Journal of Cranio-Maxillo-Facial Surgery 47 (2019) 1031–1040

Contents lists available at ScienceDirect

Journal of Cranio-Maxillo-Facial Surgery

journal homepage: www.jcmfs.com

Feasibility of iterative closest point algorithm for accuracy between virtual surgical planning and orthognathic surgery outcomes

Daniel Amaral Alves Marlière [a, *], Maurício Silva Demétrio [b], Francielle Silvestre Verner [c], Luciana Asprino [d], Henrique Duque de Miranda Chaves Netto [e]

[a] Oral and Maxillofacial Surgeon, Division of Oral and Maxillofacial Surgery, Piracicaba Dental School, State University of Campinas, Piracicaba, São Paulo, Brazil
[b] Oral and Maxillofacial Surgeon and Oral and Maxillofacial Radiologist, Private Office, São Luís, Maranhão, Brazil
[c] Department of Dentistry, Federal University of Juiz de Fora/Governador Valadares Campus, Governador Valadares, Minas Gerais, Brazil
[d] Division of Oral and Maxillofacial Surgery, Piracicaba Dental School, State University of Campinas, Piracicaba, São Paulo, Brazil
[e] Division of Oral and Maxillofacial Surgery, University Hospital of Federal University of Juiz de Fora, Minas Gerais, Brazil

ARTICLE INFO

Article history:
Paper received 23 November 2018
Accepted 21 March 2019
Available online 28 March 2019

Keywords:
Computer-assisted design
Computer-generated 3D imaging
Dimensional measurement accuracy
Orthognathic surgery

ABSTRACT

Purpose: To evaluate the feasibility of iterative closest point (ICP) algorithm for assessing the accuracy between virtual surgical planning (VSP) and outcomes in orthognathic surgery.
Materials and methods: VSP and results of surface mesh (SM0 and SM1) from CBCT scans of 25 patients who had been undergone bi-maxillary orthognathic surgery were converted into STL-format files and then imported to Geomagic software for semi-automatic alignment. ICP algorithm was used to calculate mean deviations (MD) and root mean square (3D error) at different calibrations of ±2 mm (T1), ±5 mm (T2) and ±10 mm (T3), with workflow being performed by two evaluators. Colour maps were generated to assess the 3D congruence qualitatively. Linear regression analysis was used to estimate whether SM0 or SM1 could condition the ICP and *t*-tests were used to assess whether MD and 3D error values were ≤ -2 mm and ≥ 2 mm. Descriptive statistics was used to assess the method's feasibility by comparing T2 to T1 and T3.
Results: High intra- and inter-rater correlations supported the workflow reproducibility with the software. SM0 conditioned the ICP algorithm regarding both evaluators, and *t*-tests demonstrated that MD and 3D error were > -2 mm and < 2 mm. MD and 3D error at T3 were 30% higher than those at T1.
Conclusions: ICP algorithm provided a reproducible method, but its feasibility was limited due to underestimation or overestimation of the results as they affect the validity of the actual deviations.

1. Introduction

Since the introduction of cone-beam computed tomography (CBCT) and the advances in three-dimensional (3D) technologies, a series of new computerised tools have been developed for use in virtual surgical planning (VSP) (Gaber et al., 2017). There are numerous studies on the measurement accuracy between VSP and orthognathic surgery outcomes regarding facial hard tissues (Stokbro et al., 2014; Haas et al., 2015). Gaber et al. (2017) performed a systematic review on methods for accuracy assessment and critically reported on their validation, suggesting a universal protocol for evaluation of VSP and outcomes with the least margin of error as possible. They stated that an ideal measurement of the accuracy between VSP and outcomes could be achieved through voxel-based registration (pre- and post-operative images), automated and semi-automated evaluation, and use of inter- and intra-rater reliability to validate the results.

Studies have currently shown four types of methods to evaluate 3D changes of the skeletal hard tissue after orthognathic surgery. There are studies using 3D cephalometric analysis (Terajima et al., 2008), principal component analysis and thin-plate spline analysis (Cevidanes et al., 2006), volumetric changes (Maal et al., 2012) and colour error maps (Hernández-Alfaro and Guijarro-Martínez,

* Corresponding author. Piracicaba Dental School – State University of Campinas, Division of Oral and Maxillofacial Surgery, Limeira Avenue, 901, Areião. Piracicaba, São Paulo, 13414-903, Brazil.
E-mail address: ctbmf.marliere@gmail.com (D.A.A. Marlière).

https://doi.org/10.1016/j.jcms.2019.03.025
1010-5182/

2013; Tucker et al., 2010). However, 3D cephalometric measurements require the identification of the same landmarks multiple times on the 3D surface (Gaber et al., 2017; Baan et al., 2016), which does not eliminate human error totally. Moreover, additional error might be generated from DICOM-format data (Digital Imaging and Communication in Medicine) (Jabar et al., 2015). The volume-change method does not adequately describe changes in bone position and does not quantify the measurement of direction or magnitude, whereas thin-plate spline uses statistical analysis (Jabar et al., 2015).

For analysis of 3D images (VSP and outcomes), the 3D models developed from DICOM-format images (i.e. from CBCT) can be converted into 3D surface meshes, which can simply be polygonal or triangular surface meshes (i.e. thousands of triangle's vertices or points) and submitted to 3D image superimposition method. For instance, methods of analysis have been reported to use surface-based registration or superimposition in which two 3D surface meshes are approximated and the differences between all the triangle's vertices are computed based on the Euclidean distances (Jabar et al., 2015; Khambay and Ullah, 2015). The entire workflow can be performed and computed by using an iterative closest point (ICP) algorithm, which has been widely used for various medical and research purposes, including diagnoses, treatment planning and assessment of a variety of cases involving CBCT (Almukthar et al., 2014).

The ICP algorithm is a commonly used method for assessing changes in 3D surface meshes by measuring the point-to-point distance between the surfaces and by generating colour distance maps. The question of whether ICP algorithm would be a feasible method for assessing the accuracy between 3D surface meshes of hard tissues is raised.

Therefore, our aim was to evaluate the feasibility of ICP algorithm in a reverse engineering software by using 3D surface meshes in an attempt to develop a method for assessing the accuracy between VSP and actual outcomes in orthognathic surgery.

2. Materials and methods

The present study has used DICOM-format images of 25 adult patients who had undergone bi-maxillary orthognathic surgery at the Clinics Hospital of the Federal University of Juiz de Fora between October 2015 and April 2017. The pre-operative CBCT scans were taken within one month of surgery and the post-operative scans were obtained immediately after the surgery. The images were acquired with i-CAT scanner (Imaging Sciences International LLC, Hatfield, PA, USA) operating at 120 kVp, 5 mA, FOV of 22 × 16 cm, scanning time of 40 s, isotropic voxel size of 0.4 mm, and 14-bit grey scales. The DICOM images were imported into Dolphin Imaging software (Version 11.7 Premium software, Dolphin Imaging and Management solutions, Chatsworth, CA, USA), which was performed in the same workflow of VSP. The surgical procedures were conducted by the same team of surgeons following the same process for all patients. This study was approved by the Research Ethics Committee of the Federal University of Juiz de Fora regarding the use of data and carried out according to the ethical principles and Declaration of Helsinki (Certificate of Presentation for Ethical Consideration: 69598017.2.0000.5133).

