



# A novel standardized approach for the 3D evaluation of upper eyelid area and volume

Yongwei Guo<sup>1,2,3#^</sup>, Alexander C. Rokohl<sup>3,4#^</sup>, Wanlin Fan<sup>3^</sup>, Rodothea Theodosiou<sup>3^</sup>, Xueting Li<sup>3^</sup>, Lixia Lou<sup>1,2^</sup>, Tao Gao<sup>1,2^</sup>, Ming Lin<sup>5^</sup>, Ke Yao<sup>1,2\*^</sup>, Ludwig M. Heindl<sup>3,4\*^</sup>

<sup>1</sup>Eye Center, Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China; <sup>2</sup>Zhejiang Provincial Key Lab of Ophthalmology, Hangzhou, China; <sup>3</sup>Department of Ophthalmology, University of Cologne, Faculty of Medicine and University Hospital of Cologne, Cologne, Germany; <sup>4</sup>Center for Integrated Oncology (CIO) Aachen-Bonn-Cologne-Duesseldorf, Cologne, Germany; <sup>5</sup>Department of Ophthalmology, Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

*Contributions:* (I) Conception and design: Y Guo, K Yao, LM Heindl; (II) Administrative support: LM Heindl; (III) Provision of study materials or patients: AC Rokohl, LM Heindl; (IV) Collection and assembly of data: Y Guo, W Fan, R Theodosiou; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

#These authors contributed equally to this work and should be considered first authors.

\*These authors contributed equally to this work and should be considered senior authors.

*Correspondence to:* Ke Yao, MD, PhD. Eye Center, Second Affiliated Hospital, Zhejiang University School of Medicine, 88 Jiefang Road, Hangzhou, 310009, China. Email: xlren@zju.edu.cn; Ludwig M. Heindl, MD, PhD. Department of Ophthalmology, University of Cologne, Faculty of Medicine and University Hospital Cologne, Kerpener Strasse 62, 50937, Cologne, Germany. Email: ludwig.heindl@uk-koeln.de; Yongwei Guo, MD. Eye Center, Second Affiliated Hospital, Zhejiang University School of Medicine, 88 Jiefang Road, Hangzhou 310009, China. Email: yongwei-guo@zju.edu.cn.

**Background:** Three-dimensional (3D) surface imaging and morphometrics are becoming the preferred craniofacial surface imaging modality. However, as a unique advantage of 3D imaging, areal and volumetric measurements have been rarely conducted and validated for evaluating soft tissue change in the periocular region, especially the upper eyelids. Therefore, based on an existing periocular landmark identification strategy, we proposed a novel modified method to define a standardized upper eyelid region for areal and volumetric measurements and validate its reliability for future clinical application.

**Methods:** Forty-four healthy adult volunteers were enrolled in this cross-sectional study. Three-dimensional facial images were taken with a 3D imaging system. Subsequently, the upper eyelid region selection and areal and volumetric measurements were conducted using a modified landmarks localization strategy to evaluate their intrarater, interrater, and intramethod reliability.

**Results:** Areal measurement of the upper eyelid revealed highly reliable outcomes for intraclass correlation coefficient (ICC), mean absolute difference (MAD), relative error measurement (REM), technical error of measurement (TEM), and relative technical error of measurement (%TEM) of 0.982, 0.1620 cm<sup>2</sup>, 2.9%, 0.1510 cm<sup>2</sup>, and 2.7% for intrarater reliability, respectively; 0.969, 0.2076 cm<sup>2</sup>, 3.7%, 0.1930 cm<sup>2</sup>, and 3.5% for interrater reliability, respectively; and 0.917, 0.3636 cm<sup>2</sup>, 6.5%, 0.3354 cm<sup>2</sup>, and 6.0% for intramethod reliability, respectively. Unsatisfactory results were found for the volumetric measurement of the upper eyelid: the ICC, MAD, REM, TEM, and %TEM estimates for intrarater reliability were 0.992, 0.2299 mL, 10.3%, 0.2414 mL, and 10.8%, respectively; for interrater reliability, these values were 0.985, 0.2749 mL, 12.3%,

<sup>^</sup> ORCID: Yongwei Guo, 0000-0001-9195-0770; Alexander C. Rokohl, 0000-0002-0224-3597; Wanlin Fan, 0000-0001-7143-6707; Rodothea Theodosiou, 0000-0002-6920-6854; Xueting Li, 0000-0002-4146-8991; Lixia Lou, 0000-0002-3035-8009; Tao Gao, 0000-0001-5684-0015; Ming Lin, 0000-0001-6467-4643; Ke Yao, 0000-0002-6764-7365; Ludwig M. Heindl, 0000-0002-4413-6132.

0.3253 mL, and 14.6%, respectively; and for intramethod reliability, these values were 0.433, 1.6716 mL, 77.9%, 2.0615 mL, and 96.1%, respectively.

**Conclusions:** This is the first study to propose a standardized upper eyelid region selection strategy and simultaneously validate its reliability for 3D areal and volumetric measurements. This study confirmed the high-level reliability of areal measurement and poor reliability of volumetric measurement based on direct measurements using a single image, which may provide better results when this method is combined with the image overlapping and registration procedure. However, this is subject to further validation. Nonetheless, this method could provide quantitative areal and volumetric data on the upper eyelids and might have widespread application potential in the future.

**Keywords:** Area; volume; eyelid; three-dimensional (3D); photogrammetry

Submitted Jun 11, 2022. Accepted for publication Dec 12, 2022. Published online Feb 10, 2023.

