ISAPS

ORIGINAL ARTICLES   **FACIAL SURGERY**

Development and Validation of a Comprehensive Perioral Evaluation Method Using Three-Dimensional Stereophotogrammetry

Yuyan Yang[1] · Yarong Chi[2] · Lin Jin[1] · Yuming Chong[1] · Xiao Long[1] · Ang Zeng[1] · Nanze Yu[1] · Xiaojun Wang[1]

Abstract

Background The non-invasive three-dimensional (3D) stereophotogrammetry is widely used in anthropometry for medical purpose. Yet, few studies have assessed its reliability on measuring the perioral region.

Objectives This study aimed to provide a standardized 3D anthropometric protocol for the perioral region.

Methods 38 female and 12 male Asians were recruited (mean age 31.6 ± 9.6 years). Two sets of 3D images using the VECTRA 3D imaging system were acquired for each subject, and two measurement sessions for each image were performed independently by two raters. 25 landmarks were identified, and 28 linear, 2 curvilinear, 9 angular and 4 areal measurements were evaluated for intrarater, interrater, and intramethod reliability.

Results Our results showed high reliability of 3D imaging-based perioral anthropometry by mean absolute difference (0.57 and 0.57 unit), technical error measurement (0.51 and 0.55 unit), relative error of measurement (2.18% and 2.44%), relative technical error of measurement (2.02% and 2.34%), and intraclass correlation coefficient (0.98 and 0.98) for intrarater 1 and intrarater 2 reliability; respectively 0.78 unit, 0.74 unit, 3.26%, 3.06% and 0.97 for interrater reliability; and respectively 1.01 unit, 0.97 unit, 4.74%, 4.57% and 0.95 for intramethod reliability.

Conclusions This standardized protocol utilizing 3D surface imaging technologies are feasible and highly reliable in perioral assessment. It could be further applied for diagnostic purpose, surgical planning and therapeutic effect evaluation in clinical practice in relation to perioral morphologies.

Level of Evidence IV This journal requires that authors assign a level of evidence to each article. For a full description of these Evidence-Based Medicine ratings, please refer to the Table of Contents or the online Instructions to Authors www.springer.com/00266.

Yuyan Yang and Yarong Chi contributed equally to this article.

✉ Nanze Yu
yunanze@pumch.cn

✉ Xiaojun Wang
pumchwxj@163.com

1   Department of Plastic and Reconstructive Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Dongcheng District, Beijing 100730, People's Republic of China

2   Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, People's Republic of China

Introduction

The lips represent an important part of the anthropometric features of the human face [1] and play crucial roles in vocalization, mastication and emotional expression [2]. Given that, there is a growing interest in plastic and reconstructive surgeries and aesthetic procedures of lips. A comprehensive characterization of lip morphology, including lip length and thickness, lip contouring as well as facial proportions is the premise of obtaining a harmonious and aesthetically appealing result.

Early research on lip anthropometry was mainly based on classic direct anthropometry or two-dimensional (2D) imaging tools such as standardized photographs [3, 4]. However, manual measurement requires physical contact

between the evaluator and the patient and the accuracy and precision are not guaranteed. 2D photos have major drawbacks either, including their inability to evaluate curvatures, areas and volumes, requirement of patient compliance as well as time consuming [5]. Three-dimensional (3D) imaging systems such as reconstructive computerized tomography (CT) and magnetic resonance imaging (MRI) were then utilized in lip anthropometry but CT exposes patients to excessive radiation and both of them introduced errors related to different body positions (i.e. standing or lying) and were timely and costly.

The emerging non-invasive 3D surface imaging technology has overcome above obstacles. Its advantages further lie in its acquisition speed and quickly reconstructing the 3D surface morphology, and its straightforward image presentation is particularly handy for highlighting slight asymmetries or contour defects that exist pre-operatively, which is extremely helpful during patient consultation. Thus this readily available and promising tool has been widely used in morphology assessment of facial parts such as the eyes and lips [6–9]. Weinberg et al. have established a 3D Facial Norms database, which consists of 3D craniofacial anthropometric normative data from 2454 European Caucasians ranging from 3 to 40 years old [10]. However, like many other perioral anthropometry studies [11–14], only several traditional perioral surface landmarks were identified and thus the corresponding measurements were limited. Furthermore, although it has been proven highly reliable in evaluating periocular region [6], no solid validation of the reliability and accuracy of 3D surface imaging technology in perioral anthropometry has been conducted.

Therefore, we proposed and verified a practical protocol for perioral morphology assessment with a 3D surface imaging system in an Asian population. We introduced a variety of novel perioral soft-tissue landmarks to ensure standardized and adequate perioral surface coverage, we included linear distance, curvatures, areas and angular measurements, and validated their intrarater, interrater and intramethod reliability.

Methods

Patients

A total of fifty healthy Asian individuals,12 males (mean age 32.1 ± 10.2 years; range 24–60 years) and 38 females (mean age 31.4 ± 9.4 years; range 19–57 years), were recruited from the Department of Plastic and Reconstructive Surgery, Peking Union Medical College Hospital between Apr, 2022 and Sep, 2022. Exclusion criteria include facial pathologies, congenital deformations, recent or previous trauma, procedures or surgeries influencing the perioral morphology. All participants provided written consent. This study was performed in line with the 1964 Declaration of Helsinki and its later amendments and was approved by the Institutional Review Board of the Peking Union Medical College Hospital (Reference Number I-22PJ693).

Equipment

The commercially available VECTRA H2 handheld camera (Canfield Scientific, Inc., Parsippany, NJ, USA) is a three-dimensional stereophotogrammetry system. For each set of facial imaging, three images were captured at a fixed distance and three angles (right oblique, frontal and left oblique) to ensure overlapping fields of view. Subsequently, a 3D model of the subject's face was automatically generated from these captures. The image acquisition time is 2 milliseconds per capture and the 3D model synthesis takes less than 1 min. The manufacturer claims a high geometry resolution of 1.2 mm. The 3D files were then exported in the OBJ format and then imported into the Geomagic Wrap 2017 (Geomagic, Inc., Research Triangle Park, NC, USA) for landmark identification, measurement and analysis.

### Three-Dimensional Image Acquisition

Before image acquisition, each subject was asked to wash the face with cleansing foam and fully expose the face by pulling back the hair, shaving off the beard and removing glasses and any jewelry. Participants were seated upright in a chair with a neutral facial expression. Their teeth and lips were gently closed and their eyes were gazing forward. An experienced operator performed all images acquisition according to the manufacturer's instructions in the same room under uniform lighting condition. Two sets of 3D images (Capture 1 and Capture 2) were taken for each participant at a time interval of at least 45 minutes with recalibration of the VECTRA device.