The DICOM images of each patient were selected according to the following inclusion criteria: (I) availability of pre- and post-operative CBCT data imported into Dolphin imaging software; (II) availability of VSP; (III) bi-maxillary orthognathic surgery through Le Fort I osteotomy and bi-lateral sagittal split osteotomy. Exclusion criteria were: patient's CBCT image showing (I) craniofacial syndromic abnormalities, cleft palate, degenerative condylar disease, sequels of facial trauma and previous history of Le Fort I osteotomy or bi-lateral sagittal split osteotomy; and (II) additional surgery at the time, such as multi-segment Le Fort I osteotomy, chin osteotomies, mandibular sub-apical osteotomy and trans-oral vertical ramus osteotomy. This study was limited to single-segment Le Fort osteotomy and bi-lateral sagittal split osteotomy in order to facilitate the comparative evaluation of deviations without the influence of other osseous changes or occlusal guide plate.

The 3D hard tissue models were segmented from each pre- and post-operative DICOM image by using Dolphin imaging software before being exported as 3D files of the surface mesh, whose thousands of vertices and triangles were entered into an electronic spreadsheet (Microsoft Excel, Microsoft Office 365, Redmond, WA, USA). All the 3D files of the surface mesh were acquired with the patient's head in a natural position. We used two laptops (Intel Core i7-4770 CPU, 3.2 GHz Turbo Boost, 8 Gb main memory; Intel, Acer, SP, Brazil) and Windows 8 as the operating system (Microsoft, Redmond, WA, USA).

2.1. Acquisition of 3D surface meshes

After VSP in Dolphin Imaging, each pre-operative DICOM image was converted into a 3D surface mesh and segmentedly saved as STL-format files (stereolithography format). One observer (M.S.D.) imported all segmented STL files into Materialise Magics software (Materialise NV, Lovaine, Belgium) for matching by using selected parts. Next, the reference 3D meshes were generated and then exported as STL-format files, categorized as SM0 (Fig. 1AB).

Each post-operative DICOM image was superimposed over the pre-operative DICOM image by the same observer (M.S.D.), who used Dolphin Imaging software. Axial, sagittal and coronal slices of the CBCT volumes were used to select the anatomical structures of the skull base. Next, the post-operative DICOM image was optimally aligned in relation to the pre-operative one by using the superimposition method (Fig. 1C). This voxel-based superimposition was used to maintain the same pre- and post-operative head position (Weissheimer et al., 2015; Borba et al., 2016; Koerich et al., 2016; Ritto et al., 2017). After performing the superimposition, one of the observers (M.S.D) exported the resulting surface meshes (categorized as SM1) into STL-format files by using Dolphin Imaging software, which enabled creating a surface from the volume at full resolution (Fig. 1D).

The STL-format files of both 3D surface meshes (i.e. SM0 and SM1) obtained by using Dolphin Imaging and Materialise Magics software were imported into Geomagic Wrap 2013 software (3D System, California, USA) for analysis of the deviations between SM0 and SM1.

2.2. Registration of the 3D surface mesh

For each pair of 3D surface meshes in STL format, sets of both SM0 and SM1 were imported into Geomagic Wrap software. First, a manual superimposition was performed with this software in which frontal, zygomatic and orbit regions were the references between SM0 and SM1. The software object mover tool was used to approximate the 3D surface meshes (Fig. 2). Next, the surface-based registration function of the software was used to superimpose datasets in the following steps: SM0 (as a fixed 3D model) and SM1 (as a floating 3D surface mesh) were submitted to the best fit alignment tool; calculation control options were set as follows: sample size of 5000, tolerance value of 0; and verification of symmetry (exhaustive), fine adjustments, high precision mounting and elimination of automatic diverter. The software used the ICP algorithm to adjust the position of the floating 3D surface mesh (SM1)

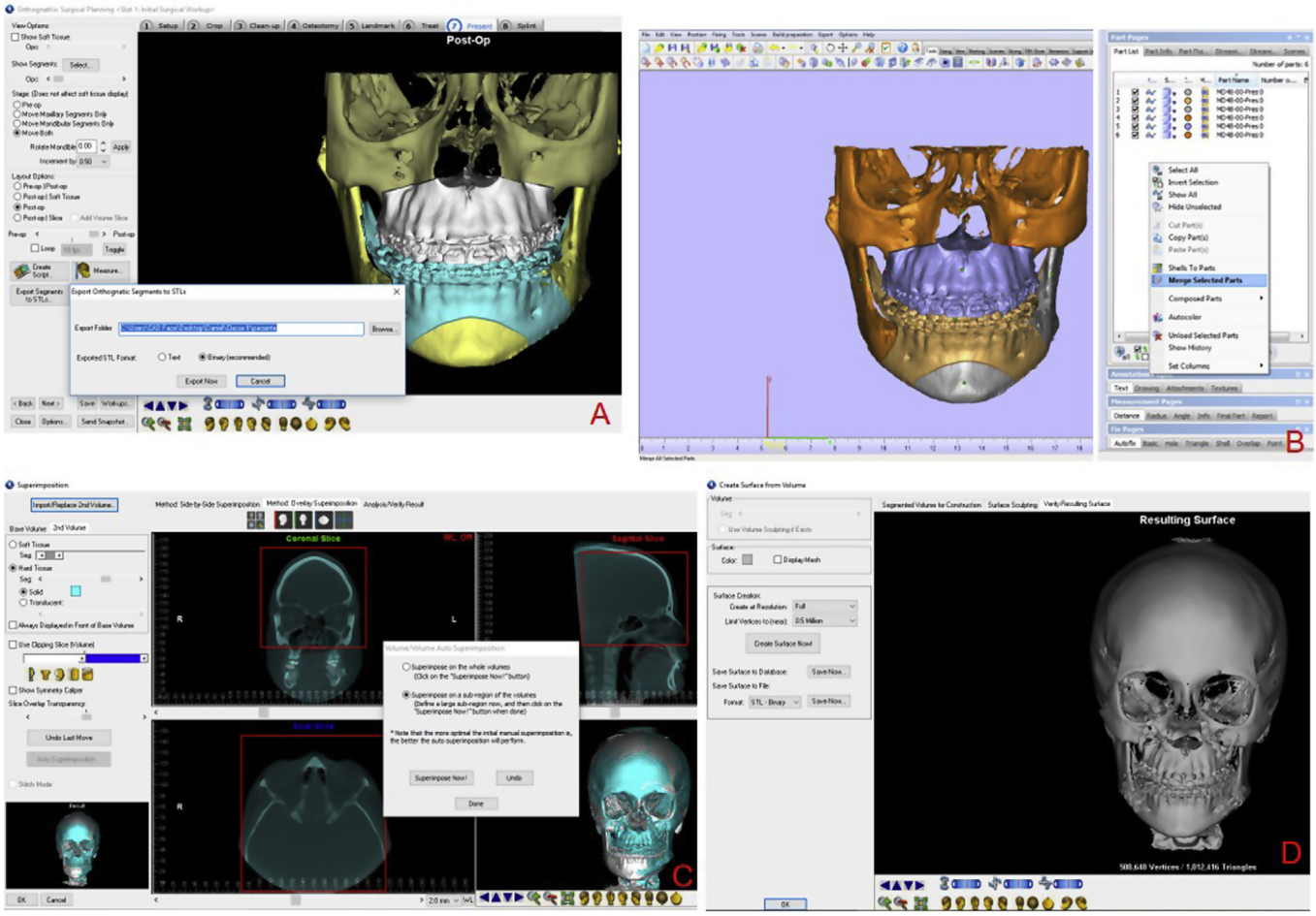

Fig. 1. (A) STL files segmented were exported from Dolphin Imaging (B) Merge selected part tool on Materialise Magics to generate SM0 (C) Voxel-based superimpose CBCT pre and postoperative on Dolphin Imaging and (D) Create surface from volume tool generated SM1.

automatically by exactly overlapping it over the fixed 3D surface mesh (SM0). This semi-automatic alignment presented maximum length (ML) of ICP algorithm and values of registration errors, such as mean error and RMS error (Fig. 2).

