by face height, upper face height, and lower face height. However, only the face height did not show clinically significant differences between the two methods. For measurements of the upper and lower face heights, the 95% limit of agreement indicated that the majority of the mean differences were not within 2 mm.

Localization of stomion, a landmark for upper face height, was challenging because this landmark was not labeled prior to direct or image acquisition, and the border of the stomion was affected by the severity of the cleft. Subnasale, which is the intersection between the nose base and upper lip, and a landmark for lower face height, was also affected by the severity of the cleft. Furthermore, some subjects have a nose that curves downwards, making landmark identifying and labeling difficult. The mean bias for face height was very small at 0.06 mm, while lower face height had a higher difference of 0.11 mm and upper face height had the most difference of 0.37 mm. Previous literature had not advocated Bland-Altman analysis widely; thus, direct comparison is difficult.

Nose height showed very minimal differences with no statistically or clinically significant differences. The nose width in the alar region was significantly different between the direct and 3D digitized methods. The mean difference obtained was 1.35 mm ± 1.91, which is much higher than what was reported by Dindaroğlu et al. [4] (0.007 mm) and Heike et al. [6] (0.3 mm). It could be recommended that there should be an adjustment of around 1.35 mm in the measurements taken via 3D images [33]. However, considering that clinical significance is 2 mm, this adjustment may not be necessary. The differences were clinically significant and generally large, with the highest difference detected to be 5.09 mm. The discrepancies might be due to the fact that alar is a landmark that has undercuts with dark holes which the 3D system may not be able to capture accurately [34]. The convexity of the alar region makes placing of the pointed caliper point difficult during direct measurement. Additionally, lack of rigidity of the soft tissue can cause distortion when the caliper's pointer is in contact with the dot [3, 21]. This is also reflected in the agreement during placement of the landmarks with the lowest ICC value of the nasal region of 0.86. The error can be minimized by marking the landmark prior to imaging and measuring and also by positioning the patient in a standardized manner [34]. The mean differences of the protrusion of the nasal tip and the dorsum nose were 0.36 mm and 0.25 mm respectively, which is in agreement with previous studies [3, 4, 6]. The differences in the protrusion of the nasal tip were within 2 mm, but the