doi: 10.21037/qims-22-589

View this article at: <https://dx.doi.org/10.21037/qims-22-589>

## Introduction

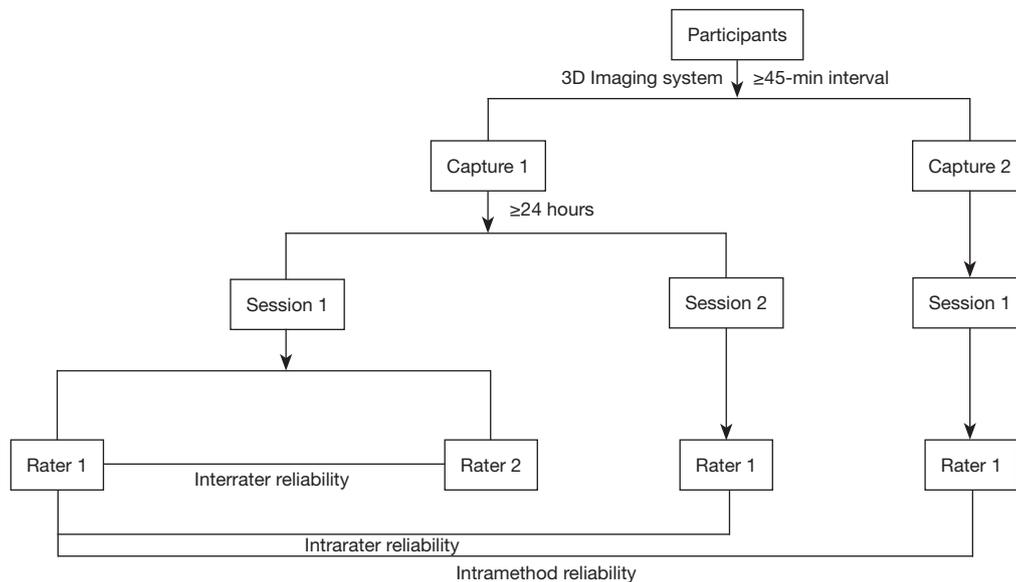
The eyes and the periocular region, especially the eyelids, are the most prominent anatomical structures in the center of the face, attracting increasing interest in surgical or nonsurgical esthetic procedures to improve an undesired appearance or correct deformities (1). As introduced by Farkas (2), standardized facial anthropometric analysis has become widespread for disease diagnosis, preoperative evaluation, surgical planning, and postoperative follow-up (3-5). Moreover, since aesthetic assessment and disease diagnosis are comprehensive and holistic assessments to a certain extent, it is necessary to evaluate the 3-dimensional (3D) holistic contours and topological surfaces (i.e., area and volume).

In recent decades, 3D surface imaging and morphometrics, especially stereophotogrammetry, have rapidly become the preferred craniofacial surface imaging modality owing to their unprecedented precision and reliability. Three-dimensional lines, curve lines, and angles can be measured on the craniofacial surface, including the periocular region (3-12). Moreover, as a unique advantage of the 3D surface imaging system, the measurement of periocular area and volume is becoming increasingly popular in clinical practice. The 3D volume changes in the lower eyelid region are evaluated after blepharoplasty with or without fat pad transposition and fat grafting (13-15). As for the upper eyelid region, 3D imaging has also shown potential applications. For instance, the 3D upper eyelid contour has been used to design and print customized conformers adapted to patients' postenucleation sockets, which might prevent conformer extrusion and forniceal

shortening (16). Moreover, the postoperative lateral brow-eyelid complex volume has been measured to assess its postoperative volume (i.e., fullness) changes following upper-eyelid blepharoplasty with internal browpexy sutures (17).

However, the accuracy and reliability of any novel measurement technique should be thoroughly validated prior to widespread clinical application. Furthermore, the standardized localization of the periocular surface area will facilitate accurate and reliable measurement of this area. Hence, more studies validating areal and volumetric analysis should be conducted to evaluate soft tissue change in the periocular region, which may possess clinically small yet significant changes. In previous studies (18,19), we printed varying square or rectangular pieces of paper with specific areas and attached them to 7 locations in the upper and lower eyelids. The accuracy and reliability of areal measurements in 3D stereophotogrammetry have been validated on these specific standardized objects; however, the actual eyelid skin area has not been directly measured. Hyer *et al.* (20) evaluated the accuracy and interobserver variability of 3D stereophotogrammetric imaging for volumetric assessment in the periocular region. However, they used the superimposition method to quantify volume changes rather than directly measuring the actual volume of the periocular region. Furthermore, more errors might be introduced while identifying the supraorbital rim as an anatomical feature for demarcating the aiming boundary of the upper eyelids because it is a bony landmark lying under the superficial soft tissues.

This study aimed to determine the feasibility and reliability of a standardized upper eyelid region localization strategy for potential periocular fullness evaluation such as the scar, tumor, eyelid retraction, or upper eyelid tone and



**Figure 1** The process of photo capture and analysis. Rater 1: YG; Rater 2: ACR.

volume loss with age as well as edema due to inflammation or operations, among others (21-24). Therefore, based on an existing standardized periocular landmark identification strategy (9-11), we described a novel modified approach for standardized evaluation of the upper eyelid area and volume and validated its intrarater, interrater, and intramethod reliability for future clinical application.

## Methods

### Participants

A total of 44 White volunteers without any history of eyelid diseases, trauma, or surgeries were recruited from the Department of Ophthalmology, University of Cologne, Cologne, Germany. There were 88 eyes in 15 (34.1%) men and 29 (65.9%) women aged 21.5–67.9 years ( $31 \pm 12.7$  years). Individuals with a medical history of facial pathologies, deformations, impairment, or surgeries influencing facial morphology were excluded. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the institutional ethics board of the University of Cologne and informed consent was taken from all individual participants.

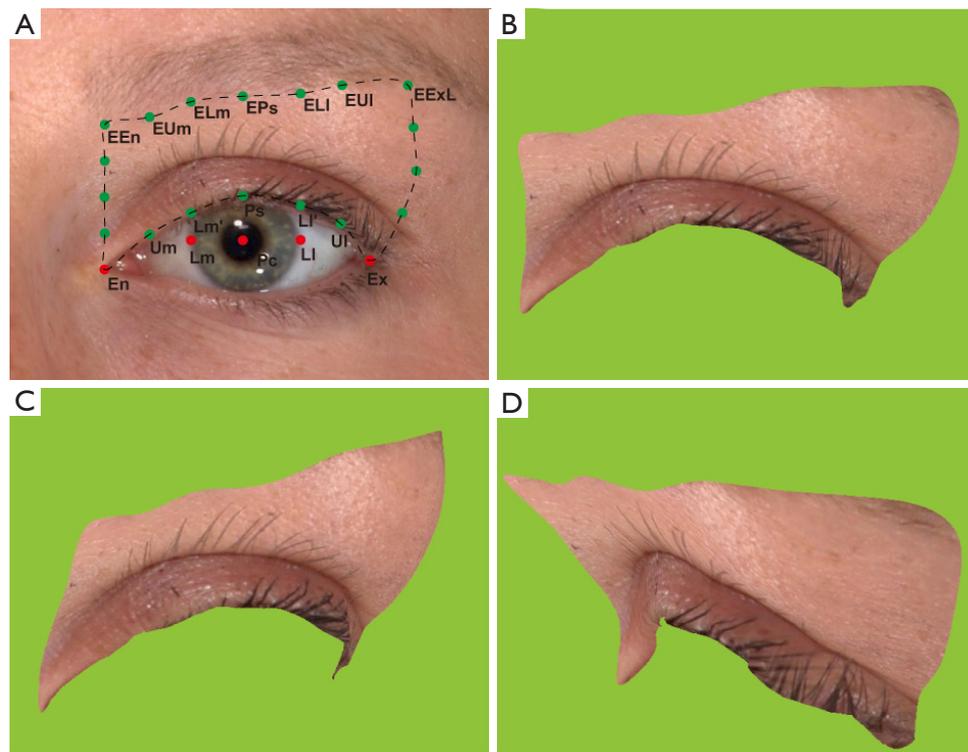
### Three-dimensional facial image acquisition