2.3. Deviation analysis for 3D comparison and accuracy evaluation

To evaluate the deviations between SM0 and SM1, the deviation analysis function was applied by using the ICP algorithm with the same software for 3D comparison and computed deviations between reference and resulting 3D surface meshes. This software function was calibrated with critical angle of 45°, fine resolution, 14-colour scale and three different maximum and minimum deviation values stipulated (±2 mm, ± 5 mm and ±10 mm, respectively). The ICP algorithm calculated the closest point distance between thousands of surface triangles in the 3D surface meshes (SM0 and SM1), providing the colour-coded surface distance maps so that the measurements could be quantified as mean deviation (MD), positive mean deviation (MD+), negative mean deviation (MD−), standard deviation (SD) and root mean square (RMS). Each pair of 3D surface meshes and the measurements of MD, MD+, MD-, SD, and RMS were categorised into three groups according to the maximum and minimum deviations stipulated (T1 for ± 2 mm; T2 for ± 5 mm; T3 for ± 10 mm). Colour difference images were obtained to examine the congruency of SM0 and SM1 qualitatively (Fig. 3).

All deviations between the closest point pairs in SM1 and SM0 were matched and calculated automatically by the algorithm of the software. The value of RMS was calculated by using the following formula:

$$\mathrm{RMS} = \sqrt{\frac{1}{2}\sum_{i=1}^{n} x_i^2}$$

If point P in SM0 has the closest point P′ in SM1, then X_n is the distance between P and P', and n is the total number of point pairs in both 3D surface meshes. RMS involved the following steps: (1) all deviation values were squared; (2) the squares were added together and their average was calculated, and (3) the square root of the resulting average was estimated. The RMS was defined as the 3D error, which could serve as a measurement indicator of how far deviations between two different datasets vary from zero.

These workflows were performed twice by two evaluators (D.A.A.M and H.D.M.C.N) and repeated after 10 days to check the alignment reproducibility of the tool for registration and deviation analysis. The results were exported into an electronic spreadsheet (Microsoft Excel, Microsoft Office 365, Redmond, WA, USA).

2.4. Statistical analysis

Statistical data analysis was performed with R Core Team software (Version 3.4.2, R Foundation for Statistical Computing,

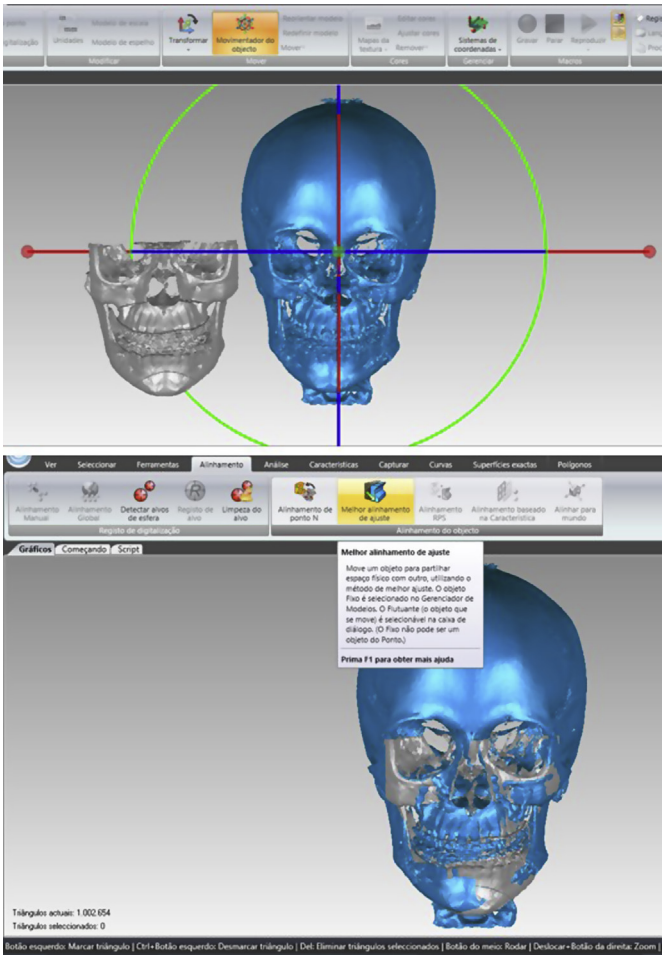

Fig. 2. (A) Illustration showing manual superimposition by object mover tool on Geomagic software. (B) Illustration showing surface-based registration (semi-automatic alignment) by best-fit alignment tool on Geomagic Wrap software.

Vienna, Austria). Pearson's correlation coefficients (Pcc – r) were calculated to assess the intra- and inter-rater agreements (evaluators 1 and 2) regarding the calibration method of automatic alignment (registration error) and deviation analysis function (MD, MD+, MD−, SD and 3D error) between SM0 and SM1.

The means and standard deviations (SD) were calculated to ML, mean error, RMS error, MD, MD+, MD-, SD and 3D error (deviations between M0 and M1). Paired t-tests were used to compare the difference in measurements between intra- and inter-raters for the null hypothesis of similarity between the above-mentioned means (H_0: $\mu = \mu_0$).

Linear regression analysis was performed to estimate whether the number of vertices in SM0 and SM1 could interfere with the ICP algorithm during automatic surface-based registration. The analysis also assessed whether ML values would be conditioned by either SM0 or SM1. The models presented coefficient values, standard error, determination coefficient (r^2) and Schwarz information criterion (SIC), which were submitted to t-tests and F-tests for both evaluators.

Descriptive statistical analysis was performed to evaluate the method's feasibility in which T2 was considered as a standard measurement to compare the difference in the percentages of T1 and T3 for MD and 3D Error. Rejection of the null hypothesis was determined by t-tests, with mean values of MD, MD+, SD and 3D Error always equal to or higher than 2 mm (H_0: $\mu \geq 2$ mm). As for an alternative null hypothesis, mean values of MD and MD-always equal to or lower than −2 mm (H_0: $\mu \leq -2$ mm) were used for evaluating whether the clinical results in orthognathic surgery were kept for deviations $\leq$ 2 mm between VSP and actual outcomes. Statistical significance was set at $p < 0.05$.

3. Results

The resulting surface meshes were superimposed on the reference surface meshes to generate different visual displays of the magnitude and location of incongruence or congruence between SM0 and SM1. Colour variances were seen in the maps generated at T1, T2 and T3. At T1, most of the skull base and frontal, zygomatic and orbital regions showed variability of colours, and maximum and minimum deviations appeared around these regions, indicating colour scales ranging from extreme (red) to no (green) deviation. In addition, maximum and minimum deviations also appeared around proximal and distal segments in the maxilla and mandible, showing a range of colours with different magnitudes (Fig. 4). For T2 and T3 colour maps, there was a similar dispensation of colour surfaces with slight magnitude difference. Regardless of

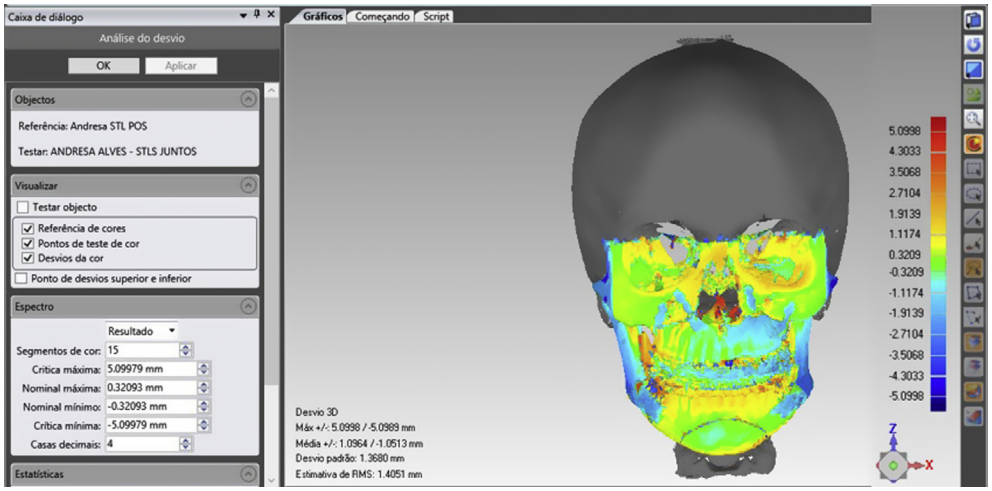

Fig. 3. Application of ICP algorithm between SM0 and SM1 and color map generation. Color maps overview qualitative deviations between SM0 and SM1 and histograms ±5 (T2) sign depicts the deviation range.