Landmark Identification and Perioral Measurement

A unified coordinate system was established in each 3D model according to a previous study [13]. Then the perioral landmarks were digitally identified (Table 1) and marked following a standard protocol (Fig. 1). In brief, landmarks easily identified around the nose, philtrum and on the middle line of lips and chin were located first. Landmarks were then identified to trisect the upper vermillion border between the cheilion and crista philtri on the left and right side, respectively. Then, labial landmarks were located vertically to them through the coordinate axes. Finally,

Table 1 Definitions of 25 anthropometric perioral landmarks

Landmark	Definition
Prn	Pronasale, the most prominent point on apex of nose
Cm	Columella, the most anterior and inferior point on apex of nose
Sn	Subnasale, the mid-point at the junction of the columella and the upper lip
Mc	The mid-point of the nasal columella crest on either side
Sbal	Subalare, the point of the lower inner limit of the alar base
Mph	The mid-point of the philtrum column
Cph	Crista philtri, peak of the Cupid's bow
Ls	Labial superius, the mid-point of the Cupid's bow
St	Stomion, the mid-point of the oral fissure when the lips are closed naturally
Li	Labial inferiors, the mid-point of the lower vermilion margin
Sl	Sublabiale, the mid-point of the mentolabial sulcus
Pog	Pogonion, the most anterior point on the soft tissue contour of the chin
Ch	Cheilion, corners of the mouth where the upper and lower vermilions meet
Lst	Lateral stomion, the point vertical to Cph at the oral fissure
Lst′	Lateral stomion, the point vertical to Lst at the lower vermilion margin
Mvm	The medial one thirds of the upper vermilion margin
FMvm	Point vertical to Mvm at the oral fissure
Mvm′	Point vertical to Mvm at the lower vermilion margin
Lvm	The lateral one thirds of the upper vermilion margin
FLvm	Point vertical to Lvm at the oral fissure
Lvm′	Point vertical to Lvm at the lower vermilion margin
Mt	The most protruded point of the median tubercle of upper lip
Pmt	The most protruded point of the paramedian tubercle of lower lip
Arc	Apex of upper arches of the upper lip
Lth	Lateral thickening of the upper lip
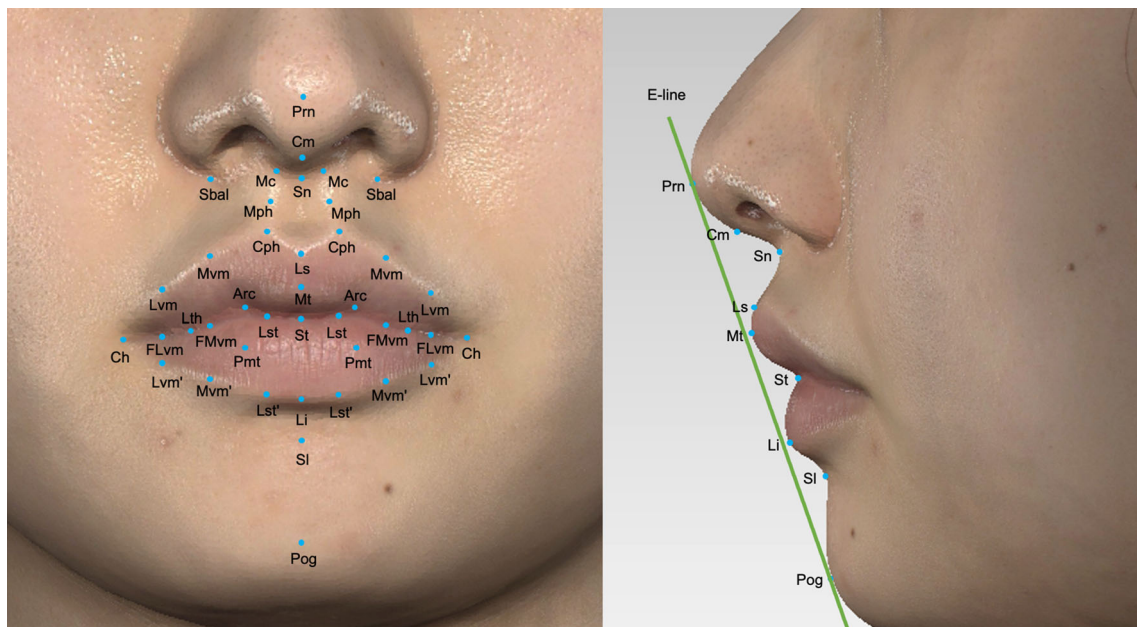
Fig. 1 The 41 three-dimensional anthropometric landmarks of the perioral region. 9 landmarks on the middle line and 16 landmarks on each side of the lip were indicated on the frontal view, and the 9 landmarks on the middle line and E-line were shown on the lateral view

upper and lower lip tubercles as well as apex of upper arches and lateral thickening of the upper lip were identified. Ricketts' E-line was drawn from the tip of the nose (pronasale) to the most anterior soft tissue point of the chin (pogonion). Subsequently, 28 linear distances, 2 curvatures, 4 areas and 9 angles were measured according to these landmarks (Table 2 and Fig. 2). Among the 28 linear distances, 6 were assessed on the lateral view as the horizontal distances of landmarks to E-line. And positive values were assigned to positions anterior to the E-line and negative values to positions posterior to the E-line.

Reliability Assessment

The first author (Yuyan Yang, rater 1) performed two measurement sessions (session 1 and session 2) with a time interval of at least 24 h for each of the two sets of 3D images. The third author (Lin Jin, rater 2) performed two measurement sessions with a time interval of at least 24 h for the first set of 3D images. For intrarater reliability (repeatability), the two measurements of the same set of 3D image by the same rater were compared. For interrater reliability (reproducibility), the second measurements of the same set of 3D image by rater 1 and rater 2 were compared. And for intramethod reliability, the second measurements of the two sets of 3D image were compared (Fig. 3).

Data Analyses

Seven statistics were evaluated to assess the reliability (Table 3) [6]. The intraclass correlation coefficient (ICC) indicates the reliability between repeated measures and represents a high reliability when close to 1 and a low reliability when close to 0. Generally, an ICC value of greater than 0.75 indicates excellent agreement [15, 16]. For the mean absolute difference (MAD) and technical error of measurement (TEM), we defined an acceptable error threshold of 1 unit (millimeter) for linear distances and curvatures because of the relatively small magnitude of perioral measurement, and an acceptable error threshold of 4 unit (square millimeter) for areas and 2 unit (degrees) for angles due to the relatively big magnitude of perioral measurement. In consideration of the influence of sample size on reliability assessment, we also calculated the relative error measurement (REM) and relative TEM (%TEM). According to previous reports, five reliability categories were defined: < 1%, excellent; 1–3.9%, very good; 4–6.9%, good; 7–9.9%, moderate; and > 10%, poor [17]. Total TEM and relative total TEM (% total TEM) were calculated to accommodate the influence from different raters when more than one rater was involved.

For statistical analysis, means and standard deviations (SDs) as well as MAD, TEM, REM, % TEM, total TEM and % total TEM were calculated using the software Microsoft Excel 2022 (Microsoft Corp., Redmond, WA, USA), and ICC using the software SPSS version 25 (IBM Corp., Armonk, NY, USA). Graphs were generated using the software GraphPad Prism 8 (GraphPad Software Inc., San Diego, CA, USA). For normally distributed measurements, statistic difference was assessed by paired-sample t tests, and for non-normally distributed measurements, Wilcoxon signed rank tests were performed. A P value of < 0.05 was considered statistically significant.