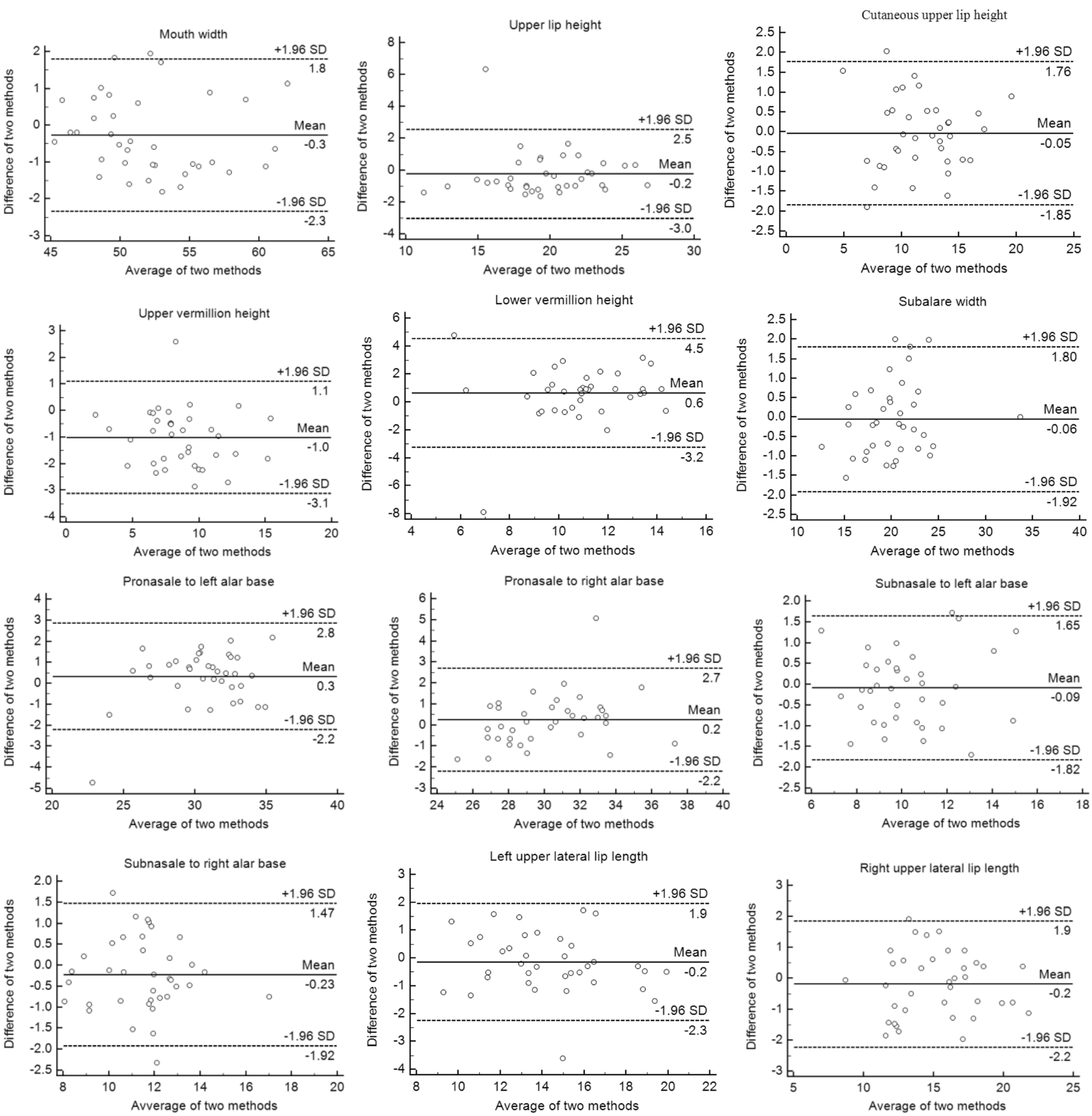


**Fig. 5** Bland-Altman plots of the orolabial region

nose dorsum length showed clinical significance. Extra caution should be employed in localization of the pronasale landmark, which was identified by viewing the patient in profile view. The angulation of the patient might affect the placement of this landmark. Light reflection may also influence digital identification of the landmarks [27]. In this current study, nose height showed a 95% limit of agreement between 1.95 and 1.93 mm, which is higher compared to the value reported by Dindaroğlu et al. [4] (− 1.14 mm to 0.96 mm). Most of the values were within the 95% limit of agreement. Similarly, the nasal tip could be incorrectly placed due to light reflections and the angulation of the image in profile view.

All of the linear landmarks measured in the orolabial region had a discrepancy below 2 mm in all of the measurements. Heike et al. [6] reported similar findings in mouth width, upper lip height, and upper and lower vermillion height, with the mean differences ranging from 0.6 to 1.3 mm. A significant difference was found in upper vermillion height with a mean difference of 1.02 mm ± 1.08 between the two methods. This area, especially the labrale superius, is affected by the severity of the cleft side