As shown in *Figure 1*, facial 3D photos were captured

by YG (rater 1) and analyzed by YG and ACR (rater 2) using the Vectra M3 3D Imaging System and Vectra software (Canfield Scientific, Inc., Parsippany, NJ, USA) at the University Hospital of Cologne (10,11). Briefly, 2 captures (captures 1 and 2) were conducted by YG for each participant at an interval of  $\geq 45$  minutes to eliminate possible errors during different captures or calibrations. Recalibration was performed before capture 2. For photos taken during capture 1, YG (rater 1) and ACR (rater 2) each performed the landmark localization and measurement procedure (session 1), and YG conducted the procedure again with an interval of  $\geq 24$  hours (session 2). We evaluated the measurements from session 1 of YG and ACR for interrater reliability. For intrarater reliability, we evaluated the measurements from sessions 1 and 2 of YG; for photos acquired during capture 2, the landmark localization and measurement procedures were performed once by YG (session 1). For intramethod reliability, we assessed the measurements from session 1 of YG during captures 1 and 2.

### Modified approach for areal and volumetric measurement

Our previously published approach identified a series of standardized landmarks (9,10) (Table S1). Briefly, 5 prime landmarks were localized on each 3D photo: the endocanthion (En), exocanthion (Ex), pupillary center (Pc), and the medial and lateral corneoscleral limbi (i.e., Lm and Ll horizontal to the Pc). Furthermore, 5 landmarks were



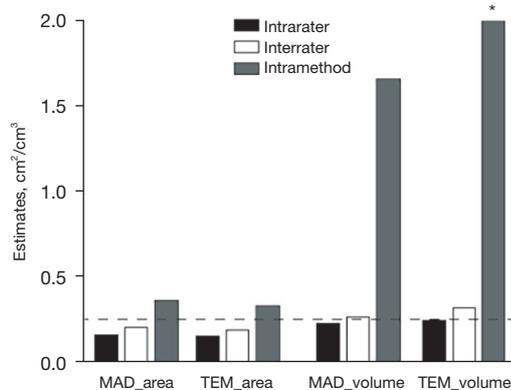
**Figure 2** The novel standardized approach for 3-dimensional evaluation of the upper-eyelid area and volume: standardized landmark identification and the upper-eyelid region. (A) The 5 prime landmarks, including the En, Ex, Pc, and the medial and lateral corneoscleral limbi (i.e., Lm and Ll, horizontal to Pc). Lash roots landmarks on the upper palpebral margin: Ps aligning to Pc, Lm' to Lm, Ll' to Ll, Um in the middle of En and Lm', and Ul in the middle of Ex and Ll'. Inferior margin landmarks of the eyebrows: EEn aligning to En, EPs to Pc, ELm to Lm', ELl to Ll', EUm to Um, EUl to Ul, and EExL to Ex in the lateral review. Medial margin of the upper eyelid: 3 novel landmarks between EEn and En equidistant to each other. Lateral margin of the upper eyelids: 3 novel landmarks between EExL and Ex equidistant to each other. These landmarks could all be connected by a click to form a closed upper-eyelid region, and the Vectra software could measure its area and volume. (B) The frontal view of the selected upper-eyelid region. (C) The medial view of the selected upper-eyelid region. (D) The lateral view of the selected upper-eyelid region. Definition of landmarks were listed in Table S1.

set at the eyelash roots of the upper palpebral margin and 7 at the inferior margin of the eyebrows. Finally, 6 novel landmarks were added at the upper eyelid's medial and lateral margins for areal and volumetric measurements. All these landmarks could be connected with a click to form a closed upper eyelid region, and the area and volume could be measured using Vectra software. *Figure 2* depicts these landmarks (*Figure 2A*) and the identified upper eyelid region in the frontal, medial, and lateral views (*Figure 2B-2D*). The detailed definitions of these landmarks are as follows:

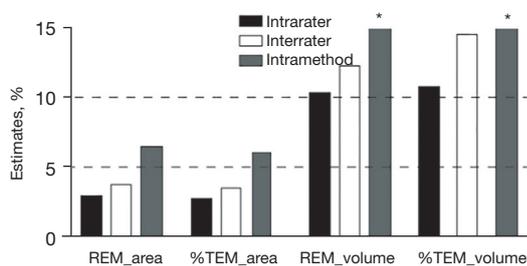
- ❖ Five prime landmarks: the En, Ex, Pc, and the medial and lateral corneoscleral limbi (i.e., Lm and Ll horizontal to the Pc).
- ❖ Lash roots landmarks on the upper palpebral

margin: Ps aligning to the Pc, Lm' to the medial corneoscleral limbus (Lm), Ll' to the lateral corneoscleral limbus (Ll), Um in the middle of the En and Lm', Ul in the middle of the Ex and Ll'.

- ❖ Inferior margin landmarks of the eyebrow: EEn aligning to the En, EPs to the Pc, ELm to Lm, ELl to Ll, EUm to the Um, EUl to the Ul, and EExL to the Ex in the lateral review.
- ❖ Medial margin of the upper eyelid: 3 novel landmarks between the EEn and En at an equal distance from each other.
- ❖ Lateral margin of the upper eyelid: 3 novel landmarks between the EExL and Ex at an equal distance from each other in the lateral review.



**Figure 3** The MAD and TEM of intrarater, interrater, and intramethod reliability of areal and volumetric measurements. \*, data points are outside the axis limits of 2 units. MAD, mean absolute difference; TEM, technical error of measurement.



**Figure 4** The REM and %TEM of intrarater, interrater, and intramethod reliability of areal and volumetric measurements. \*, data points are outside the axis limits of 15%. REM, relative error of measurement; %TEM, relative technical error of measurement.

### Statistical analyses

Reliability, also known as precision or repeatability, demonstrates how repeated measures acquire the same value (9,25). As shown below, 5 statistics were calculated to evaluate the intrarater, interrater, and intramethod reliability (9,25,26). (I) The intraclass correlation coefficient (ICC) ranges from 0 to 1, representing low to high-level reliability. Poor, satisfactory, and excellent consistency were defined as  $<0.4$ ,  $0.4$  to  $0.75$ , and  $\geq 0.75$ , respectively; (II) the mean absolute difference (MAD) is calculated as the average of absolute differences between 2 sets of measurements; (III) the technical error of measurement (TEM) is expressed as the square root of the measurement error variance; (IV–V) the relative error of measurement (REM) and relative TEM (%TEM) are calculated by dividing the MAD or TEM

by the grand mean of both sets of repeated measurements and multiplying the result by 100%. *Excellent consistency* is defined as REM or %TEM  $<1\%$ , very good 1–3.9%, good 4–6.9%, moderate 7–9.9%, and poor  $\geq 10\%$ .