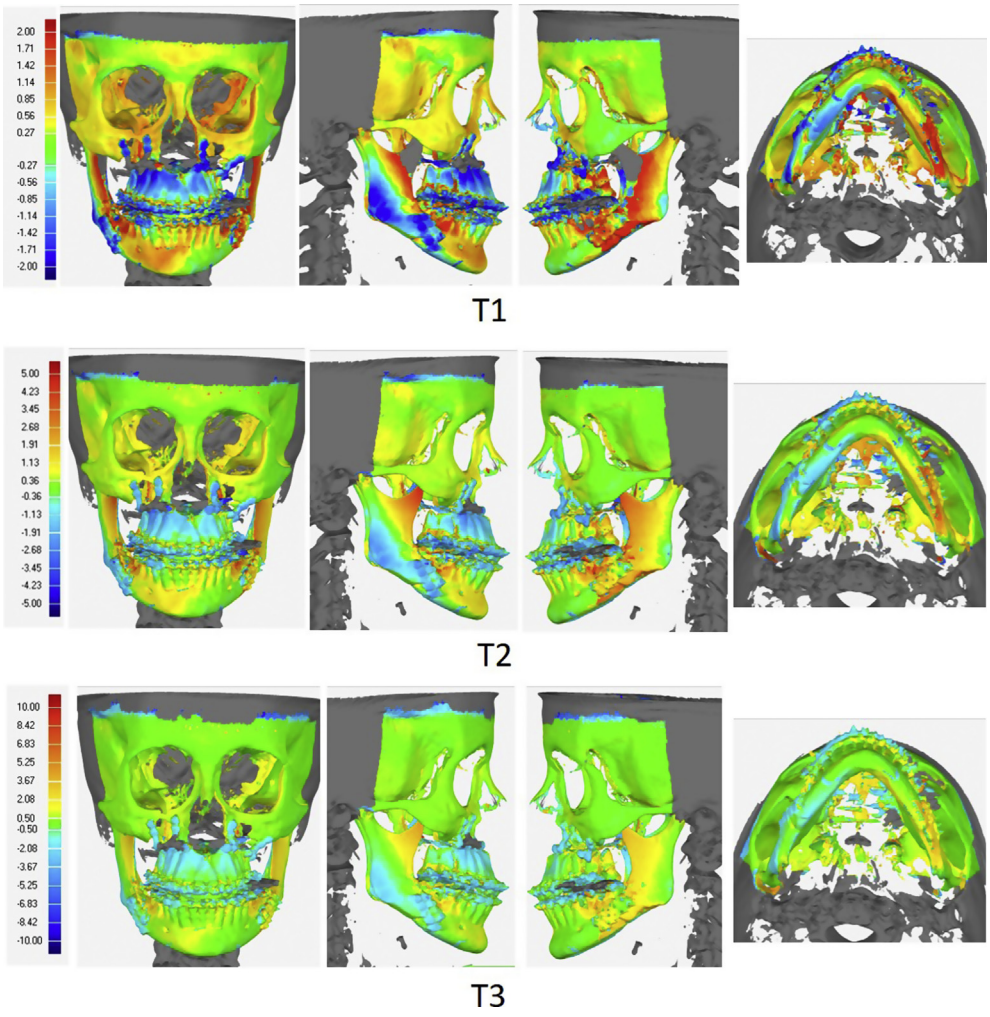

Fig. 4. Colour maps generated at different calibrations (T1, T2 and T3).

the colours between T1, T2 and T3, there was a resemblance among deviations when they were verified as values in the colour-coded scales (Fig. 4).

Table 1 shows the Pcc and paired *t*-test results, with all values for Pcc being higher than 94% ($r \geq 0.94$). The paired *t*-test did not reject the null hypothesis of similarity (H_0: $\mu_0 = \mu$) between means of registration errors and deviations when comparing the two evaluators. All results were not statistically significant ($p > 0.05$). These results confirmed the excellent reproducibility of the intra-rater and inter-rater reliability methods for deviation measurements.

In Table 1, one can also see the means of RMS error (1.18 mm) and different 3D Error(s) according to T1 (0.88 mm), T2 (1.27 mm) and T3 (1.42 mm). When comparing the mean of 3D error generated at T1 to RMS error, it was observed that the values of 3D error were lower than RMS error. The scatter plots at T1 showed this behaviour for which the values of RMS error were higher than 3D error for both evaluators. For T2 and T3, the scatter plots showed congruence of results between the two evaluators, there being most values of RMS error lower than 3D error (Fig. 5).

The 3D surface meshes in STL-format files comprised thousands of triangle's vertices in which SM0 and SM1 presented means and standard deviations of 773528 (±447652) and 1111270 (±429091), respectively. When the ICP algorithm was used to align semi-automatically SM0 and SM1 (surface-based registration), it resulted in ML values with means and standard deviations (Table 1). The results of ML from the ICP algorithm between SM0 and SM1 showed high intra- and inter-rater agreements (r = 0.98 and r = 0.99, respectively). In linear regression models by which the results were obtained, *t*-test for coefficients and standard error were statistically significant in SM0 for both evaluators, but not in SM1, regardless of the results being statistically significant for *F*-test (Table 2). When SM1 was removed from the linear regression analysis, all results were statistically significant (Table 2). It could be suggested that ML values might be more conditioned by SM0 than by SM1.

Table 3 shows the difference in the percentage between MD and 3D error at T1 and T3. The results indicated that deviations were more underestimated at T1 than at T2, being overestimated or even exceeded at T3. Both results may highlight the difference between T3 and T1, which was higher than 30%.

To confirm that, all the mean values met the criterion of clinical success (Stokbro et al., 2014). Table 4 shows the null hypothesis which was rejected ($p < 0.05$), with *t*-test results showing that all mean values of the deviation variables were higher than - 2 mm and lower than 2 mm, regardless of T1, T2 or T3.

4. Discussion

This study has evaluated the feasibility of ICP algorithm by using 3D surface meshes generated from CBCT DICOM data to validate an accuracy assessment method between VSP and surgical outcomes

Table 1
Pearson correlation coefficients (Pcc − r) and Paired *t*-test intra-evaluator and inter-evaluators to check the reproducibility.