**Table 6** Mean difference, SD, *p* values, and ICC for the intra-observer

| Measurements | Paired *t* test (T1–T2) | | | | | ICC | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean difference | SD | 95% CI | | *p* value | | 95% CI | | *p* value |
| | | | Lower | Upper | | | Lower | Upper | |
| Face Height | 0.02 | 2.75 | − 1.95 | 1.99 | 0.98 | 0.94 | 0.76 | 0.98 | 0.00*** |
| Upper face height | − 0.26 | 1.74 | − 1.51 | 0.98 | 0.64 | 0.96 | 0.86 | 0.99 | 0.00*** |
| Lower face height | − 1.25 | 3.96 | − 4.08 | 1.59 | 0.35 | 0.65 | 0.11 | 0.90 | 0.02* |
| Nose height | − 0.41 | 2.68 | − 2.33 | 1.51 | 0.64 | 0.77 | 0.30 | 0.94 | 0.00** |
| Nose width | − 0.70 | 1.07 | − 1.47 | 0.06 | 0.07 | 0.95 | 0.79 | 0.99 | 0.00*** |
| Protrusion of nasal tip | − 0.71 | 1.06 | − 1.47 | 0.05 | 0.07§ | 0.61 | 0.05 | 0.88 | 0.01* |
| Nose dorsum length | 0.63 | 2.03 | − 0.82 | 2.08 | 0.35 | 0.88 | 0.61 | 0.97 | 0.00*** |
| Mouth width | − 0.48 | 2.34 | − 2.15 | 1.19 | 0.53 | 0.88 | 0.59 | 0.97 | 0.00*** |
| Upper lip height | − 0.13 | 2.03 | − 1.58 | 1.32 | 0.84 | 0.87 | 0.56 | 0.97 | 0.00*** |
| Cutaneous upper lip height | − 0.03 | 1.04 | − 0.77 | 0.71 | 0.93 | 0.96 | 0.85 | 0.99 | 0.00*** |
| Upper vermillion height | − 0.03 | 1.12 | − 0.83 | 0.78 | 0.95 | 0.92 | 0.71 | 0.98 | 0.00*** |
| Lower vermillion height | 0.22 | 1.48 | − 0.85 | 1.28 | 0.66 | 0.72 | 0.20 | 0.92 | 0.01** |
| Subalare width | 1.02 | 1.99 | − 0.40 | 2.25 | 0.11§ | 0.93 | 0.73 | 0.98 | 0.00*** |
| Pronasale to left alar base | 0.28 | 2.00 | − 1.15 | 1.72 | 0.67 | 0.85 | 0.50 | 0.96 | 0.00*** |
| Pronasale to right alar base | − 0.83 | 1.63 | − 2.00 | 0.34 | 0.14 | 0.65 | 0.12 | 0.90 | 0.01* |
| Subnasale to left alar base | 0.37 | 2.11 | − 1.14 | 1.87 | 0.60 | 0.38 | − 0.33 | 0.80 | 0.14 |
| Subnasale to right alar base | 0.24 | 0.89 | − 0.40 | 0.87 | 0.43 | 0.79 | 0.39 | 0.94 | 0.00** |
| Left upper lateral lip length | − 0.33 | 1.44 | − 1.36 | 0.70 | 0.49 | 0.89 | 0.65 | 0.72 | 0.00*** |
| Right upper lateral lip length | 0.77 | 1.98 | − 0.65 | 2.19 | 0.65§ | 0.85 | 0.52 | 0.96 | 0.00*** |

ICC *p* value = * < 0.05, ** < 0.01, *** < 0.001; § = Wilcoxon rank test was performed

and the fact that the stomion is unmarked [18]. Furthermore, this area consists of mobile soft tissues which could be affected by the posture of the lips that could occur during direct measurement or during image acquisition. Unfortunately, this type of error occurs in all types of anthropometry techniques and no technique to date could avoid such error completely [35]. A higher median difference of more than 2.50 mm was found in a study done by Ghoddousi et al. [27]. As no human being can maintain a natural facial expression, changes in facial expressions during direct measurement and image capturing will increase the chance of error [36]. The Bland-Altman analysis showed that only four out of 12 linear measurements of the orolabial region were within the 95% limit of agreement. Nonetheless, the mean bias in the measurements between methods in the orolabial region is less than 2 mm, being the highest difference in upper vermillion height of 1 mm. Thus, no adjustment to compensate for bias is necessary for the 3D measurements.

## Reproducibility of the 3D VECTRA imaging system

The second objective of this study was to evaluate the reproducibility of the digital measurements obtained using the VECTRA imaging system intra- and inter-operatively. Reproducibility has been referred to as the variation or consistency in measurements made on a subject under changing conditions [37] and has been defined as the closeness of successive measurements of the same object [38].

### Intra-observer reproducibility

Plooij et al. [26] investigated the reproducibility of multiple landmarks with the use of a stereophotography image using based reference frame and soft tissue landmarks. The mean differences in the alar region for the 3D reference frames were 0.99 mm (observer 1) and 0.67 mm (observer 2). By using the 3D coordinate system, it was reported that intra-observer error was 1.00 mm (observer 1) and 0.65 mm (observer 2). Though they employed a normal subject without a cleft and used different types of camera, interestingly, a similar finding was obtained in the current study by comparing the 3D reference frame difference whereby a difference of 0.70 mm was acquired for the alar region. Othman et al. [39] also investigated the reproducibility of 24 facial landmarks on a normal subject using the 3D VECTRA dual module camera. They found that for all the landmarks, the ICC ranged from 0.68 to 0.97 and there was no significant difference in locating the landmarks.