Following the measurements, a database was established using Microsoft Excel for Mac 15.33 (Microsoft Corp., Redmond, WA, USA), in which the statistics mentioned above were calculated (except for the ICC). Bar graphs were created using GraphPad Prism 8 (GraphPad Software Inc., San Diego, CA, USA). The ICC and further statistical analyses were conducted using SPSS 28 (IBM Corp., Armonk, NY, USA). Analysis of variance (ANOVA) of repeated measurement data was applied to analyze the statistical significance of the differences between the 4 measurements.  $P \leq 0.05$  was considered statistically significant, and Bonferroni adjustment was conducted for multiple comparisons.

### Results

To determine the intrarater, interrater, and intramethod reliability of areal and volumetric measurements, the MAD and TEM are shown in *Figure 3*, and the REM and %TEM are demonstrated in *Figure 4*. *Table 1* presents the descriptive statistics (mean and SD) for each measurement, and *Table 2* displays the statistical analyses for all sets of measurements.

#### Reliability of the areal measurement

For intrarater reliability, the mean area was  $5.5730 \text{ cm}^2$  for session 1 and  $5.5656 \text{ cm}^2$  for session 2, as evaluated by YG (rater 1). The differences between both sets of measurements were statistically insignificant. Furthermore, the ICC was 0.982, which indicates excellent intrarater consistency. The MAD and TEM estimates ( $0.1620$  and  $0.1510 \text{ cm}^2$ , respectively) were less than  $1 \text{ cm}^2$ . Additionally, to eliminate the positive influence of measurement size, REM and %TEM were introduced to compare the reliability between various variables, populations, or studies with different measurement sizes combined with the MAD and TEM. Very good reliability was demonstrated for the REM and %TEM (2.9% and 2.7%, respectively).

As for the interrater reliability, the mean area was  $5.6071 \text{ cm}^2$  as evaluated by ACR (rater 2). Statistically nonsignificant differences were observed between areas measured in session 1 by both raters. Moreover, excellent interrater consistency (0.969) was also found in the

**Table 1** Descriptive statistics for the area and volume of the upper eyelid

Variable	Capture 1						Capture 2	
	Rater 1				Rater 2		Rater 1	
	Session 1		Session 2		Session 1		Session 1	
	M (SD)	95% CI						
Area (cm <sup>2</sup> )	5.573 (1.1133)	5.3396, 5.8135	5.5656 (1.1170)	5.3447, 5.7993	5.6071 (1.0799)	5.3822, 5.8562	5.6914 (1.2015)	5.4258, 5.9205
Volume (mL)	2.2405 (2.6709)	1.671, 2.7843	2.2105 (2.6852)	1.6397, 2.7829	2.2191 (2.6562)	1.6644, 2.7968	2.0487 (2.8140)	1.4574, 2.6194

M, mean; SD, standard deviation; CI, confidence interval.

**Table 2** Statistical analyses for all sets of measurements

Variable	Multivariate tests P value	Pairwise comparison <sup>a</sup>			
		Intrarater	Intramethod	Intramethod	Interrater
		R1:S1 vs. S2	R1S1: C1 vs. C2	R1:C1S2 vs. C2S1	S1:R1 vs. R2
Area	0.059	1.000	0.11	0.049	1.000
Volume	0.817	1.000	1.000	1.000	1.000

<sup>a</sup>, adjustment for multiple comparisons: Bonferroni; \*,  $P \leq 0.05$ . R1, rater 1, i.e., YG; R2, rater 2, i.e., ACR; S1, session 1; S2, session 2; C1, capture 1; C2, capture 2; R1:S1 vs. S2, measurement comparison of rater 1's session 1 and session 2 from capture 1; R1S1:C1 vs. C2, measurement comparison of rater 1's session 1 from capture 1 and capture 2; R1:C1S2 vs. C2S1, measurement comparison of rater 1's session 2 from capture 1 and session 1 from capture 2; S1:R1 vs. R2, measurement comparison of session 1 from rater 1 and rater 2 from capture 1.

ICC. Similar to intrarater reliability, less than 1 cm<sup>2</sup> was observed in the MAD and TEM (0.2076 and 0.1930 cm<sup>2</sup>, respectively). Furthermore, very good reliability was also found in the REM and %TEM (3.7% and 3.5%, respectively) despite it being slightly higher than the intrarater values.

Concerning intramethod reliability, a mean area of 5.6914 cm<sup>2</sup> was obtained in capture 2 by YG (rater 1). Furthermore, the differences were statistically significant between the measurement of rater 1's session 1 in capture 2 and the other 3 measures; however, a statistically insignificant difference was found between the 4 measures according to ANOVA ( $P=0.059$ ). Moreover, the ICC estimate (0.917) indicated excellent intramethod consistency. Values of 0.3636 cm<sup>2</sup> and 0.3354 cm<sup>2</sup> were observed in the MAD and TEM, respectively. In addition, REM and %TEM showed good reliability, at 6.5% and 6.0%, respectively, which were higher values than those of the intra- and interrater reliability.

### Reliability of volumetric measurement

For intrarater reliability, the mean volume was 2.2405 mL for session 1 and 2.2105 mL for session 2, and the difference was not statistically significant, as evaluated by YG (rater 1). Furthermore, excellent intrarater consistency (0.992) was indicated in the ICC. The MAD and TEM estimates were 0.2299 and 0.2414 mL, respectively. However, poor agreements were found in the REM and %TEM values (10.3% and 10.8%, respectively).

Regarding interrater reliability, the mean volume was 2.2191 mL, as evaluated by ACR (rater 2), and was not statistically different to rater 1's session 1 in capture 1. In terms of the ICC, excellent interrater consistency (0.985) was observed. Although the MAD and TEM values were less than 1 mL (0.2749 and 0.3253 mL, respectively), poor agreements were found in the REM and %TEM values, with 12.3% and 14.6%, respectively.

Concerning intramethod reliability, a mean volume of

2.0487 mL was acquired in session 1 of capture 2 by rater 1. The differences between session 1 of the 2 captures were also not significant. Conversely, the ICC estimate demonstrated poor consistency (0.433). Additionally, the MAD and TEM showed higher errors than those of the intra- and interrater reliability, with 1.6716 and 2.0615 mL, respectively. Additionally, the REM and %TEM exhibited poor reliability (77.9% and 96.1%, respectively), which was significantly higher than those of the intra- and interrater reliability. Converse plus-minus signs in volume were found in 18 of the 88 eyes between volume measurements in session 1 of capture 2 and the other sessions.