Results

General Results

Fifty healthy individuals with a mean age of 31.6 years (range 19–60 years) were recruited, of which 12 (24%) were men and 38 (76%) were women. Descriptive statistics (means and standard deviations, SDs) of all perioral measurements are shown in Table 4. Intra- and inter-rater reliability as well as intramethod reliability assessed by MAD, TEM, REM, %TEM, total TEM, % total TEM and ICC were shown in Supplementary Figure 1, 2, 3, 4, 5 and Table 5, respectively. In brief, the majority of perioral measurements had high reliability with an ICC estimate larger than 0.95, a MAD and a TEM estimate smaller than our defined limits and a REM and a %TEM estimate less than 7%.

Intrarater Reliability

Vast majority of the intrarater ICC estimates for all perioral measurements were larger than or equal to 0.95, except for SULW, SPW, PCW, PW and LVMLH in one or both raters, which still were larger than 0.90.

All intrarater MAD and TEM estimates were smaller than our defined limits—1 unit for linear distances and curvatures, 4 unit for areas and 2 unit for angles. Among the intrarater REM estimates, 2 curvatures (UVML and LVML), 2 linear distances (HSNE and HSTE), all areal measurements (SACUL, SALV, SAUV and SAPH) and 6 of the angles (MLSA, NA, CBA, ULVA, LSA and CBA′) had excellent reliability of less than 1%. 21 linear distances and the remaining 3 angles including ULA, LLA and CBAA had very good reliability (1–3.9%). And 5 linear distances including HLSE, CLLH, LVMLH, PCW and SPW had good reliability (4–6.9%). Similarly, of all intrarater %TEM estimates, 2 curvatures (UVML and LVML), 3 linear distances (HSNE, LW and HSTE),

all areal measurements (SACUL, SALV, SAUV and SAPH) and 6 of the angles (MLSA, ULVA, NA, CBA, LSA and CBA′) had excellent reliability. 23 linear distances and the remaining 3 angles had very good reliability and only 2 linear distances (CLLH and SPW) had good reliability.

Interrater Reliability

Most interrater ICC estimates were larger than or equal to 0.95, except for SULW, SPW, PW, UVH, LLH, UVMLH, UVMMH, ULST and LW ranging from 0.91 to 0.94 and the lowest ICC value 0.83 for PCW.

Table 2 List of 43 perioral measurements

Measurement	Abbreviation	Definition
Linear distance		
Frontal view		
Superior upper lip width	SULW	Sbal (left)–Sbal (right)
Superior philtrum width	SPW	Mc (left)–Mc (right)
Philtrum column width	PCW	Mph (left)–Mph (right)
Philtrum width	PW	Cph (left)–Cph (right)
Philtrum height	PH	Sn–Cph
Upper lip height	ULH	Sn–St
Cutaneous upper lip height	CULH	Sn–Ls
Upper vermilion height	UVH	Ls–St
Lower lip height	LLH	St–Sl
Cutaneous lower lip height	CLLH	Li–Sl
Lower vermilion height	LVH	St–Li
Upper vermilion margin lateral height	UVMLH	Lvm–FLvm
Lower vermilion margin lateral height	LVMLH	FLvm–Lvm′
Upper vermilion margin medial height	UVMMH	Mvm–FMvm
Lower vermilion margin medial height	LVMMH	FMvm–Mvm′
Upper lateral stomion height	ULST	Cph–Lst
Lower lateral stomion height	LLST	Lst–Lst′
Apex of upper arches width	AUAW	Arc (left)–Arc (right)
Lateral thickening width	LTHW	Lth (left)–Lth (right)
Paramedian tubercle width	PMTW	Pmt (left)–Pmt (right)
Lip width	LW	Ch (left)–Ch (right)
Lip thickness	LT	Ls–Li
Lateral view		
Horizontal distance of Sn to E-line	HSNE	Sn–(Prn–Pog)
Horizontal distance of Ls to E-line	HLSE	Ls–(Prn–Pog)
Horizontal distance of Mt to E-line	HMTE	Mt–(Prn–Pog)
Horizontal distance of St to E-line	HSTE	St–(Prn–Pog)
Horizontal distance of Li to E-line	HLIE	Li–(Prn–Pog)
Horizontal distance of Sl to E-line	HSLE	Sl–(Prn–Pog)
Curvature		
Upper vermilion margin length	UVML	Ch (left)–Lvm (left)–Mvm (left)–Cph (left)–Ls–Cph (right)–Mvm (right)–Lvm (right)–Ch (right)
Lower vermilion margin length	LVML	Ch (left)–Lvm′ (left)–Mvm′ (left)–Lst′ (left)–Li–Lst′ (right)–Mvm′ (right)–Lvm′ (right)–Ch (right)
Area		
Surface area of philtrum	SAPH	Mc (right)–Mph (right)–Cph (right)–Ls–Cph (left)–Mph (left)–Mc (left)

Table 2 continued

Measurement	Abbreviation	Definition
Surface area of upper vermilion	SAUV	Ch (left)–Lvm (left)–Mvm (left)–Cph (left)–Ls–Cph (right)–Mvm (right)–Lvm (right)–Ch (right)–FLvm (right)–FMvm (right)–Lst (right)–St–Lst (left)–FMvm (left)–FLvm (left)
Surface area of lower vermilion	SALV	Ch (right)–FLvm (right)–FMvm (right)–Lst (right)–St–Lst (left)–FMvm (left)–FLvm (left)–Ch (left)–Lvm′ (left)–Mvm′ (left)–Lst′ (left)–Li–Lst′ (right)–Mvm′ (right)–Lvm′ (right)
Surface area of cutaneous upper lip	SACUL	Ch (left)–Lvm (left)–Mvm (left)–Cph (left)–Ls–Cph (right)–Mvm (right)–Lvm (right)–Ch (right)–Sbal (right)–Mc (right)–Sn–Mc (left)–Sbal (left)
Angle		
Frontal view		
Lip spread angle	LSA	Ch (left)–Ns–Ch (right)
Cupid's bow angle	CBA	Ls–Cph–Mvm
Central bow angle	CBA′	Cph (left)–Ls–Cph (right)
Cupid's bow arm angle	CBAA	Cph–St–Ls
Upper lip angle	ULA	Lvm–Ch–FLvm
Lower lip angle	LLA	FLvm–Ch–Lvm′
Lateral view		
Nasolabial angle	NA	Cm–Sn–Ls
Upper and lower vermilion angle	ULVA	Ls–St–Li
Mentolabial sulcus angle	MLSA	Li–Sl–Pog

Vast majority of interrater MAD and TEM estimates were smaller than our defined limits, except for LW (MAD = 1.11 mm). Similarly, most measurements had high reliabilities in term of total TEM, with the exception of LW (total TEM = 1.10 mm), CBA′ (total TEM = 2.24°) and LLA (total TEM = 2.37°).