Variables		Mean (standard deviation)		Pcc (r) intra and inter Eva(s)			Paired *t*-test					
							EVA 1		EVA 2		Inter[a]	
		EVA 1	EVA 2	EVA 1 (*r*)	EVA 2 (*r*)	Inter[a] (*r*)	*p*	H_o	*p*	H_o	*p*	H_o
Registration	Mean Error	0.82 (±0.26)	0.82 (±0.27)	0.99	0.99	0.99	0.35	NR	0.81	NR	0.66	NR
	RMS Error	1.18 (±0.43)	1.18 (±0.43)	0.99	0.99	0.99	0.93	NR	0.84	NR	0.06	NR
	ML	204.78 (±12.61)	204.88 (±12.34)	0.98	0.98	0.99	0.05	NR	0.08	NR	0.67	NR
T1	MD	0.05 (±0.16)	0.05 (±0.16)	0.99	0.99	0.99	0.28	NR	0.26	NR	0.33	NR
	MD+	0.72 (±0.11)	0.72 (±0.11)	0.98	0.97	0.99	0.66	NR	0.52	NR	0.64	NR
	MD-	−0.72 (±0.08)	−0.72 (±0.07)	0.94	0.95	0.99	0.71	NR	0.64	NR	0.77	NR
	SD	0.87 (±0.08)	0.87 (±0.07)	0.96	0.95	0.99	0.61	NR	0.53	NR	0.17	NR
	3D Error	0.88 (±0.09)	0.88 (±0.08)	0.97	0.97	0.99	0.43	NR	0.61	NR	0.99	NR
T2	MD	0.08 (±0.23)	0.08 (±0.22)	0.99	0.99	0.99	0.49	NR	0.88	NR	0.58	NR
	MD+	0.96 (±0.19)	0.96 (±0.2)	0.98	0.97	0.99	0.49	NR	0.53	NR	0.99	NR
	MD-	−0.94 (±0.13)	−0.94 (±0.13)	0.97	0.97	0.99	0.49	NR	0.72	NR	0.74	NR
	SD	1.25 (±0.16)	1.25 (±0.16)	0.98	0.98	0.99	0.52	NR	0.52	NR	0.29	NR
	3D Error	1.27 (±0.18)	1.27 (±0.18)	0.98	0.98	0.99	0.49	NR	0.50	NR	0.21	NR
T3	MD	0.09 (±0,24)	0.09 (±0.24)	0.99	0.99	0.99	0.08	NR	0.66	NR	0.17	NR
	MD+	1.02 (±0,24)	1.02 (±0.25)	0.99	0.99	0.99	0.46	NR	0.66	NR	0.75	NR
	MD-	−1 (±0,19)	−1 (±0.19)	0.99	0.99	0.99	0.92	NR	0.57	NR	0.77	NR
	SD	1.4 (±0.28)	1.4 (±0.28)	0.99	0.99	0.99	0.90	NR	0.79	NR	0.87	NR
	3D Error	1.42 (±0.29)	1.42 (±0.3)	0.99	0.99	0.99	0.69	NR	0.78	NR	0.86	NR

[a] Inter Evaluater; NR: Not rejected; Statistical significance ($p < 0.05$).

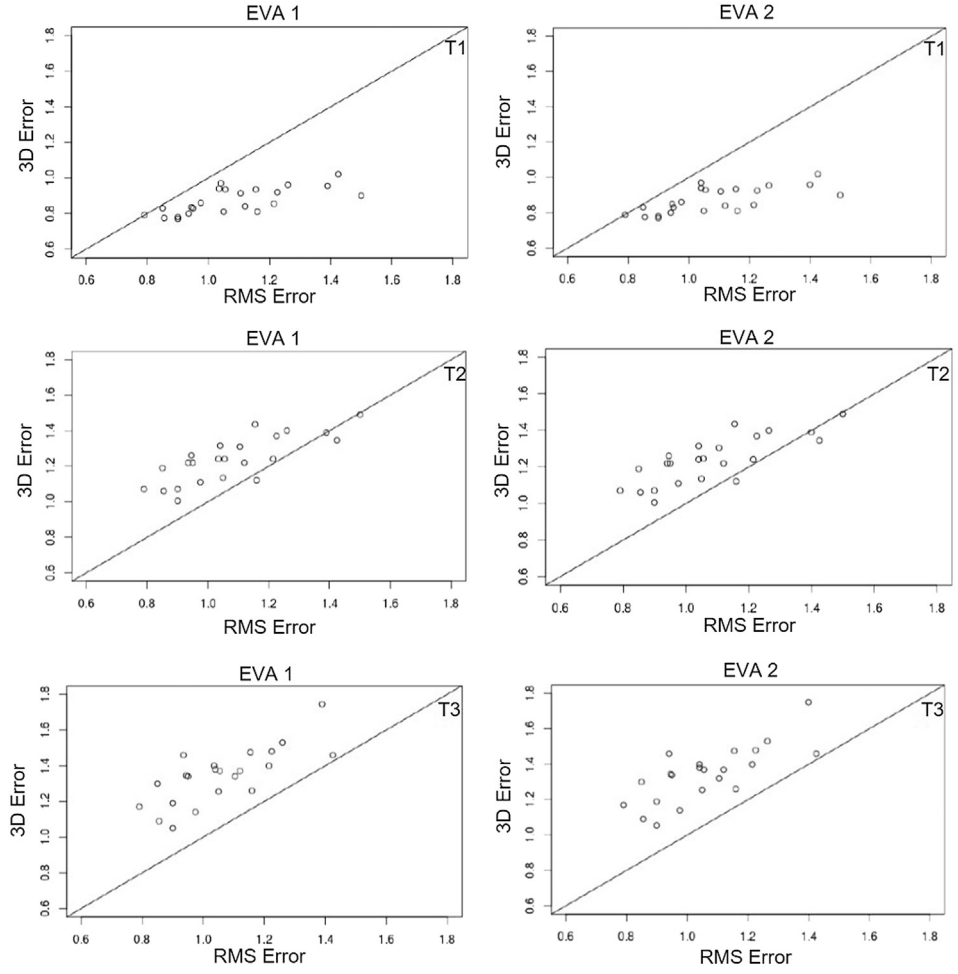

Fig. 5. Illustration showing scatterplot comparing RMS Error and 3D Error according to T1, T2 and T3.

from 25 pairs of surface meshes from patients treated with orthognathic surgery. Accordingly, the method was based on ICP algorithm for alignment and registration of SM0 to the corresponding SM1, including deviation analysis. The ICP algorithm has become the standard method for rigid registration between 3D models or surface meshes for biometric applications, such as

Table 2
Linear regression analysis estimated if SM0 and SM1 conditioned values of Maximum length in surface-based registration.

Variables		Coefficient	Standard Error	*t*-Test	*p*	r^2[b]	SIC[b]	*F* test (*p*)
EVA 1	ML	186.87	8.32	22.44	$<2.2 \times 10^{-16}$	0.22	202.98	0.06
	SM0	1.26×10^{-5}	5.34×10^{-6}	2.36	0.03			
	SM1	7.34×10^{-6}	5.57×10^{-6}	1.31	0.20			
EVA 2	ML	186.4	7.95	23.43	$<2.2 \times 10^{-16}$	0.26	200.7	0.03
	SM0	1.33×10^{-5}	5.1×10^{-6}	2.62	0.01			
	SM1	7.31×10^{-6}	5.32×10^{-6}	1.37	0.18			
Inter[a]	ML	186.63	8.13	22.95	$<2.2 \times 10^{-16}$	0.24	201.81	0.04
	SM0	1.29×10^{-5}	5.21×10^{-6}	2.49	0.02			
	SM1	7.33×10^{-6}	5.44×10^{-6}	1.34	0.19			
EVA 1	ML	195.93	4.76	41.16	$<2.2 \times 10^{-16}$	0.16	201.66	0.04
	SM0	1.14×10^{-5}	5.35×10^{-6}	2.13	0.04			
EVA 2	ML	195.68	4.56	42.83	$<2.2 \times 10^{-16}$	0.19	199.54	0.02
	SM0	1.22×10^{-5}	5.13×10^{-6}	2.38	0.02			
Inter[a]	ML	195.68	4.65	42.01	$<2.2 \times 10^{-16}$	0.18	200.57	0.03
	SM0	1.18×10^{-5}	5.23×10^{-6}	2.25	0.03			

***SIC: Schwarz information criterion.
[a] Inter: Inter-evaluators.
[b] r^2: Determination Coefficient.

Table 3
Absolut and percental difference between T1, T2 and T3 for means of MD and 3D Error.