## Discussion

Compared with traditional linear or angular measurements, areal or volumetric measurements have rarely been studied in the periocular region, even though area and volume changes potentially play a critical role in clinical practice, for instance, in cases of eyelid edema, deformities, tumors, thyroid proptosis, etc. The present study introduced a novel standardized method modified from our previous approach (9-11), defining a standardized upper eyelid region for direct measurement of area and volume. Moreover, intrarater, interrater, and intramethod reliability were validated to assess its feasibility in clinical practices. This study adds to the prior literature describing the 3D areal and volumetric measurements and sets the foundation for the application of 3D imaging in the periocular region. This might prompt the development of novel techniques for diagnosing and treating upper eyelid-related diseases and digital innovations for global eye care equity (27,28).

The standardized measurement of area and volume has not been extensively evaluated in the periocular region. Standardization is essential for anthropometric measurements, facilitating comparisons between measurements from various studies, investigators, and even the same investigator. More recently, researchers have been attempting to develop methods to standardize the periocular region for volume measurement. As shown in *Table 3*, volume analysis has been studied in lower blepharoplasty in clinical practice. The medial, central, and lateral lower eyelid compartments have also been divided by the vertical lines tangential to the medial canthus, medial corneal limbus, lateral corneal limbus, and lateral canthus, respectively (15). The upper and lower limits of the lower eyelid compartments have been considered as the ciliary margin of the lower eyelid and the lower margin

of the orbit, respectively. However, the lower margin of the orbit is difficult to define in 3D photos, as it lays beneath the superficial soft tissues. Furthermore, the target region is also expanded to the upper midface. Miller (13) evaluated the 3D volumetric differences between the upper cheek's tear trough and the horizontal line originating from the alar crease. Cristel and Caughlin (14) selected a standardized region for volume analysis, which included the medial canthus to the oral commissure medially, the oral commissure to the anterior hairline directly lateral from the lateral canthus laterally, and the ciliary margin of the lower eyelid to the anterior hairline superiorly.

In contrast, few studies have been carried out for areal and volumetric measurements of the upper eyelid region. As shown in *Table 3*, the supraorbital rim has been considered the superior border of the upper eyelid (20). However, identifying the bony landmarks of the orbital rims might introduce more errors since it is more or less challenging to recognize the bony rims lying under the superficial soft tissue, especially for upper eyelids with a fuller contour. Figueiredo *et al.* (17) studied the lateral brow-eyelid complex, involving the eyelid crease inferiorly, the superior brow follicle line superiorly, a vertical line drawn at the temporal corneal limbus medially, and an oblique line drawn between the lateral canthus and the edge of the lateral brow inferolaterally. Nonetheless, the edge of the lateral brow might vary due to blurred or shaped eyebrows, especially in women. Furthermore, they only researched a part of the upper-eyelid region (i.e., the lateral brow-eyelid complex) rather than the whole upper eyelid. To standardize the areal and volumetric measurement of the whole upper eyelid, we added 6 points to modify an existing standardized method in the current study: 3 landmarks between the medial canthus and the lower brow margin and 3 between the lateral canthus and the lower brow margin in the lateral view. The medial and lateral borders become nearly straight with these 6 points, with minor variation.

Furthermore, to the best of our knowledge, this is the first study to validate the reliability of directly measuring the area and volume of the upper-eyelid region. Direct measurement of the eyelid area and volume allows researchers to define the average reference values in different populations (gender, age, race, etc.), characterize pathological lesions, design surgical approaches, and evaluate treatment outcomes. The current study showed that, for area measurement, excellent agreement was observed in the ICC for intrarater, interrater, and intramethod reliability; very good agreement in the REM

**Table 3** Previous studies on periocular areal and volumetric measurements

Study [year]	Imaging system	Eyelid region definition	Area/volume changes	Study aims
Miller [2016]	Vectra M3	Superior border: the tear trough (and eyelid-cheek junction); medial and lateral borders: perpendicular lines from the medial and lateral canthi; inferior border: a horizontal line from the alar crease intersecting the medial and lateral borders	Average volume gain: 0.64 mL (left side, 0.61 mL; right side, 0.67 mL; P=0.49)	Determining the volumetric effects to the tear trough and deep fat compartments of the upper cheek after lower blepharoplasty with fat repositioning
Cristel and Caughlin [2020]	Vectra XT	Superior border: from the lower eyelid margin to the anterior hair line directly lateral from the lateral canthus; medial and lateral borders: the oral commissure to the medial canthus or the anterior hair line	Mean volume gain: 2.84 mL on the right and 2.87 mL on the left (P=0.95)	Determining quantitative volumetric effects on the lower periorbital region and tear trough (upper midface) after lower blepharoplasty in combination with fat pad transposition and autologous fat grafting
Miranda and Matayoshi [2021]	Vectra H1	Superior border: the ciliary margin of the lower eyelid; inferior border: the lower margin of the orbit; medial border: the vertical line tangential to the endocanthion; lateral border: the vertical line tangential to the lateral canthus	The difference between the postoperative volume and the simulated volume decreased over time. After 6 months, the actual volume was still 0.30 mL (right eye) and 0.24 mL (left eye), and larger than the simulated volume	Assessing the level of agreement between 3D simulations and the outcome of lower eyelid blepharoplasties with fat pad volume removal based on these simulations
Figueiredo <i>et al.</i> [2021]	Vectra H1	Superior border: the superior brow follicle line; inferior border: the eyelid crease; medial border: a vertical line drawn at the temporal corneal limbus; lateral border: an oblique line drawn between the lateral canthus and the lateral brow limit	Mean volume gain: 0.19 mL (control group with upper blepharoplasty alone) and 0.85 mL (internal browpepsy intervention group)	Comparing the lateral brow-eyelid complex volume change before and after blepharoplasty with or without internal browpepsy sutures
Liu <i>et al.</i> [2021] & Liu <i>et al.</i> [2021]	Vectra M3	Five flat paper objects with varying areas (0.16 cm <sup>2</sup> , 0.36 cm <sup>2</sup> , 0.64 cm <sup>2</sup> , 1.44 cm <sup>2</sup> , and 2.56 cm <sup>2</sup> )	ICC: excellent in 21.2%, good in 45.5%, moderate in 27.3%, and poor in 6.1%; MAD <0.3 mm <sup>2</sup> , TEM <1 mm <sup>2</sup> , REM and %TEM <2%	Evaluating the accuracy and reliability of stereophotography for area measurements with standardized objects in varying sizes at different periocular locations
Hyer <i>et al.</i> [2021]	Vectra M3	(I) Upper eyelid to the supraorbital rim, (II) lower eyelid to the infraorbital rim, (III) lower eyelid extended to the tear trough and palpebromalar groove, (IV) both eyelids to the infraorbital rim, and (V) both eyelids extended to the tear trough and palpebromalar groove	Mean volume change: 0.11±0.13 mL for the most accurate registration method of surface area selection; highest ICC: 0.63 (moderate) in interobserver variability	Evaluating the accuracy and interobserver variability of 3D stereophotogrammetric imaging for volumetric assessment in the periorbital region