As for interrater REM estimates, 2 curvatures (UVML and LVML), 3 areas (SACUL, SALV and SAUV) and 2 angles (NA and MLSA) had a REM of less than 1%. 14 linear distances, SAPH and 6 angles (ULVA, CBA, LSA, CBAA, LLA and ULA) had a REM between 1 and 3.9% and 13 linear distances and CBA′ had a REM between 4 and 6.9%. PCW had the lowest but still moderate reliability (REM = 7.40%). The results of interrater %TEM estimates were similar. 2 curvatures (UVML and LVML), 3 areas (SACUL, SALV and SAUV) and 3 angles (NA, MLSA and ULVA) had excellent reliability. 17 linear distances, SAPH and 3 angles (CBA, LSA and CBAA) had very good reliability, the remaining 11 linear distances and 3 angles (ULA, LLA and CBA′) had good reliability. Furthermore, most % total TEM estimates were less than 7%, with the exception of LVMLH (% total TEM = 7.12%), PCW (% total TEM = 7.17%), HLSE (% total TEM = 7.27%), UVMLH (% total TEM = 7.44%),

HMTE (% total TEM = 7.58%) and CBA′ (% total TEM = 8.74).

Intramethod Reliability

16 of the 43 perioral measurements had an intramethod ICC estimates larger than or equal to 0.95, 13 were between 0.90–0.94 and 3 were below 0.90 (PCW, ICC = 0.88; PW, ICC = 0.89; MLSA, ICC = 0.89).

The intramethod MAD and TEM estimates were less than our defined limits in vast majority of perioral measurements, except for LW (MAD = 1.01 mm), ULVA (MAD = 2.32°, TEM = 2.03°), CBA′ (MAD = 2.37°, TEM = 2.43°) and MLSA (MAD = 3.12°, TEM = 3.10°). 3 (SACUL, SALV and SAUV) of the 43 perioral measurements had an intramethod REM of less than 1%, 17 were between 1-3.9%, 18 were between 4–6.9%, 1 was between 7–9.9% and 4 were larger than 10% (HSTE, HLSE, HLIE and HMTE). As for intramethod %TEM, all 4 areal measurements were less than 1%, 2 curvatures, 9 linear distances and 7 angles were between 1–3.9%, 14 linear distances and 2 angles were between 4–6.9%, 1 was between 7–9.9%, and similar to intramethod REM, 4 linear distances were larger than 10% (HSTE, HLSE, HSLE, HMTE).

Fig. 2 Perioral measurements.
a Linear distances on the frontal view (left) and the lateral view (right). **b** Curvatures on the frontal view. **c** Areas on the frontal view. **d** Angles on the frontal view (left) and the lateral view (right)
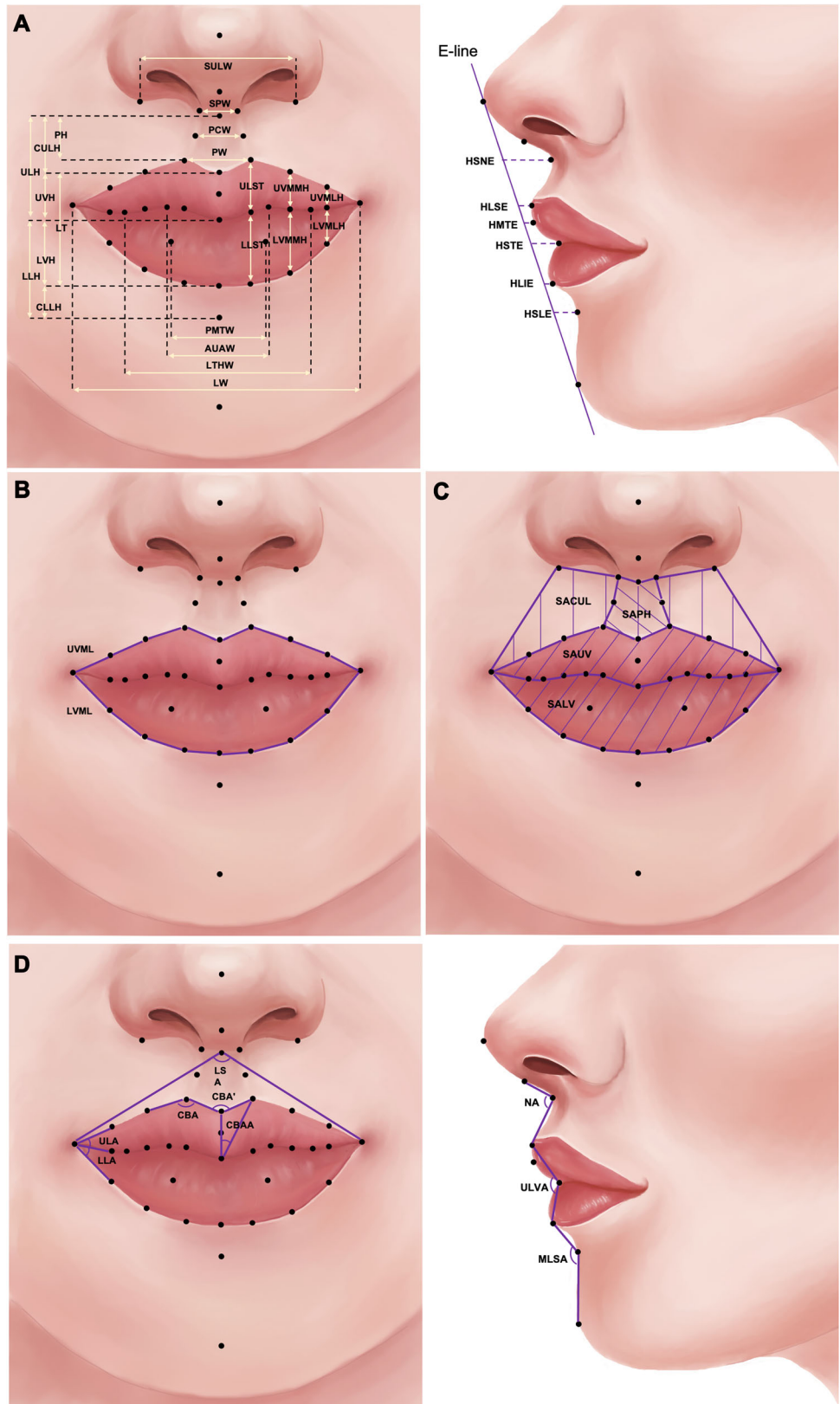
Discussion

Unlike the periocular region, no research has thoroughly evaluated the reliability and accuracy of 3D stereophotogrammetry in perioral anthropometry which restricts its utilization to full potential in clinical practice, especially in reconstructive and aesthetic plastic surgeries of the lower face. To address this, the present study developed a feasible and repeatable protocol of evaluating perioral region with the VECTRA H2 3D imaging system, involving identification of 25 anthropometric perioral landmarks in a sequential order and subsequent generation of corresponding 28 linear distances, 2 curvatures, 4 areas and 9 angles. These measurements covered the entire perioral region between the nose base and mentolabial sulcus and involve both frontal and lateral views. Our protocol showed high intrarater, interrater and intramethod reliability as reflected by several statistics including MAD, REM, TEM, %TEM, total TEM, % total TEM as well as ICC.