**Table 7** Mean difference, SD, $p$ values, and ICC for the inter-observer

| Measurements | Paired $t$ test (OP1–OP2) | | | | | ICC | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean difference | SD | 95% CI | | $p$ value | | 95% CI | | $p$ value |
| | | | Lower | Upper | | | Lower | Upper | |
| Face height | − 1.79 | 3.77 | − 4.49 | 0.90 | 0.17 | 0.85 | 0.53 | 0.96 | 0.00*** |
| Upper face height | 2.13 | 3.07 | − 0.07 | 4.32 | 0.06 | 0.81 | 0.34 | 0.95 | 0.00*** |
| Lower face height | − 3.31 | 4.46 | − 6.51 | − 0.12 | 0.04 | 0.51 | − 0.05 | 0.85 | 0.02* |
| Nose height | 0.77 | 3.59 | − 1.80 | 3.34 | 0.51 | 0.69 | 0.16 | 0.91 | 0.01* |
| Nose width | 2.62 | 2.93 | 0.53 | 4.72 | 0.02 | 0.68 | 0.05 | 0.92 | 0.00** |
| Protrusion of nasal tip | 0.95 | 1.74 | − 0.30 | 2.19 | 0.12 | 0.37 | − 0.18 | 0.78 | 0.10 |
| Nose dorsum length | − 0.71 | 3.39 | − 3.13 | 1.72 | 0.53 | 0.72 | 0.22 | 0.92 | 0.00** |
| Mouth width | 1.72 | 3.00 | − 0.42 | 3.87 | 0.10 | 0.82 | 0.43 | 0.95 | 0.00*** |
| Upper lip height | 0.66 | 1.79 | − 0.62 | 1.94 | 0.28 | 0.89 | 0.64 | 0.97 | 0.00*** |
| Cutaneous upper lip height | 1.54 | 1.36 | 0.57 | 2.51 | 0.01 | 0.83 | 0.11 | 0.96 | 0.00*** |
| Upper vermillion height | − 0.49 | 2.18 | − 2.05 | 1.07 | 0.50 | 0.71 | 0.19 | 0.92 | 0.00** |
| Lower vermillion height | 0.09 | 1.30 | − 0.84 | 1.02 | 0.83 | 0.77 | 0.29 | 0.94 | 0.00** |
| Subalare width | 0.51 | 2.89 | − 1.56 | 2.58 | 0.86§ | 0.88 | 0.59 | 0.97 | 0.00*** |
| Pronasale to left alar base | 2.47 | 2.75 | 0.50 | 4.44 | 0.02 | 0.53 | − 0.06 | 0.86 | 0.01* |
| Pronasale to right alar base | 1.52 | 3.32 | − 0.86 | 3.89 | 0.18 | 0.42 | − 0.16 | 0.81 | 0.08 |
| Subnasale to left alar base | 0.24 | 2.20 | − 1.34 | 1.81 | 0.74 | 0.34 | − 0.39 | 0.79 | 0.17 |
| Subnasale to right alar base | 0.66 | 1.25 | − 0.24 | 1.55 | 0.13 | 0.67 | 0.15 | 0.90 | 0.00** |
| Left upper lateral lip length | 0.59 | 2.04 | − 0.87 | 2.05 | 0.38 | 0.78 | 0.36 | 0.94 | 0.00** |
| Right upper lateral lip length | 1.76 | 1.76 | 0.50 | 3.02 | 0.01 | 0.83 | 0.18 | 0.96 | 0.00*** |

ICC $p$ value = * < 0.05, ** < 0.01, *** < 0.001; § = Wilcoxon rank test was performed

The most reproducible landmark was the nasion (ICC = 0.97; mean difference = 0.23 mm). In the present study, the ICC values with the linear distances involving the nasion (face height, upper face height, nose height, and nose dorsum length) ranged from 0.77 to 0.96. The slightly lower value observed may be due to the cleft subjects used in this study. Wong et al. [24] reported that the intra-observer agreement ranged from 0.66 to 0.99 with a grand mean of 0.91. They concluded that the inconsistency of landmark identification is due to a poorly defined upper philtral crest, presence of facial hair, and flash from the camera obscuring these landmarks.