Variables			MD (%)	3D Error (%)
Evaluation 1	EVA 1	T1 *vs.* T2	−0.0340 (−39.17)	−0.3848 (−30.36)
		T3 *vs.* T2	0.0084 (9.68)	0.1576 (12.47)
	EVA 2	T1 *vs.* T2	−0.0352 (−41.12)	0.3852 (−30.38)
		T3 *vs.* T2	0.0096 (11.21)	0.1580 (12,47)
Evaluation 2	EVA 1	T1 *vs.* T2	−0.03752 (−43.26)	−0,3864 (−30.38)
		T3 *vs.* T2	0.0104 (12.09)	0,1540 (12.46)
	EVA 2	T1 *vs.* T2	−0.0376 (−43.72)	−0.3876 (−30.46)
		T3 *vs.* T2	0.0096 (11.16)	0.1524 (11,94)

medical diagnostic support tools and computer-aided interventions (Mora et al., 2016). In other words, rigid registration can be described as a mathematical term involving all points or vertices of triangles within surface meshes and whose relationships are kept constant. These iterative points allowed the resulting surface mesh to be translated while its orientation and location are maintained based on the reference surface mesh, with single numerical values being considered the Euclidean distance between points (Jabar et al., 2015).

Previous studies applied the ICP algorithm to surface meshes of soft tissues of maxillofacial regions to obtain surface-based registration and colour-coded maps distances, thus allowing analysis of facial morphological changes (Cheung et al., 2016), variations of facial features in several populations (Kau et al., 2010) and facial changes following surgical procedures (Verzé et al., 2014) as well as assessment of the accuracy of 3D prediction of soft tissue changes following orthognathic surgery (Resnick et al., 2016; Mundluru et al., 2017). However, their results could not be compared to outcomes as in the present study.

Studies using surface meshes of hard tissues and similar methods could allow better comparisons, such as reliability and reproducibility of measurements, by means of Pearson's correlation or intra-class coefficient. In this sense, Almukthar et al. (2014) performed a comparison between voxel-based and surface-based registrations for which both methods used the ICP algorithm for superimposition of 3D surfaces models of the hard tissues. These methods showed a strong positive Pearson's correlation ($r \geq 0.88$). Jabar et al. (2015) used surface models of the hard tissue to check the validity assessment of 3D maxillary and mandibular changes following orthognathic surgery. The reproducibility of this method was estimated by a high-reliability coefficient ($r = 0.99$). Koerich et al. (2016) assessed the accuracy and reproducibility of the voxel-based superimposition method of CBCT by using ICP algorithm and colour maps for which RMS measurements between 3D surface models were computed. Their results of intra-class correlation coefficient were higher than 0.98, indicating reproducibility of the method. Baan et al. (2016) also performed surface-based registration of the 3D maxillary and mandibular models from pre-operative, planned and post-operative points in orthognathic surgery. This study evaluated the reproducibility of a semi-automatic approach to quantify the accuracy of the surgical outcome performed by two independent observers at multiple times. For validation of the method, reliability coefficient was higher than 0.88. In the present study, all values of Pearson's coefficient showed high intra- and inter-rater correlations for our evaluators ($r \geq 0.94$). Therefore, we believe that there was reproducibility of the results from the ICP algorithm when alignment tools for registration and deviation analysis were applied to surface meshes of the hard tissue following the same workflow.

In this study, the surface-based registration involved the ICP algorithm applied to reverse engineering software, which was performed to overlap the reference and resulting surface meshes and then the iterative measurements of the deviations between the triangle's vertices of SM1 and corresponding SM0. The software registration technique using the ICP algorithm has produced an important integration between the engineering software and dentistry areas, thus being mainly important for several computer-assisted treatments using 3D. As reported in other studies, the ICP algorithm of Geomagic software has already been used to evaluate the reproducibility of the lip at rest position by using stereophotogrammetric images (Dindaroglu et al., 2016), to compare the accuracy of printed models from intra-oral scans (Carmadella et al., 2017), and to quantify the accuracy of 3D facial models

Table 4
t-Test analysis to assess all means of registration error and deviations measures rejecting or not rejecting the hypotheses.

Variables		*t*-Test (H_o: $\mu \geq 2$ mm)			*t*-Test (H_o: $\mu \geq 2$ mm)		
Eva 1		*t*	*p*	H_o	*t*	*p*	H_o
Registration Error	Mean Error	−21.88	$<2.2 \times 10^{-16}$	R	−22.31	$<2.2 \times 10^{-16}$	R
	RMS Error	−9.2	1.21×10^{-9}	R	−9.41	7.87×10^{-10}	R
T1	MD	−59.25	$<2.2 \times 10^{-16}$	R	−61.07	$<2.2 \times 10^{-16}$	R
	MD+	−59.8	$<2.2 \times 10^{-16}$	R	−54.94	$<2.2 \times 10^{-16}$	R
	SD	−81.4	$<2.2 \times 10^{-16}$	R	−69.94	$<2.2 \times 10^{-16}$	R
	3D Error	−69.29	$<2.2 \times 10^{-16}$	R	−61.65	$<2.2 \times 10^{-16}$	R
T2	MD	−41.99	$<2.2 \times 10^{-16}$	R	−41.7	$<2.2 \times 10^{-16}$	R
	MD+	−28.53	$<22 \times 10^{-16}$	R	−23.91	$<2.2 \times 10^{-16}$	R
	SD	−25.04	$<2.2 \times 10^{-16}$	R	−21.32	$<2.2 \times 10^{-16}$	R
	3D Error	−21.81	$<2.2 \times 10^{-16}$	R	−19	2.83×10^{-15}	R
T3	MD	−38.96	$<2.2 \times 10^{-16}$	R	−38.80	$<2.2 \times 10^{-16}$	R
	MD+	−20.44	$<2.2 \times 10^{-16}$	R	−17.76	$<2.2 \times 10^{-16}$	R
	SD	−10.42	1.07×10^{-10}	R	−10.03	2.28×10^{-10}	R
	3D Error	−9.59	5.45×10^{-10}	R	−9.26	1.06×10^{-9}	R
Eva 2							
Registration Error	Mean Error	−21.88	$<2.2 \times 10^{-16}$	R	−22.31	$<2.2 \times 10^{-16}$	R
	RMS Error	−9.2	1.21×10^{-9}	R	−9.39	8.28×10^{-10}	R
T1	MD	−60.8	$<2.2 \times 10^{-16}$	R	−61.28	$<2.2 \times 10^{-16}$	R
	MD+	−62.59	$<2.2 \times 10^{-16}$	R	−55.13	$<2.2 \times 10^{-16}$	R
	SD	−82.06	$<2.2 \times 10^{-16}$	R	−69.94	$<2.2 \times 10^{-16}$	R
	3D Error	−71.06	$<2.2 \times 10^{-16}$	R	−61.58	$<2.2 \times 10^{-16}$	R
T2	MD	−42.96	$<2.2 \times 10^{-16}$	R	−41.96	$<2.2 \times 10^{-16}$	R
	MD+	−28.56	$<2.2 \times 10^{-16}$	R	−23.96	$<2.2 \times 10^{-16}$	R
	SD	−24.36	$<2.2 \times 10^{-16}$	R	−21.36	$<2.2 \times 10^{-16}$	R
	3D Error	−21.53	$<2.2 \times 10^{-16}$	R	−18.84	3.42×10^{-16}	R
T3	MD	−38.96	$<2.2 \times 10^{-16}$	R	−38.77	$<2.2 \times 10^{-16}$	R
	MD+	−19.83	$<2.2 \times 10^{-16}$	R	−17.66	1.47×10^{-15}	R
	SD	−10.32	1.3×10^{-10}	R	−10.03	2.3×10^{-10}	R
	3D Error	−9.48	6.87×10^{-10}	R	−9.21	1.17×10^{-10}	R
Variables		*t*-Test (H_o: $\mu \leq -2$ mm)			*t*-Test (H_o: $\mu \leq -2$ mm)		
Eva 1		*t*	*p*	H_o	*t*	*p*	H_o
T1	MD	62,46	$<2,2 \times 10^{-16}$	R	64,12	$<2,2 \times 10^{-16}$	R
	MD-	90,08	$<2,2 \times 10^{-16}$	R	79,79	$<2,2 \times 10^{-16}$	R
T2	MD	45,80	$<2,2 \times 10^{-16}$	R	45,45	$<2,2 \times 10^{-16}$	R
	MD-	41,53	$<2,2 \times 10^{-16}$	R	37,02	$<2,2 \times 10^{-16}$	R
T3	MD	42,85	$<2,2 \times 10^{-16}$	R	42,734	$<2,2 \times 10^{-16}$	R
	MD-	25,58	$<2,2 \times 10^{-16}$	R	24,3	$<2,2 \times 10^{-16}$	R
Eva 2							
T1	MD	63.94	$<2.2 \times 10^{-16}$	R	64.32	$<2.2 \times 10^{-16}$	R
	MD-	88.2	$<2.2 \times 10^{-16}$	R	78.39	$<2.2 \times 10^{-16}$	R
T2	MD	46.8	$<2.2 \times 10^{-16}$	R	45.44	$<2.2 \times 10^{-16}$	R
	MD-	41.43	$<2.2 \times 10^{-16}$	R	36.61	$<2.2 \times 10^{-16}$	R
T3	MD	42.85	$<2.2 \times 10^{-16}$	R	42.66	$<2.2 \times 10^{-16}$	R
	MD-	26.94	$<2.2 \times 10^{-16}$	R	24.13	$<2.2 \times 10^{-16}$	R