The precision error of limits on measurements was usually defined as 1 [6, 18] or 2 unit [15, 19]. In this study, we defined different error limits for different measurements because of the huge difference in the value magnitude. The mean value of linear distances and curvatures in our perioral measurement ranged from 0.35 to 67.87 mm, the areas

ranged from 135.03 to 664.64 mm^2, while the angles ranged from 26.32° to 139.04°. Our findings showed highly reliable results of mean values for MAD, TEM, REM, %TEM and ICC in rater 1 (0.57 unit, 0.51 unit, 2.18%, 2.02% and 0.98) and rater 2 (0.57 unit, 0.55 unit, 2.44%, 2.34% and 0.98). Such consistency also extended to between raters (0.78 unit, 0.74 unit, 3.26%, 3.06% and 0.97) and between methods (1.01 unit, 0.97 unit, 4.74%, 4.57% and 0.95).

We found a highest reliability in intrarater measurements and a lowest but still good reliability in intramethod measurements. Ideal evaluation of intramethod reliability should only involve errors from camera recalibration while maintaining consistency of patients' position and face expression. However, the way we calculated intramethod reliability in this study introduced intrarater errors as well because the rater had to perform two sets of landmark identification on the two sets of 3D images captured at different times. To address this, the study by Gibelli D. et al. pre-marked 50 facial landmarks on participants' face using liquid eyeliner and compared the linear, angular and surface area measurements between two captures of one 3D imaging device and captures of two devices (VECTRA H1 and VECTRA M3). The intramethod reliability was shown to be high in M3–M3, H1–H1 and M3–H1 comparisons (TEM range: 0.3–2.0 mm, 0.4–1.8°; REM range: 0.2–3.1%) [5].

Consistent with others' study [6, 16], we found that measurements with a small magnitude of value have small MAD and TEM estimates but large REM and %TEM estimates, and vise versa. In particular, 4 measurements relating lips' distance to E-line (HLSE, HMTE, HSTE and HLIE) had small MAD and TEM estimates in intrarater, interrater and intramethod reliability assessment, but had high REM and %TEM estimates especially for intramethod reliability (> 10%). This is most likely due to the very small value of these measurements in our study population.
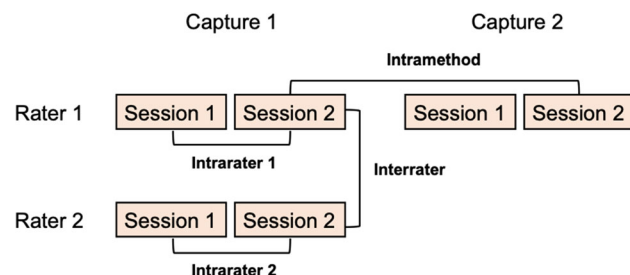
Fig. 3 Schematic of intrarater, interrater and intramethod reliability assessment

Table 3 List of statistics for reliability evaluation

Statistic	Equation
Intraclass correlation coefficient (ICC)	B/(B+W)
Mean absolute difference (MAD)	$\|X1 - X2\|$
Technical error of measurement (TEM)	$\sqrt{(\sum D^2/2N)}$
Relative error measurement (REM)	$(MAD/X3) \times 100$
Relative TEM (% TEM)	$(TEM/X3) \times 100$
Total TEM	$\sqrt{\left(\frac{TEM(intra1)^2 + TEM(intra2)^2}{2}\right) + TEM(inter)^2}$
Relative total TEM (% total TEM)	(% total TEM/X3) × 100

B, between-measurement variance; W, within measurement variance; D, difference between measurements; N, number of lips or subjects measured; X1, session 1 or session 2 of capture 1; X2, session 2 of capture 1 or capture 2; X3, grand mean; TEM (intra1) is the intrarater TEM for the rater 1; TEM (intra2) is the intrarater TEM for the rater 2; TEM (inter) is the interrater TEM between the two raters.

Table 4 Mean and standard deviations (SDs) of all measurements

Variable	Rater 1								Rater 2			
	Capture 1				Capture 2				Capture 1			
	Session 1		Session 2		Session 1		Session 2		Session 1		Session 2	
	μ	SD	μ	SD	μ	SD	μ	SD	μ	SD	μ	SD
SULW	25.61	2.02	25.74	2.20	25.58	2.32	25.64	2.33	25.61	2.28	25.60	2.29
SPW	5.68	0.99	5.67	0.98	5.60	1.02	5.60	0.98	5.88	1.02	5.87	0.93
PCW	7.88	1.06	7.89	0.99	7.98	0.95	7.98	1.19	8.25	1.09	8.30	1.00
PW	10.81	1.53	10.68	1.43	10.77	1.54	10.76	1.57	11.09	1.51	10.97	1.40
PH	12.73	2.30	12.73	2.30	12.95	2.13	12.83	2.11	12.71	2.39	12.72	2.31
ULH	22.69	1.97	22.70	1.95	22.77	2.00	22.74	1.92	22.77	2.02	22.81	1.99
CULH	15.17	2.42	15.11	2.44	15.35	2.28	15.21	2.18	15.28	2.51	15.24	2.43
UVH	7.57	1.52	7.57	1.51	7.45	1.65	7.55	1.68	7.49	1.60	7.56	1.63
LLH	16.66	2.48	16.65	2.31	16.63	2.29	16.52	2.29	16.85	2.51	16.90	2.38
CLLH	7.29	2.22	7.27	2.21	7.27	2.17	7.30	2.18	7.35	2.25	7.41	2.07
LVH	9.53	2.04	9.57	2.01	9.45	1.83	9.48	1.84	9.51	2.05	9.50	2.01
UVMLH	4.31	0.95	4.31	0.99	4.25	0.89	4.32	0.91	4.35	0.96	4.39	0.96
LVMLH	4.68	1.17	4.65	1.16	4.65	1.09	4.61	1.12	4.78	1.08	4.75	1.09
UVMMH	7.32	1.27	7.31	1.34	7.19	1.34	7.24	1.39	7.14	1.29	7.18	1.30
LVMMH	7.84	1.61	7.90	1.63	7.81	1.46	7.81	1.50	7.72	1.59	7.71	1.58
ULST	9.67	1.54	9.68	1.62	9.54	1.66	9.62	1.78	9.74	1.53	9.79	1.51
LLST	9.66	2.03	9.64	1.99	9.54	1.83	9.55	1.79	9.46	1.96	9.45	1.95
AUAW	13.93	2.64	14.10	2.70	13.91	2.67	13.98	2.70	14.22	2.89	14.13	2.89
LTHW	31.66	3.58	31.48	3.59	31.60	3.79	31.76	3.74	30.97	3.56	31.05	3.48
PMTW	15.81	2.38	15.81	2.34	15.65	2.37	15.67	2.41	16.24	2.47	16.29	2.34
LW	49.03	3.35	49.09	3.32	48.95	3.62	49.21	3.47	48.68	3.33	48.42	3.33
LT	17.06	3.31	17.23	3.39	16.92	3.24	17.06	3.27	17.00	3.38	17.07	3.40
HSNE	− 9.69	1.33	− 9.68	1.34	− 9.70	1.40	− 9.71	1.41	− 9.60	1.33	− 9.61	1.34
HLSE	− 1.40	2.02	− 1.40	2.02	− 1.47	2.05	− 1.45	2.03	− 1.34	2.04	− 1.36	2.02
HMTE	− 0.93	2.12	− 0.94	2.14	− 0.97	2.19	− 0.98	2.18	− 0.92	2.14	− 0.96	2.11
HSTE	− 4.31	2.05	− 4.30	2.06	− 4.46	2.06	− 4.47	2.06	− 4.26	2.07	− 4.25	2.06
HLIE	− 0.38	2.58	− 0.39	2.59	− 0.61	2.54	− 0.59	2.53	− 0.35	2.61	− 0.36	2.58
HSLE	− 4.44	1.80	− 4.44	1.79	− 4.64	1.72	− 4.64	1.72	− 4.41	1.79	− 4.42	1.79
UVML	67.68	4.95	67.68	4.91	67.53	4.95	67.73	4.92	67.87	4.96	67.80	4.93
LVML	60.07	5.11	60.15	5.05	59.91	4.97	60.00	5.03	60.11	4.97	60.12	5.05
SAPH	135.10	27.86	135.03	27.89	135.55	27.51	135.33	27.44	135.72	27.72	135.27	27.33
SAUV	450.54	80.77	450.41	80.73	450.54	80.22	450.96	80.82	449.34	80.46	449.23	80.72
SALV	423.65	97.89	424.06	97.88	423.13	96.99	423.30	97.52	423.52	97.82	423.19	98.05
SACUL	663.63	96.76	663.62	97.13	664.64	97.10	664.55	97.19	664.18	96.21	663.92	96.14
LSA	84.15	5.73	84.33	5.60	83.93	5.13	84.26	5.10	83.71	5.65	83.46	5.69
CBA	134.42	6.85	134.39	6.93	134.34	6.98	134.11	6.96	133.26	6.86	133.28	6.61
CBA′	129.25	8.79	129.28	8.57	128.88	9.06	128.98	8.79	128.61	8.56	129.27	8.88
CBAA	26.34	4.61	26.32	4.65	26.52	4.83	26.38	4.77	26.71	4.63	26.54	4.59
ULA	35.63	6.00	35.96	5.85	35.71	6.09	35.88	6.32	35.11	5.87	35.00	5.83
LLA	35.36	8.73	35.41	8.75	35.27	8.58	35.40	8.72	35.19	8.78	34.66	9.20
NA	101.98	9.68	102.00	9.72	102.02	10.05	101.98	10.37	101.96	9.59	101.77	9.72
ULVA	138.62	10.11	138.62	10.25	138.75	10.23	138.89	10.33	138.47	9.54	138.30	9.54
MLSA	138.79	9.31	139.04	9.10	138.03	9.16	137.94	9.21	139.16	9.01	138.89	8.71