ICC, intraclass correlation coefficients; MAD, mean absolute difference; TEM, technical error of measurement; REM, relative error of measurement; %TEM, relative TEM; 3D, 3-dimensional.

and %TEM for intra- and interrater reliability; and good agreement in the REM and %TEM for intramethod reliability. These results indicated that the novel method is reliable for area measurement and has great potential for clinical application. Although high accuracy has been validated for measuring the flat objectives' area using this 3D imaging system (18,19), the accuracy of the current approach still needs to be evaluated. In the future, we may compare this standardized landmark localization approach with other standardized approaches and measurement methods (e.g., ImageJ for 2D areal measurement and direct measurement for its true value). Only then can we determine whether the area measurement of this imaging system has value.

Conversely, relatively poor reliability was observed for volumetric measurement. Although an excellent agreement was found in the ICCs for intra- and interrater reliability, there was poor agreement in the intra- and interrater REM and %TEM and poor agreement in the ICC, REM, and %TEM for intramethod reliability. Thus far, Hyer *et al.* have conducted the only study to evaluate the reliability of upper-eyelid volume (20). In their study, the orbital rim was defined as the superior limit of the upper eyelid. Furthermore, the volume analyses that differed from ours were evaluated, which were based on the superimposition of paired images registered by the two methods; that is, surface area selection (the forehead and bridge of the nose) and landmark selection (the bilateral subalar, alar curvature point, subnasale, pronasale, and nasion). The mean volume (mL) changes in upper eyelid volume between the paired images were  $0.44 \pm 0.37$  when the eyes were open and  $0.51 \pm 0.32$  when the eyes were closed during landmark registration; and  $0.15 \pm 0.17$  when the eyes opened and  $0.16 \pm 0.20$  when eyes closed during surface area selection. Compared to landmark registration ( $0.55 \pm 0.47$  mL), the surface area selection showed the most accurate mean change in upper-eyelid volume ( $0.15 \pm 0.20$  mL;  $P < 0.001$ ) despite the small effect size ( $\eta^2 < 0.260$ ). Surprisingly, our study achieved worse results for measuring volume than it did for area. For direct measurement, our intra- and interrater results indicated that the MAD and TEM estimates were 0.2299 and 0.2414 mL for intrarater reliability, 0.2749 and 0.3253 mL for interrater reliability, and 1.6716 and 2.0615 mL for intramethod reliability, respectively. The intra- and interrater volume differences were slightly higher than the volume changes found when the surface area selection procedure mentioned above was used. However, the intramethod results showed substantial

errors.

The less reliable volumetric value might be attributable to various factors. For one, we found many pairs of cases with opposite arithmetic signs (22/88, 25%) when evaluating the intramethod volume differences (77.9% and 96.1% for REM and %TEM, respectively). In contrast, no pairs were found in the intrarater volume differences (10.3% and 10.8% for REM and %TEM, respectively), and only 1 pair was observed in the interrater volume differences (12.3% and 14.6% for REM and %TEM, respectively). The opposite arithmetic signs directly lead to significantly larger MAD and TEM values and subsequently, higher REM and %TEM. We speculated that this might be due to the anatomical characteristics of the broad eyelid base, including the outer and inner surfaces, which make it challenging to select the entire eyelid region in 3 dimensions. Hence, the accuracy and reliability of volumetric measurements are higher when assessing an object with a relatively narrow and independent base (e.g., the palm, arm, thigh, auricle, etc.) For another, this might also be attributable to the algorithms of the 3D imaging systems for 3D structural reconstruction and direct volumetric measurement techniques. In direct volumetric measurements, the primary condition is that the target has a datum plane on which the elevation or depression will be deemed a positive or negative volume value, and the volume on this plane will be zero. The lack of an efficient, fixed datum plane for volumetric measurements in the eyelid region results in some reconstructed 3D soft tissue datum plane deviations in intramethod errors. In addition, coupled with errors in the placement of landmarks resulting in different selected structural regions, there may be significant variations in the datum plane, leading to differences in positive and negative values. To address the insufficient reliability of volume measurement, we propose the following. First, the area change can be considered an indirect indicator of volume change since the area can reflect the volume change to some extent. Second, for volume and surface area measurement of eyelid skin tumors, an upper-lid skin surface may be simulated at the tumor base as a reference plane for volumetric measurement. Third, fixed third-party reference plane on which the upper lid protrudes or depresses can be identified as the basis for volumetric measurement, which can be used as a relative measure for volume calculation. Unfortunately, we have not put them into practice to date due to the complexity of the eyelid's anatomical characteristics and the professionalization of volumetric measurement algorithm. Therefore, more collaboration

should be built with engineers or professionals in computer science.

Moreover, results of previous studies suggest that it may also be more clinically feasible for surgical outcome assessment to superimpose two 3D images before and after surgery, (13-15,20). Additionally, the reliability of this novel modified area selection strategy and the method mentioned above based on the orbital rim should be compared in future study. The direct measurement might also be compared with registration-based volumetric change measurement to provide further information on the applicability of 3D volume measurement. Finally, areal and volumetric measurements are the most promising function of the 3D imaging system, and the various potential applications should be explored. The current 3D photographic system is not perfect. A professional and reliable landmark system is the first step to accurate measurement. The next step should be to seek collaboration with professionals to ascertain how to improve the accuracy and reliability issues of volumetric measurements. In the current direct volumetric measurement results, although there was no statistically significant difference in the overall measurement value, the individual measurement value gap was substantial. Hence, remeasurement is recommended when the gap is overly large, especially when positive and negative signs occur.