Heike et al. [6] assessed the reliability of the 3dMDface System. The participants were children and adults who had abnormal facial characteristics, with and without 22q11.2 deletion syndrome. They found the intra-rater reliability ranged from 0.65 to 0.99. Their intra-rater reliability for the mouth region ranged from 0.94 to 0.98, indicating good agreement for the areas. In contrast with the recent study, the ICC values were lower, 0.56 to 0.96 for the mouth regions. In the nose region, both studies had similar acceptable to good agreement (0.65 to 0.99 for Heike et al.; 0.42 to 0.95 for the current study); however, it can be observed that the current study had slightly lower ICC values. The landmarks' reproducibility did not exceed the cutoff 2 mm clinical significance threshold; thus, it can be noted that the landmarks' placement error does not significantly affect the measurement for intra-observer reliability [34]. Most of the errors in measurement were made during the direct anthropometry and were affected by the ability and experience of the operator.

### Inter-observer reproducibility

The general assumption from the result obtained in this study is that the ICC values for reproducibility of the linear landmarks ranged from 0.34 to 0.89. Dindaroglu et al. [4] conducted a study evaluating the accuracy and reliability of the 3D stereophotogrammetry compared to the direct method. They found a very high ICC value of inter-observer reliability (0.90 to 0.99). Plooij et al. [26] on the other hand reported slightly low inter-observer agreement of 0.69 to 0.99; however, they utilized the reference frame and the 3D coordinates.

The upper and lower face heights have mean differences of 2.13 mm ± 3.07 and − 3.31 mm ± 4.46, respectively, which indicated clinically significant differences. These discrepancies could be attributed to the difficulty in identifying landmarks which involve palpation such as the nasion and gnathion [28]; however, these potential errors do not affect the outcome [27]. de Menezes et al. [28] also reported that the discrepancies in the measurements might be due to the digitization of unmarked landmarks. Inter-observer reliability in this study had similar findings with Heike et al. [6]. Baysal et al. [40] evaluated the intra- and inter-examiner reproducibility of the soft tissue landmarks using 3-DMD Face (3-DM TM Ltd, Atlanta, GA). The ICC ranges for intra- and inter-examiners were from 0.99 to 1.00 and 0.99 to 1.00, respectively. The identification of the landmarks by different raters showed a mean difference below 0.10 mm except for the left alar (0.11 mm). However, direct comparison to the current study is impossible because they used coordinates for landmark localization, while in this current study, linear measurements were used.

There are various papers that have evaluated the accuracy and reliability of 3D stereophotogrammetry systems. However, direct comparison is difficult due to the differences in the study design, measurement protocols, and statistical analyses [24]. Furthermore, there is a degree of human error when physical and digital measurements were made [21].

In summary, nose width and the upper vermillion height measured from the nose and the orolabial regions, respectively, exhibited slightly lower accuracy when measured using the VECTRA-M5 360° Imaging System (Fig. 6). This could be due to the fact that the alar is a landmark that has undercuts creating shadows that the 3D system is not able to capture accurately. The convexity of the alar region and the lack of rigidity of the soft tissue made it difficult to locate the landmarks. The labrale superius landmark is affected by the severity of the cleft side, which could affect the accuracy of the upper vermillion height. Furthermore, the mobile soft tissue in this area could also be affected by lip posture that could occur during the procedures.

Inter-observer reproducibility of the nasal tip protrusion was low and the subnasale to the left alar base had the lowest intra- and inter-observer reproducibility (Fig. 6). Both measurements are from the nose region where the area was affected by the cleft side. These discrepancies could be attributed to the difficulty in identifying the landmarks due to the convexity and distorted soft tissue from the surgical scar. An inexperienced operator could also contribute to the discrepancies.