Table 5 Intrarater, interrater and intramethod intraclass correlation coefficient (ICC) and mean differences of all measurements

Variable	Intrarater 1			Intrarater 2			Interrater			Intramethod		
	ICC (CI 95%)	Dmean	P value	ICC (CI 95%)	Dmean	P value	ICC (CI 95%)	Dmean	P value	ICC (CI 95%)	Dmean	P value
SULW	0.93 (0.89–0.96)	−0.13	0.230	0.98 (0.97–0.99)	0.01	0.859	0.91 (0.84–0.95)	0.14	0.312	0.93 (0.88–0.96)	0.10	0.393
SPW	0.92 (0.86–0.95)	0.01	0.927*	0.96 (0.93–0.98)	0.02	0.703	0.93 (0.80–0.97)	−0.20	<0.001	0.90 (0.84–0.94)	0.06	0.305
PCW	0.92 (0.87–0.96)	−0.01	0.887	0.92 (0.87–0.96)	−0.05	0.408	0.83 (0.72–0.9)	−0.41	<0.001	0.88 (0.8–0.93)	−0.09	0.725
PW	0.93 (0.88–0.96)	0.13	0.111	0.96 (0.93–0.98)	0.12	0.047	0.91 (0.84–0.95)	−0.29	0.002	0.89 (0.82–0.94)	−0.07	0.461
PH	0.98 (0.97–0.99)	0.00	0.964	0.99 (0.98–0.99)	−0.02	0.763	0.98 (0.97–0.99)	0.00	0.952	0.95 (0.91–0.97)	−0.10	0.307
ULH	0.98 (0.97–0.99)	−0.01	0.668	0.99 (0.97–0.99)	−0.04	0.592	0.96 (0.93–0.98)	−0.11	0.981	0.95 (0.91–0.97)	−0.04	0.686
CULH	0.98 (0.97–0.99)	0.06	0.375	0.98 (0.97–0.99)	0.04	0.543	0.98 (0.96–0.99)	−0.13	0.104	0.93 (0.88–0.96)	−0.10	0.430
UVH	0.97 (0.96–0.99)	0.00	0.966	0.97 (0.95–0.98)	−0.07	0.195	0.94 (0.90–0.97)	0.01	0.927	0.91 (0.85–0.95)	0.02	0.811
LLH	0.97 (0.95–0.98)	0.02	0.845	0.96 (0.93–0.98)	−0.06	0.552	0.94 (0.90–0.97)	−0.26	0.027	0.90 (0.83–0.94)	0.13	0.389
CLLH	0.98 (0.97–0.99)	0.02	0.709	0.97 (0.95–0.98)	−0.06	0.347	0.98 (0.96–9.99)	−0.14	0.043	0.96 (0.93–0.98)	−0.03	0.736
LVH	0.99 (0.98–0.99)	−0.04	0.368	0.99 (0.98–0.99)	0.01	0.855	0.98 (0.97–0.99)	0.07	0.172	0.95 (0.92–0.97)	0.10	0.203
UVMLH	0.97 (0.95–0.98)	0.00	0.913	0.95 (0.92–0.97)	−0.05	0.272	0.93 (0.88–0.96)	−0.08	0.133	0.94 (0.89–0.96)	−0.01	0.849
LVMLH	0.98 (0.96–0.99)	0.03	0.387	0.94 (0.90–0.97)	0.04	0.482	0.95 (0.91–0.97)	−0.09	0.067	0.94 (0.89–0.96)	0.04	0.445
UVMMH	0.98 (0.97–0.99)	0.01	0.878	0.98 (0.96–0.99)	−0.03	0.415	0.93 (0.88–0.96)	0.14	0.055	0.94 (0.90–0.97)	0.07	0.296
LVMMH	0.98 (0.97–0.99)	−0.06	0.173	0.98 (0.96–0.99)	0.01	0.875	0.97 (0.92–0.98)	0.19	0.001	0.95 (0.91–0.97)	0.09	0.203
ULST	0.99 (0.97–0.99)	−0.01	0.858	0.99 (0.98–0.99)	−0.05	0.182	0.94 (0.90–0.97)	−0.11	0.152	0.94 (0.90–0.97)	0.05	0.515
LLST	0.99 (0.99–1.00)	0.02	0.578	0.99 (0.98–0.99)	0.01	0.847	0.99 (0.98–0.99)	0.20	<0.001	0.96 (0.93–0.98)	0.09	0.249
AUAW	0.98 (0.96–0.99)	−0.17	0.040	0.98 (0.96–0.99)	0.09	0.336	0.97 (0.95–0.99)	−0.04	0.672	0.98 (0.96–0.99)	0.12	0.168
LTHW	0.97 (0.95–0.98)	0.18	0.163	0.98 (0.96–0.99)	−0.08	0.439	0.97 (0.95–0.98)	0.42	0.001	0.96 (0.94–0.98)	−0.28	0.054
PMTW	0.95 (0.91–0.97)	0.00	0.896*	0.96 (0.94–0.98)	−0.05	0.377*	0.95 (0.91–0.97)	−0.49	<0.001*	0.97 (0.94–0.98)	0.14	0.109
LW	0.98 (0.97–0.99)	−0.06	0.466*	0.97 (0.95–0.98)	0.26	0.017*	0.94 (0.89–0.96)	0.67	<0.001*	0.93 (0.88–0.96)	−0.12	0.516
LT	0.99 (0.98–0.99)	−0.17	0.032	0.99 (0.98–1.00)	−0.06	0.332	0.98 (0.96–0.99)	0.17	0.131	0.96 (0.93–0.98)	0.17	0.238
HSNE	1.00 (1.00–1.00)	−0.01	0.365	1.00 (0.99–1.00)	0.01	0.643	0.99 (0.99–1.00)	−0.07	0.011	0.97 (0.94–0.98)	0.03	0.515