R: Rejected.

acquired via different optical scans of dentofacial deformities (Zhao et al., 2017).

Therefore, a wide set of applications using the ICP algorithm for different fields allow registration errors to be computed automatically by software. In the present study, the ICP algorithm was required for a semi-automatic approach when the evaluators initialised the algorithm and selected parameter settings. The registration error means were 0.82 mm (mean error) and 1.18 mm (RMS error). As a comparison with our study, Zhou et al. (2016) aimed to evaluate both feasibility and accuracy of the ICP algorithm applied to registration and fusion of soft tissue between 3D photorealistic surface images and multislice spiral computed tomography (CT) scans. The authors highlighted that registration errors were lower than 0.8 mm and ICP algorithm was an optimal registration technique for enabling accurate fusion between 3D facial images and CT re-constructed data. In addition, Khambay and Ullah (2015) compared the predicted 3D soft tissue image to post-operative outcomes. Regardless of the results and limitations of the current methods of analysis, the alignment between 3D surface meshes of the soft tissues had been obtained from rigid registration of pre- and post-operative outcomes regarding the hard tissues, whose images were superimposed over the anterior skull or other stable structures. Although the results showed that VSP software has the predictive ability of 3D soft tissue analysis, the authors emphasised that hard tissues were used as template for alignment because the soft tissue would also be aligned as there is a close relationship between hard and soft tissues.

Our method was conducted with a heterogeneous set of surface meshes consisting of different numbers of triangle's vertices, and, even so, the software presented behaviour of alignment and deviation analysis, which automatically considered as references the skull base and frontal, orbital and zygomatic regions. Furthermore, the paired *t*-test showed that the difference between measurements was not statistically significant and did not reject the hypothesis of similarity between intra- and inter-rater agreements, suggesting that the specified deviations could be accurately measured by both evaluators.

We have found no previous study presenting ML values after surface-based registration or semi-automatic alignment between surface meshes. ML could suggest the way SM0 and SM1 behave themselves, mainly regarding heterogeneous surface meshes of the hard tissues (different number of triangle vertices). From regression linear analysis, it was possible to estimate which ML values were conditioned to a given number of triangles vertices in SM0 ($p \leq 0.04$). These results may be due to the SM0 being defined as a reference when applied to the workflow. These estimation results could suggest evidence of which streak artefacts in SM1 (caused by dental metals, orthodontics appliances or rigid internal fixation) could be minimised during iterative workflow in the surface-based registration and deviation analysis, thus not interposing on the ICP algorithm. Nevertheless, erroneous data on SM1 were enveloped by ICP algorithm because streak artefacts may be viewed in colour maps, and thus it could have a marked effect on measurements or deviations (Jabar et al., 2015).

As shown in the results listed in Table 1, the means of MD were lower than those of 3D error. This bias in the results had never been reported in other studies (Khambay and Ullah, 2015; Jayaratne et al., 2012) and was not considered as a validly signed average of the deviation because positive values would cancel out any negative values underestimating the mean deviation. In an attempt to overcome this bias, our method used the 3D error (RMS) provided by the deviation analysis. Zhao et al. (2017) also believed that 3D error might be considered as a measurement indicator of shape congruence to quantify the accuracy between VSP and actual outcomes in orthognathic surgery. This method was more comprehensive than those used by previous studies, which were based on calculations of linear and angular differences between cephalometric landmarks (Zinser et al., 2012; De Riu et al., 2014) or intraclass coefficients of reference points and reference angles (Hernández-Alfaro and Guijarro-Martínez, 2013). As for the larger SD at T1, T2 and T3, the ICP algorithm demonstrated variability in the deviation analysis. According to Almukthar et al., 2014, the surface-based registration also demonstrated variability in the superimposition of surface meshes of the hard tissue. The authors suggested that variability in superimposition may be due to application of ICP algorithm to features of heterogeneous surfaces for registration.

Jabar et al. (2015) highlighted that ICP algorithm provided numerical values (i.e. mean distance and RMS) which were Euclidean distances between the points of surface meshes. The authors emphasised that a drawback of this method should be taken into account when trying to assess 3D hard tissue changes (i.e. between pre- and post-operative surface meshes of plastic skull 3D models) because their results showed an underestimation of the distances of simulated surgical movement by about 50–70%. In this study, the results listed in Table 3 indicated that the values of MD and 3D error at T2 were underestimated by 30% when analysis deviation was calibrated at T1 and overestimated by 12% when calibrated at T3. Therefore, we could suggest that this software limitation was minimised because the ICP algorithm had been used in similar surface meshes. However, this bias on the results may affect the feasibility of the method and the validity of actual deviations.

As shown in Table 4, the results indicated that all values were lower than 2 mm and higher than −2 mm, with 3D error only showing antero-posterior direction and magnitude of the deviations, which did not fully describe any direction of the deviations. There are limitations regarding the ICP algorithm of the software in the assessment of the accuracy between surface meshes of the hard tissue (Baan et al., 2016; Jabar et al., 2015; Almukthar et al., 2015), which biased the results. Firstly, the shortcoming of this approach lies in the fact that measurements or deviations were between the two nearest points of the two surface meshes (i.e. shortest deviations between triangle vertices of the adjacent meshes), with no actual correspondence (Cheung et al., 2016) or not corresponding to the same anatomical points (Jabar et al., 2015; Almukthar et al., 2015). Secondly, the deviation analysis did not stratify the deviations by using the three Cartesian frames of reference (x, y and z), nor indicated all quantitatively potential errors which needed to be addressed to compare the VSP to the post-operative outcomes without answering how to best eliminate such errors.