As shown in *Table 3*, several 3D imaging systems have been studied for volumetric measurement in the periocular region, especially the lower eyelid. Although superimposition for volume change evaluation rather than direct measurement has been performed, these studies may provide some inspiration and ideas for volumetric analysis of the upper eyelids. The Vectra M3 3D imaging system is the most commonly studied 3D camera. Miller reported an average upper cheek volume gain of 0.64 mL (left side, 0.61 mL; right side, 0.67 mL) in patients who had undergone lower blepharoplasty with fat repositioning (13). Hyer *et al.* evaluated its accuracy and interobserver variability for volumetric assessment in the upper and lower eyelid regions (20); they found a mean change in volume of  $0.11 \pm 0.13$  mL for the surface-based registration method and confirmed the agreement between observers (highest ICC, 0.63). However, considering that the postoperative volume changes might be minimal and expressed in tenths of a milliliter in the periocular region, the abovementioned statistical estimates for accuracy and reliability might still require further improvement for future applications. Regarding the Vectra XT 3D imaging and analysis system, Cristel and Caughlin's study showed a mean volume gain

of 2.84 and 2.87 mL on the periocular region's right and left side after lower blepharoplasty combined with fat pad transposition and fat grafting (14). Therefore, they concluded that this is an effective procedure for improving the tear trough deformity and the periocular area. They also proposed that 3D volume assessment can accurately measure the volume changes of the postoperative results even though they did not evaluate the reliability and accuracy of this technique. The Vectra H1 3D imaging and analysis system has also been evaluated for its reliability in volumetric simulation (15). Preoperatively, the researchers simulated the volume reduction procedure during a lower eyelid blepharoplasty until a satisfactory aesthetic outcome was achieved. Postoperatively, they juxtaposed the postoperative images with the respective preoperative simulation images to calculate volume differences in the lower eyelids. The Vectra H1 software led to postoperative values systematically larger than the preoperative simulation values. Furthermore, a low level of agreement was observed between the simulated volume using Vectra H1 mirror software and the actual volume resected during lower eyelid blepharoplasties (i.e., in the right eye: 0.253 *vs.* 0.279 mL;  $P=0.046$ ). Hence, they concluded that the accuracy of the tested software (Vectra H1 mirror software) is insufficient for reliably simulating the volume of fat during lower lid blepharoplasty, making the Vectra H1 simulation software an inadequate tool for planning surgery and managing the patient's expectations. At present, there is a lack of studies comparing the different 3D imaging systems in periocular region volumetric measurement. However, our standardized approach for upper-eyelid region selection facilitates comparisons of different studies or 3D imaging systems. Therefore, given the varying results in terms of volumetric accuracy and reliability, 3D imaging systems might be improved to predict the volume required for adequate surgical outcomes and to reduce the risk of complications.

Concerning the limitations of this study, it should be pointed out that only healthy White adult volunteers were enrolled. It worth noting that the differences between various populations are probably more significant than those of interobserver differences despite the fact that all observers in the study received professional training and were assessed to minimize interobserver differences when measuring the same population. Moreover, the measurement error may be larger when measuring the facial structural features of different populations. For one, the facial characteristics of different ethnic groups are disparate, resulting in differing sizes of the measured values of the

same structure. Thus, the sizes of the reliability evaluation indicators are different. For example, large-sized measured values have large MAD and TEM values but small REM and %TEM values (12). For another, it is important to note that different-sized anatomical structures will lead to differences in the quality of 3D imaging, and thus, dissimilarities in imaging errors. For example, due to the occlusion of the tall nose bridge and the prominent eyebrow arch of White individuals, the 3D periorbital structure images of the medial canthus and the lower part of the eyebrows captured and synthesized by the 3D camera are easily distorted or blurred, thereby increasing the imaging errors. Therefore, it remains to be further investigated whether the results can be maintained in populations with different ethnicities and diseases. Furthermore, we only evaluated the reliability of direct volume measurement using this modified standardized area selection strategy. In contrast, it might be more reliable to measure volume changes after superimposing and registering perioperative images. Therefore, further studies are needed to evaluate the accuracy and reliability of the standardized strategy when it is combined with image superimposition and registration. It will also be of clinical interest to compare the reliability of various current eyelid area selection procedures during eyelid volume measurement. Our group is also measuring periorbital morphologic changes, including lines, curves, angles, areas, and volumes, related to different ages in a White population. In the future, we propose to apply this method in a clinical setting to assess its feasibility and evaluate its treatment outcomes by measuring changes in the upper-eyelid area and volume before and after treatment in patients with eyelid deformities, tumors, thyroid proptosis, and other conditions.

## Conclusions

This study adds to the previously published literature describing the areal and volumetric measurements of the periorbital region. We introduced and validated a novel modified approach for standardized 3D evaluation of upper-eyelids area and volume, and confirmed the excellent or very good reliability of areal measurement and the poor reliability of volumetric measurements based on the direct measurement approach on a single image. Nonetheless, better results may be achieved by combining the present method with image overlapping and registration processing methods, which is subject to further validation. Moreover, a comparative study with other eyelid area selection methods

using the orbital rim as the boundary should be performed in the future. Regardless, it can be concluded that the novel method presented in this study could provide quantitative data for areal and volumetric measurement of the upper eyelids and might have widespread clinical application potential in the future.

## Acknowledgments

*Funding:* This work was supported by the National Natural Science Foundation of China (No. 82102346 to Y.G., and No. 82000948 to L.L.) and the Fundamental Research Funds for the Central Universities (No. 2021FZZX005-15 to Y.G.).

## Footnote

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-22-589/coif>). The authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the institutional ethics board of the University of Cologne, and informed consent was obtained from all individual participants.

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

## References

1. Sforza C, Dolci C, Grandi G, Tartaglia GM, Laino A, Ferrario VF. Comparison of soft-tissue orbital morphometry in attractive and normal Italian subjects. *Angle Orthod* 2015;85:127-33.