HLSE	1.00 (1.00–1.00)	0.00	0.909*	1.00 (1.00–1.00)	0.01	0.377*	1.00 (1.00–1.00)	−0.04	0.064*	0.98 (0.97–0.99)	0.05	0.783*
HMTE	1.00 (1.00–1.00)	0.01	0.372*	1.00 (1.00–1.00)	0.04	0.454*	1.00 (1.00–1.00)	0.02	0.554*	0.95 (0.91–0.97)	0.04	0.927*
HSTE	1.00 (1.00–1.00)	−0.01	0.061*	1.00 (1.00–1.00)	−0.01	0.242*	1.00 (1.00–1.00)	−0.05	0.006*	0.95 (0.91–0.97)	0.17	0.160*
HLIE	1.00 (1.00–1.00)	0.01	0.532*	1.00 (1.00–1.00)	0.01	0.247*	1.00 (1.00–1.00)	−0.03	0.080*	0.97 (0.94–0.98)	0.21	0.146*
HSLE	1.00 (1.00–1.00)	0.00	0.851*	1.00 (1.00–1.00)	0.01	0.489*	1.00 (1.00–1.00)	−0.03	0.454*	0.94 (0.89–0.97)	0.20	0.011*
UVML	1.00 (1.00–1.00)	0.00	0.978	0.99 (0.99–1.00)	0.02	0.415	0.99 (0.99–1.00)	−0.13	0.167	0.98 (0.96–0.99)	−0.06	0.703
LVML	1.00 (1.00–1.00)	−0.08	0.214	0.99 (0.99–1.00)	−0.01	0.950	0.99 (0.99–1.00)	0.04	0.700	0.96 (0.93–0.98)	0.15	0.457
SAPH	1.00 (1.00–1.00)	0.07	0.735	1.00 (1.00–1.00)	0.45	0.079	1.00 (1.00–1.00)	−0.24	0.406	1.00 (1.00–1.00)	−0.31	0.227
SAUV	1.00 (1.00–1.00)	0.13	0.696	1.00 (1.00–1.00)	0.11	0.780	1.00 (1.00–1.00)	1.18	0.043	1.00 (1.00–1.00)	−0.55	0.464
SALV	1.00 (1.00–1.00)	−0.41	0.224	1.00 (1.00–1.00)	0.34	0.312	1.00 (1.00–1.00)	0.87	0.036	1.00 (1.00–1.00)	0.76	0.154
SACUL	1.00 (1.00–1.00)	0.01	0.983	1.00 (1.00–1.00)	0.27	0.501	1.00 (1.00–1.00)	−0.30	0.581	1.00 (1.00–1.00)	−0.93	0.137
LSA	0.99 (0.98–0.99)	−0.18	0.152	0.99 (0.98–0.99)	0.26	0.034	0.98 (0.96–0.99)	0.88	<0.001	0.93 (0.89–0.96)	0.07	0.798

Table 5 continued

Variable	Intrarater 1 ICC (CI 95%)	Dmean	P value	Intrarater 2 ICC (CI 95%)	Dmean	P value	Interrater ICC (CI 95%)	Dmean	P value	Intramethod ICC (CI 95%)	Dmean	P value
CBA	0.98 (0.97–0.99)	0.03	0.889*	0.99 (0.98–0.99)	−0.02	0.900	0.97 (0.95–0.98)	1.11	<0.001	0.95 (0.92–0.97)	0.28	0.700*
CBA'	0.98 (0.97–0.99)	−0.03	0.899	0.97 (0.95–0.98)	−0.66	0.041	0.96 (0.93–0.98)	0.01	0.984	0.92 (0.87–0.96)	0.30	0.626*
CBAA	0.98 (0.96–0.99)	0.02	0.779*	0.98 (0.97–0.99)	0.18	0.128*	0.98 (0.97–0.99)	−0.22	0.052*	0.97 (0.95–0.98)	−0.06	0.954
ULA	0.99 (0.98–0.99)	−0.34	0.014	0.97 (0.95–0.98)	0.11	0.570	0.95 (0.91–0.97)	0.97	0.001	0.90 (0.84–0.95)	0.08	0.825
LLA	0.99 (0.99–1.00)	−0.05	0.728	0.96 (0.94–0.98)	0.53	0.139	0.96 (0.93–0.98)	0.75	0.049	0.96 (0.93–0.98)	0.01	0.966
NA	0.99 (0.99–1.00)	−0.02	0.891	0.99 (0.99–1.00)	0.19	0.144	0.99 (0.99–1.00)	0.23	0.171	0.98 (0.96–0.99)	0.03	0.933
ULVA	0.99 (0.99–1.00)	−0.01	0.970	0.99 (0.99–1.00)	0.17	0.273	0.98 (0.97–0.99)	0.32	0.211	0.96 (0.93–0.98)	−0.27	0.515
MLSA	0.99 (0.99–1.00)	−0.26	0.105	0.99 (0.98–0.99)	0.27	0.142	0.98 (0.97–0.99)	0.15	0.515	0.89 (0.82–0.94)	1.10	0.076

Compared to Caucasians, Asians tend to have a less prominent nose and chin [20], and thus the upper and lower lips are in close proximity to the E-line. Therefore, we believe that these measurements should have higher reliability in Caucasian population with a concave facial profile which needs further validation. On the contrary, because of the big value magnitude, the four areal measurements showed relatively large MAD and TEM estimates (1.02–3.71 mm^2) but small REM and %TEM estimates (0.34–1.28%).