The data of this interesting patient cohort used in the current study has potential values in the clinical setting. CLP patients' treatment protocol normally continues to adulthood, especially for those who need orthognathic surgery. To achieve a good treatment outcome, the characteristics of the defects, especially from the facial area, must be completely understood by the operator. Pre- and post-soft tissue surgical treatment assessment and evaluation of these patients are very important. On that account, data from the 3D soft tissue analysis of adult CLP patients could be used in the initial treatment planning phase and may be beneficial in the evaluation of the surgical outcome.

One of the limitations of the current study is that it relies on the operator's skills to manually identify landmarks, so accuracy may be affected by the operator's skills and exhaustion of frequently repeated measurements. Currently, automated landmark localization using curvature maps and profile analysis is in development, which has the potential to reduce inaccuracy caused by operator error [41]. Such a method is currently not easily accessible. However, it may potentially be used for future studies when commercially available.

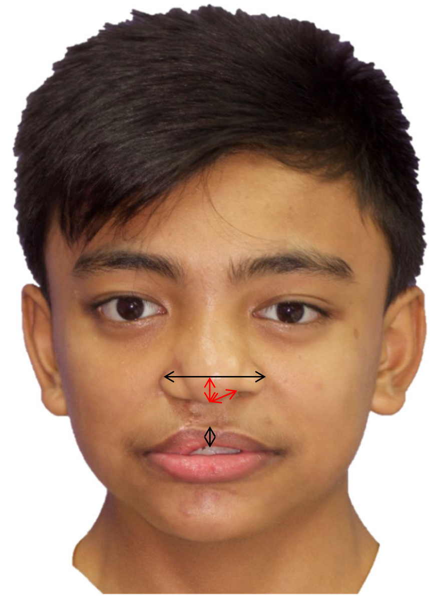

**Fig. 6** Red arrows—areas with low reproducibility (subnasale to the left alar and protrusion of the nasal tip). Black arrows—areas with low accuracy (nose width and the upper vermillion height)

## Conclusions

The validation and the reproducibility of the VECTRA-M5 360° Imaging System on CLP subjects have been analyzed and compared with direct anthropometry linear measurements. From the results obtained, most of the linear measurements show no statistically significant differences between the proposed method and direct anthropometry linear measurements. Nevertheless, bias of the 3D imaging system is present in the linear measurements of the nose width and the upper vermillion height. The measurements' mean biases were within 2 mm, but the 95% limit of agreement was more than 2 mm. Intra- and inter-observer measurements generally showed good reproducibility. Four inter-observer measurements were clinically significant; they were the upper and lower face heights, nose width, and pronasale to left alar base. It can be concluded that the measurements obtained from the 3D imaging system are valid and reproducible for evaluating adult CLP patients and are suitable to be used in the clinical setting. However, training of the operator is strictly advisable. In addition, extra attention and caution must be taken when measuring the areas around the nose and the upper lips, specifically the subnasale to the left alar, protrusion of the nasal tip, nose width, and the upper vermillion height.

**Acknowledgments** We would like to thank Miss Najihah Lokman for her assistance in statistical analysis, Miss Roshahida Ahmad for her expertise in 3D imaging, and Madam Zuraini Ghazali, a representative from Cleft Lip and Palate Association of Malaysia (CLAPAM), who contributed to the enrolment of potential subjects.

**Funding** This research was supported by the Postgraduate Research Fund by Coursework, Faculty of Dentistry, University of Malaya (PPPC/C1-2015/DGD/21), and by the University of Malaya High Impact Research Grant (UM.C/625/1/HIR/MOHE/DENT/13).

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflicts of interest.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards (IRB Reference Number: DF CD1410/0086(P)).