2. Farkas LG. Basic morphological data of external genitals in 177 healthy Central European men. *Am J Phys Anthropol* 1971;34:325-8.
3. Hou X, Rokohl AC, Meinke MM, Li S, Liu J, Fan W, Lin M, Jia R, Guo Y, Heindl LM. A novel standardized distraction test to evaluate lower eyelid tension using three-dimensional stereophotogrammetry. *Quant Imaging Med Surg* 2021;11:3735-48.
4. Hou X, Rokohl AC, Meinke MM, Liu J, Li S, Fan W, Lin M, Jia R, Guo Y, Heindl LM. Standardized Three-Dimensional Lateral Distraction Test: Its Reliability to Assess Medial Canthal Tendon Laxity. *Aesthetic Plast Surg* 2021;45:2798-807.
5. Hou X, Rokohl AC, Meinke MM, Zhang M, Guo Y, Heindl LM. Digital Photogrammetry for Assessing Medial Canthal Tendon Laxity: Novel Standardized Three-Dimensional Versus Traditional Two-Dimensional Lateral Distraction Test. *J Oral Maxillofac Surg* 2022;80:1033-9.
6. Bartella AK, Laser J, Kamal M, Krause M, Neuhaus M, Pausch NC, Sander AK, Lethaus B, Zimmerer R. Accuracy of low-cost alternative facial scanners: a prospective cohort study. *Oral Maxillofac Surg* 2022. [Epub ahead of print]. doi: 10.1007/s10006-022-01050-5.
7. Kwon SH, Choi JW, Kim HJ, Lee WS, Kim M, Shin JW, Na JI, Park KC, Huh CH. Three-Dimensional Photogrammetric Study on Age-Related Facial Characteristics in Korean Females. *Ann Dermatol* 2021;33:52-60.
8. Chong Y, Li J, Liu X, Wang X, Huang J, Yu N, Long X. Three-dimensional anthropometric analysis of eyelid aging among Chinese women. *J Plast Reconstr Aesthet Surg* 2021;74:135-42.
9. Guo Y, Rokohl AC, Schaub F, Hou X, Liu J, Ruan Y, Jia R, Koch KR, Heindl LM. Reliability of periocular anthropometry using three-dimensional digital stereophotogrammetry. *Graefes Arch Clin Exp Ophthalmol* 2019;257:2517-31.
10. Guo Y, Liu J, Ruan Y, Rokohl AC, Hou X, Li S, Jia R, Koch KR, Heindl LM. A novel approach quantifying the periorbital morphology: A comparison of direct, 2-dimensional, and 3-dimensional technologies. *J Plast Reconstr Aesthet Surg* 2021;74:1888-99.
11. Guo Y, Schaub F, Mor JM, Jia R, Koch KR, Heindl LM. A Simple Standardized Three-Dimensional Anthropometry for the Periocular Region in a European Population. *Plast Reconstr Surg* 2020;145:514e-23e.
12. Guo Y, Hou X, Rokohl AC, Jia R, Heindl LM. Reliability of Periocular Anthropometry: A Comparison of Direct, 2-Dimensional, and 3-Dimensional Techniques. *Dermatol Surg* 2020;46:e23-31.
13. Miller TR. Long-term 3-Dimensional Volume Assessment After Fat Repositioning Lower Blepharoplasty. *JAMA Facial Plast Surg* 2016;18:108-13.
14. Cristel RT, Caughlin BP. Lower Blepharoplasty Three-Dimensional Volume Assessment after Fat Pad Transposition and Concomitant Fat Grafting. *Facial Plast Surg* 2020;36:478-83.
15. Miranda RE, Matayoshi S. Vectra 3D Simulation in Lower Eyelid Blepharoplasty: How Accurate is it? *Aesthetic Plast Surg* 2022;46:1241-50.
16. Mourits DL, Remmers JS, Tan SH, Moll AC, Hartong DT. An Individualized 3-Dimensional Designed and Printed Conformer After Dermis Fat Grafting for Complex Sockets. *Ophthalmic Plast Reconstr Surg* 2018;34:390-2.
17. Figueiredo MN, Limongi RM, Tao JP. A Prospective Randomized Study of Lateral Brow-Eyelid Complex Volume after Internal Browpepxy Using Three-Dimensional Stereophotogrammetry. *Facial Plast Surg Aesthet Med* 2021. [Epub ahead of print]. doi: 10.1089/fpsam.2020.0586.
18. Liu J, Rokohl AC, Guo Y, Li S, Hou X, Fan W, Formuzal M, Lin M, Heindl LM. Reliability of Stereophotogrammetry for Area Measurement in the Periocular Region. *Aesthetic Plast Surg* 2021;45:1601-10.
19. Liu J, Guo Y, Arakelyan M, Rokohl AC, Heindl LM. Accuracy of Areal Measurement in the Periocular Region Using Stereophotogrammetry. *J Oral Maxillofac Surg* 2021;79:1106.e1-9.
20. Hyer JN, Murta F, Juniat VAR, Ezra DG. Validating three-dimensional imaging for volumetric assessment of periorbital soft tissue. *Orbit* 2021;40:9-17.
21. Wang L, Fang H, Shao A, Zhang H, Ye J. Eyelid squamous cell carcinoma in the setting of epidermodysplasia verruciformis (EV) diagnosed by next-generation sequencing: A case report and literature review. *Adv Ophthalmol Pract Res* 2022;2:100066.
22. Hou X, Guo Y, Li S, Lin M, Jia R, Rokohl A, Heindl LM. Lateral tarsal strip procedure for involutional ectropion: A retrospective analysis of 85 cases and a comprehensive literature review. *Adv Ophthalmol Pract Res* 2021;1:100004.
23. Kwok SY. A simple procedure for involutional ectropion repair. *Adv Ophthalmol Pract Res* 2022;2:100045.
24. Liu J, Rokohl AC, Liu H, Fan W, Li S, Hou X, Ju S, Guo Y, Heindl LM. Age-related changes of the periocular

- morphology: a two- and three-dimensional anthropometry study in Caucasians. *Graefes Arch Clin Exp Ophthalmol* 2023;261:213-22.
25. Guo Y, Rokohl AC, Lin M, Heindl LM. Three-dimensional anthropometry in periorbital region. *Annals of Eye Science* 2020;6:8.
  26. Camison L, Bykowski M, Lee WW, Carlson JC, Roosenboom J, Goldstein JA, Losee JE, Weinberg SM. Validation of the Vectra H1 portable three-dimensional photogrammetry system for facial imaging. *Int J Oral Maxillofac Surg* 2018;47:403-10.
  27. Zheng S, Choo J, Chen J, Joshi S, Sun Z. The barriers to global eye care equity and the role of digital innovations. *Adv Ophthalmol Pract Res* 2021;1:100021.
  28. Jin K, Ye J. Artificial intelligence and deep learning in ophthalmology: Current status and future perspectives. *Adv Ophthalmol Pract Res* 2022;2:100078.

**Cite this article as:** Guo Y, Rokohl AC, Fan W, Theodosiou R, Li X, Lou L, Gao T, Lin M, Yao K, Heindl LM. A novel standardized approach for the 3D evaluation of upper eyelid area and volume. *Quant Imaging Med Surg* 2023;13(3):1686-1698. doi: 10.21037/qims-22-589

**Table S1** Definition of abbreviations for landmarks

Abbreviation	Definition
En	Endocanthion
EEn	Inferior margin landmark of the eyebrow aligning to the En
Lm	Medial corneoscleral limbus
Lm'	Lash roots landmark on the upper palpebral margin aligning to the Lm
ELm	Inferior margin landmark of the eyebrow aligning to the Lm
Um	Middle point between En and Lm'
EUm	Inferior margin landmark of the eyebrow aligning to Um
Pc	Pupillary center
Ps	Lash roots landmark on the upper palpebral margin aligning to the Pc
EPs	Inferior margin landmark of the eyebrow aligning to the Ps
LI	Lateral corneoscleral limbus
LI'	Lash roots landmark on the upper palpebral margin aligning to the LI
ELI	Inferior margin landmark of the eyebrow aligning to the LI
UI	Middle point between Ex and LI'
EUI	Inferior margin landmark of the eyebrow aligning to UI
Ex	Exocanthion
EExL	Inferior margin landmark of the eyebrow aligning to the Ex in the lateral review