Overall, we demonstrated a standardized method of perioral assessment that shows high precision. This study lays the foundation of applying this new imaging technology to many clinical situations. For example, certain diseases concerning perioral morphology change can be easily diagnosed with the establishment of normative and pathologic 3D perioral anthropometric database. Attractive lips can be better defined with detailed measurements. And with such knowledge, planning of lip surgeries or aesthetic procedures (e.g. fillers and/or botulinum injection) can be more individualized and precise, the therapeutic or aesthetic effects can be objectively assessed and better expected. Accordingly, a clearer and more reliable clinician-patient relationship might be built.

This study has several limits. First, it is hard to keep an exactly same lips position for participants at different captures and this had resulted in a relatively big intramethod error compared to the intrarater and interrater errors. The human mouth is a very muscular region with many muscles stemming from or terminate at orbicularis oris, and the lip morphology is further influenced by the occlusion status. Hence, although each participant has been required to gently close their mouth and teeth with a relaxed neutral face expression, involuntary lips movement was unable to be completely eliminated. Indeed, the study of Othman et al. showed that bigger evaluation errors were recorded in the mouth area compared to other parts of the face of patients with cleft lip and palate [14].

Second, no external validation was conducted in this study. Comparison of 3D stereophotogrammetry to direct anthropometry or digital photogrammetry may further confirm its validity and accuracy. Third, the sample size of this study is relatively small and the majority of them were female. This method needs to be evaluated in different ethnics as well. For example, aged Caucasians (especially Fitzpatrick skin types I, II and III) tend to have effaced vermilion border due to actinic damage which might impairs the reliability of assessment. Besides, a chief complaint in predominately Caucasians is perioral lines. As our subjects were all Asians and relatively young, perioral rhytides were not measured in this study. Future studies should encompass perioral pathologies (such as cleft lip and palate, acromegalia and congenital

macrostomia), more males, more elderly people as well as other ethnic groups, and perioral rhytides assessment should be taken into account during the evaluation.

Conclusions

In this study, we proposed a novel and thorough evaluation protocol for perioral anthropometry using the 3D digital stereophotogrammetry technology with linear, curvilinear, angular and areal measurements. And we proved it highly reliable and repeatable for the analysis of perioral morphological characteristics. These results imply a great potential for its application in instructing clinical practice, although further validation is required.

Declarations

Informed Consent All participants provided written consent for the use of their 3D facial images.

Ethical Approval This study was approved by the Institutional Review Board of the Peking Union Medical College Hospital (Reference Number I-22PJ693) and conducted in accordance with the Declaration of Helsinki.

References

1. Wu SQ, Pan BL, An Y et al (2019) Lip morphology and aesthetics: study review and prospects in plastic surgery. Aesthetic Plast Surg 43(3):637–643. https://doi.org/10.1007/s00266-018-1268-x
2. Naini FB, Gill DS (2008) Facial aesthetics: 1. Concepts and canons. Dent Update 35(2):102–104, 106–107. https://doi.org/10.12968/denu.2008.35.2.102
3. Werschler WP, Fagien S, Thomas J et al (2015) Development and validation of a photographic scale for assessment of lip fullness. Aesthet Surg J 35(3):294–307. https://doi.org/10.1093/asj/sju025
4. Kim H, Lee M, Park SY et al (2019) Age-related changes in lip morphological and physiological characteristics in Korean women. Skin Res Technol 25(3):277–282. https://doi.org/10.1111/srt.12644
5. Gibelli D, Pucciarelli V, Cappella A et al (2018) Are portable stereophotogrammetric devices reliable in facial imaging? A validation study of VECTRA H1 device. J Oral Maxillofac Surg 76(8):1772–1784. https://doi.org/10.1016/j.joms.2018.01.021
6. Guo Y, Rokohl AC, Schaub F et al (2019) Reliability of periocular anthropometry using three-dimensional digital stereophotogrammetry. Graefes Arch Clin Exp Ophthalmol 257(11):2517–2531. https://doi.org/10.1007/s00417-019-04428-6

7. Patel AA, Schreiber JE, Gordon AR et al (2022) Three-dimensional perioral assessment following subnasal lip lift. Aesthet Surg J 42(7):733–739. https://doi.org/10.1093/asj/sjac070
8. Gibelli D, Codari M, Rosati R et al (2015) A quantitative analysis of lip aesthetics: the influence of gender and aging. Aesthetic Plast Surg 39(5):771–776. https://doi.org/10.1007/s00266-015-0495-7
9. Tsai LC, Lin ET, Chang CC et al (2022) Quantitative and objective measurements of facial aging process with anatomical landmarks. J Cosmet Dermatol 21(3):1317–1320. https://doi.org/10.1111/jocd.14221
10. Weinberg SM (2022) 3D Facial Norms Database. https://www.facebase.org/facial_norms/summary/
11. Sawyer AR, See M, Nduka C (2009) 3D stereophotogrammetry quantitative lip analysis. Aesthetic Plast Surg 33(4):497–504. https://doi.org/10.1007/s00266-008-9191-1
12. Sforza C, Grandi G, Binelli M et al (2010) Age- and sex-related changes in three-dimensional lip morphology. Forensic Sci Int 200(1–3):182.e1–7. https://doi.org/10.1016/j.forsciint.2010.04.050
13. Chong Y, Dong R, Liu X et al (2021) Stereophotogrammetry to reveal age-related changes of labial morphology among Chinese women aging from 20 to 60. Skin Res Technol 27(1):41–48. https://doi.org/10.1111/srt.12906
14. Othman SA, Saffai L, Wan Hassan WN (2020) Validity and reproducibility of the 3D VECTRA photogrammetric surface imaging system for the maxillofacial anthropometric measurement on cleft patients. Clin Oral Investig 24(8):2853–2866. https://doi.org/10.1007/s00784-019-03150-1
15. Weinberg SM, Scott NM, Neiswanger K et al (2004) Digital three-dimensional photogrammetry: evaluation of anthropometric precision and accuracy using a Genex 3D camera system. Cleft Palate Craniofac J 41(5):507–518. https://doi.org/10.1597/03-066.1
16. Andrade LM, Rodrigues da Silva AMB, Magri LV et al (2017) Repeatability study of angular and linear measurements on facial morphology analysis by means of stereophotogrammetry. J Craniofac Surg 28(4):1107–1111. https://doi.org/10.1097/scs.0000000000003554
17. Camison L, Bykowski M, Lee WW et al (2018) Validation of the vectra H1 portable three-dimensional photogrammetry system for facial imaging. Int J Oral Maxillofac Surg 47(3):403–410. https://doi.org/10.1016/j.ijom.2017.08.008
18. de Menezes M, Rosati R, Ferrario VF et al (2010) Accuracy and reproducibility of a 3-dimensional stereophotogrammetric imaging system. J Oral Maxillofac Surg 68(9):2129–2135. https://doi.org/10.1016/j.joms.2009.09.036
19. De Stefani A, Barone M, Hatami Alamdari S et al (2022) Validation of vectra 3D imaging systems: a review. Int J Environ Res Public Health. https://doi.org/10.3390/ijerph19148820
20. Zhao XG, Hans MG, Palomo JM et al (2013) Comparison of Chinese and white Bolton standards at age 13. Angle Orthod 83(5):809–816. https://doi.org/10.2319/110412-849.1