**Informed consent** Informed consent was obtained from all individual participants included in the study.

## References

1. Othman SA, Ahmad R, Asi SM, Ismail NH, Rahman ZAA (2014) Three-dimensional quantitative evaluation of facial morphology in adults with unilateral cleft lip and palate, and patients without clefts. Br J Oral Maxillofac Surg 52:208–213
2. Krimmel M, Kluba S, Bacher M, Dietz K, Reinert S (2006) Digital surface photogrammetry for anthropometric analysis of the cleft infant face. Cleft Palate Craniofac J 43:350–355
3. Koo MS, Jung S, Park H, Oh HK, Ryu SY, Cho JH, Shin HK (2014) A comparison study of different facial soft tissue analysis methods. J Craniomaxillofac Surg 42:648–656
4. Dindaroğlu F, Kutlu P, Duran GS, Görgülü S, Aslan E (2016) Accuracy and reliability of 3D stereophotogrammetry: a comparison to direct anthropometry and 2D photogrammetry. Angle Orthod 86:487–494
5. Fourie Z, Damstra J, Gerrits PO, Ren Y (2011) Evaluation of anthropometric accuracy and reliability using different three-dimensional scanning systems. Forensic Sci Int 207:127–134
6. Heike CL, Cunningham ML, Hing AV, Stuhaug E, Starr JR (2009) Picture perfect? Reliability of craniofacial anthropometry using three-dimensional digital stereophotogrammetry. Plast Reconstr Surg 124:1261–1272
7. Ort R, Metzler P, Kruse AL, Matthews F, Zemann W, Grätz KW, Luebbers HT (2012) The reliability of a three-dimensional photo system-(3dMDface-) based evaluation of the face in cleft lip infants. Plast Surg Int. https://doi.org/10.1155/2012/138090
8. Brons S, Darroudi A, Nada R, Bronkhorst EM, Vreeken R, Berge SJ, Maal T, Kuijpers-Jagtman AM (2018) Influence of involuntary facial expression on reproducibility of 3D stereophotogrammetry in children with and without complete unilateral cleft lip and palate from 3 to 18 months of age. Clin Oral Investig 23:1041–1050. https://doi.org/10.1007/s00784-018-2520-0
9. Tse R, Booth L, Keys K, Saltzman B, Stuhaug E, Kapadia H, Heike C (2014) Reliability of nasolabial anthropometric measures using three-dimensional stereophotogrammetry in infants with unrepaired unilateral cleft lip. Plast Reconstr Surg 133:530e–542e
10. Khambay B, Nairn N, Bell A, Miller J, Bowman A, Ayoub AF (2008) Validation and reproducibility of a high-resolution three-dimensional facial imaging system. Br J Oral Maxillofac Surg 46: 27–32
11. Meyer-Marcotty P, Alpers GW, Gerdes ABM, Stellzig-Eisenhauer A (2010) Impact of facial asymmetry in visual perception: a 3-dimensional data analysis. Am J Orthod Dentofac Orthop 137: 168.e1–168.e8
12. Faul F, Erdfelder E, Lang AG, Buchner A (2007) G* Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 39:175–191
13. Asi SM, Ismail NH, Rahman ZA (2012) Validity and reliability evaluation of data acquisition using Vectra 3D compare to direct method. In Biomedical Engineering and Sciences (IECBES), 2012 IEEE EMBS Conference on (pp. 883-887). IEEE
14. Zreaqat M, Hassan R, Halim AS (2012) Facial dimensions of Malay children with repaired unilateral cleft lip and palate: a three dimensional analysis. Int J Oral Maxillofac Surg 41:783–788
15. Walter SD, Eliasziw M, Donner A (1998) Sample size and optimal designs for reliability studies. Stat Med 17:101–110
16. Roberts CT, Richmond S (1997) The design and analysis of reliability studies for the use of epidemiological and audit indices in orthodontics. Br J Orthod 24:139–147
17. Urbaniak GC, Plous S (2013) Research Randomizer (Version 4.0) [Computer software]. Retrieved on June 22, 2013
18. Aynechi N, Larson BE, Leon-Salazar V, Beiraghi S (2011) Accuracy and precision of a 3D anthropometric facial analysis with and without landmark labeling before image acquisition. Angle Orthod 81:245–252
19. Farkas LG, Hajniš K, Posnick JC (1993) Anthropometric and anthroposcopic findings of the nasal and facial region in cleft patients before and after primary lip and palate repair. Cleft Palate Craniofac J 30:1–12
20. Weber DW, Fallis DW, Packer MD (2013) Three-dimensional reproducibility of natural head position. Am J Orthod Dentofacial Orthop 143:738–744
21. Winder RJ, Darvann TA, McKnight W, Magee JD, Ramsay-Baggs P (2008) Technical validation of the Di3D stereophotogrammetry surface imaging system. Br J Oral Maxillofac Surg 46:33–37
22. Lee JY, Han Q, Trotman CA (2004) Three-dimensional facial imaging: accuracy and considerations for clinical applications in orthodontics. Angle Orthod 74:587–593
23. Weinberg SM, Naidoo S, Govier DP, Martin RA, Kane AA, Marazita ML (2006) 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 17:477–483
24. Wong JY, Oh AK, Ohta E, Hunt AT, Rogers GF, Mulliken JB, Deutsch CK (2008) Validity and reliability of craniofacial anthropometric measurement of 3D digital photogrammetric images. Cleft Palate Craniofac J 45:232–239
25. Metzler P, Sun Y, Zemann W, Bartella A, Lehner M, Obwegeser JA, Lübbers HT (2014) Validity of the 3D VECTRA photogrammetric surface imaging system for cranio-maxillofacial anthropometric measurements. J Oral Maxillofac Surg 18:297–304
26. Plooij JM, Swennen GRJ, Rangel FA, Maal TJJ, Schutyser FAC, Bronkhorst EM, Berge SJ (2009) Evaluation of reproducibility and reliability of 3D soft tissue analysis using 3D stereophotogrammetry. Int J Oral Maxillofac Surg 38:267–273
27. Ghoddousi H, Edler R, Haers P, Wertheim D, Greenhill D (2007) Comparison of three methods of facial measurement. Int J Oral Maxillofac Surg 36:250–258
28. de Menezes M, Rosati R, Ferrario VF, Sforza C (2010) Accuracy and reproducibility of a 3-dimensional stereophotogrammetric imaging system. J Oral Maxillofac Surg 68:2129–2135

29. Aksu M, Kaya D, Kocadereli I (2010) Reliability of reference distance used in photogrammetry. Angle Orthod 80:670–677
30. Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. Stat Methods Med Res 8:135–160
31. Ngeow WC, Aljunid ST (2009) Craniofacial anthropometric norms of Malays. Singap Med J 50:525–528
32. Aldridge K, Boyadjiev SA, Capone GT, DeLeon VB, Richtsmeier JT (2005) Precision and error of three-dimensional phenotypic measures acquired from 3dMD photogrammetric images. Am J Med Genet A 15:247–253
33. Bland JM, Altman D (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 327: 307–310
34. Van Loon B, Maal TJ, Plooij JM, Ingels KJ, Borstlap WA, Kuijpers-Jagtman AM, Berge SJ (2010) 3D Stereophotogrammetric assessment of pre-and postoperative volumetric changes in the cleft lip and palate nose. Int J Oral Maxillofac Surg 39:534–540
35. Jayaratne YS, Deutsch CK, Zwahlen RA (2013) A 3D anthropometric analysis of the orolabial region in Chinese young adults. Bri J Oral Maxillofac Surg 51:908–912
36. Hajeer MY, Millett DT, Ayoub AF, Siebert JP (2004) Current products and practices: applications of 3D imaging in orthodontics: part 1. J Orthod 31:62–70
37. Bartlett JW, Frost C (2008) Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. Ultrasound Obstet Gynecol 31:466–475
38. Houstan WJ (1983) The analysis of errors in orthodontic measurements. Am J Orthod 83:382–390
39. Othman SA, Ahmad R, Merican AF, Jamaludin M (2013) Reproducibility of facial soft tissue landmarks on facial images captured on a 3D camera. Aust Orthod J 29:58–65
40. Baysal A, Sahan AO, Ozturk MA, Uysal T (2016) Reproducibility and reliability of three-dimensional soft tissue landmark identification using three-dimensional stereophotogrammetry. Angle Orthod 86:1004–1009
41. Lippold C, Liu X, Wangdo K, Drerup B, Schreiber K, Kirschneck C, Tatjana M, Danesh G (2014) Facial landmark localization by curvature maps and profile analysis. Head Face Med 10:54. https://doi.org/10.1186/1746-160X-10-54


**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.