The method has partially satisfied the criteria suggested in the systematic review protocol for 3D accuracy evaluation of VSP in orthognathic surgery (Gaber et al., 2017), which consisted in reducing the likelihood of human error as the evaluators had chosen the software's parameters (i.e. semi-automatically surface-based registration and deviation analysis) and the results were considered reproducible by intra and inter-rater agreements. We believe that colour maps were generated easily and could be able to provide 3D comparisons between planning and actual results. However, the method has not become feasible yet because there were methodological limitation and drawbacks. We look forward to the method described being used to guide other studies in an attempt to eliminate the bias of results when being applied in clinical practices by oral and maxillofacial surgeons.

5. Conclusion

ICP algorithm and colour maps were reproducible on Geomagic software. The method was able to show a deviation value (3D error), but it did not stratify the deviation axes (x, y and z coordinates), which could show further inaccuracy and sources of errors.

The method presented a limited feasibility due to either underestimation or overestimation of the results, thus affecting the validity of the actual deviations. Therefore, colour maps could not be used for 3D comparisons between treatment planning and post-operative outcomes in orthognathic surgery.

Ethical approval

Anonymous data from cone beam computed tomography of patients that had undergone orthognathic surgery were analyzed after approval by humans' research Committee in Brazil (ethical standards applicable in 1964 Helsinki Declaration). The data were analyzed without personal identifying details. Ethical approval was supported by the Ethics committee of the Federal University of Juiz de Fora, Minas Gerais, Brazil (reference number: CAAE 695.98017.2.0000.5133; electronic address: http://plataformabrasil.saude.gov.br/login.jsf).

Funding

There was no funding for this study.

Conflicts of interest

The authors declare that they have no conflict of interest.

References

Almukthar A, Xiangyang J, Khambay B, McDonald J, Ayoub A: Comparison of the accuracy of voxel-based registration for 3D assessment of surgical change following orthognathic surgery. PLoS One 9: e93402, 2014

Almukthar A, Khambay B, Ayoub A, Ju X, Al-Hiyali A, Macdonald J, et al: "Direct Dicom Slice Landmarking" A novel research technique to quantify skeletal changes in orthognathic surgery. PLos One 10: e0131540, 2015

Baan F, Liebreqts J, Xi T, Schreurs R, de Koning M, Bergé S, et al: A new tool for assessing the accuracy of bimaxillary surgery: the OrthoGnathicAnalyser. PLoS One 11: e0149625, 2016

Borba AM, Haupt D, de Almeida Romualdo LT, da Silva AL, da Graça Naclério-Homem M, Miloro M: How many oral and maxillofacial surgeons does it take to

perform virtual orthognathic surgical planning? J Oral Maxillofac Surg 74: 1807–1826, 2016
Carmadella LT, Vilella OV, Breuning H: Accuracy of printed dental models made with 2 prototyp technologies and different designs of models bases. Am J Orthod Dentofacial Orthop 151: 1178–1187, 2017
Cevidanes LH, Styner MA, Proffit WR: Image analysis and superimposition of 3-dimensional cone-beam computed tomography models. Am J Orthop Dentofacial Orthop 129: 611–618, 2006
Cheung MY, Almukhtar A, Keeling A, Hsung TC, Ju X, McDonald J, et al: The accuracy of conformation of a generic surface mesh for the analysis of facial soft tissue changes. PLoS One 11: e0152381, 2016
De Riu G, Meloni SM, Baj A, Corda A, Soma D, Tullio A: Computer-assisted orthognathic surgery for correction of facial asymmetry: results of a randomized controlled clinical trial. Br J Oral Maxillofac Surg 52: 251–257, 2014
Dindaroglu F, Duran GS, Görgülü S: Reproducibility of the lip at rest: a 3-dimensional perspective. Am J Orthod Dentofacial Orthop 149: 757–765, 2016
Gaber RM, Shareen E, Falter B, Araya S, Politis C, Swennen GRJ, et al: A systematic review to uncover a universal protocol for accuracy assessment of 3-dimensional virtually planned orthognathic surgery. J Oral Maxillofac Surg 75: 2430–2440, 2017
Haas Jr OL, Becker OE, de Oliveira RB: Computer-aided in orthognathic surgery – systematic review. Int J Oral Maxillofac Surg 44: 329–342, 2015
Hernández-Alfaro F, Guijarro-Martínez R: New protocol for three-dimensional surgical planning and CAD/CAM splint generation in orthognathic surgery: an in vitro and in vivo study. Int J Oral Maxillofac Surg 42: 1547–1556, 2013
Jabar N, Robinson W, Goto TK, Kambay BS: The validity of using surface meshes for evaluation of three-dimensional maxillary surgical changes. Int J Oral Maxillofac Surg 44: 914–920, 2015
Jayaratne YSN, McGrath C, Zwahlen RA: How accurate are the fusion of cone-beam CT ad 3D Stereophotographic images? PLoS One 7: e49585, 2012
Kau CH, Richmond S, Zhurou A, Ovsenik M, Tawfik W, Borbely P, et al: Use of three-dimensional surface acquisition to study facial morphology in 5 population. Am J Orthod Dentofacial Orthop 137(Suppl. 4): e1–e9, 2010
Khambay B, Ullah R: Current methods of assessing the accuracy of three-dimensional soft tissue facial predictions: technical and clinical considerations. Int J Oral Maxillofac Surg 44: 132–138, 2015
Koerich L, Burns D, Weissheimer A, Claus JDP: Three-dimensional maxillary and mandibular regional superimposition using cone beam computed tomography: a validation study. Int J Oral Maxillofac Surg 45: 662–669, 2016
Maal TJJ, de Koning MJ, Plooij JM, Verhamme LM, Rangel F, Bergé SJ, et al: One year postoperative hard and soft tissue volumetric changes after a BSSO mandibular advancement. Int J Oral Maxillofac Surg 41: 1137–1145, 2012
Mora H, Mora-Pascual JM, García-Gárcia A, Martínez-González: Computacional analysis of distance operators for the iterative closest point algorithm. PLoS One 11: e0164694, 2016
Mundluru T, Almukhtar A, Ju X, Ayoub A: The accuracy of three-dimensional prediction of soft tissue changes following the surgical correction of facial asymmetry: an innovative concept. Int J Oral Maxillofac Surg 46: 1517–1524, 2017
Resnick CM, Dang RR, Glick SJ, Padwa BL: Accuracy of three-dimensional soft tissue prediction for Le Fort I osteotomy using Dolphin 3D software: a pilot study. Int J Oral Maxillofac Surg 46: 289–295, 2016
Ritto FG, Schmitt ARM, Pimentel T, Canellas JV, Medeiros PJ: Comparison of the accuracy of maxillary position between conventional model surgery and virtual surgical planning. Int Oral Maxillofac Surg 47: 160–166, 2017
Stokbro K, Aagaard E, Torkov P, Bell RB, Thygesen: Virtual planning in orthognathic surgery. Int J Oral Maxillofac Surg 43(8): 957–965. https://doi.org/10.1016/j.ijom.2014.03.011, 2014 Aug
Terajima M, Yanagita N, Ozeki K, Hoshino Y, Mori M, Goto TK, et al: Three-dimensional analysis system for orthognathic surgery patients with jaw deformities. Am J Orthod Dentofacial Orthop 134: 100–111, 2008
Tucker S, Cevidanes LHS, Styner M, Kim H, Reyes M, Proffit W, et al: Comparison of actual surgical outcomes and 3-dimensional surgical simulations. J Oral Maxillofac Surg 68: 2412–2421, 2010
Verzé L, Bianchi FA, Schellino E, Ranieri G: Soft tissue changes after orthodontic surgical correction of jaws asymmetry evaluated by 3D surface laser scanner. J Craniofac Surg 23: 1448–1452, 2014
Weissheimer A, Menezes LM, Koerich L, Pham J, Cevidanes LHS: Fast three-dimensional superimposition of cone beam computed tomography for orthopaedics and orthognathic surgery evaluation. Int J Oral Maxillofac Surg 44: 1188–1196, 2015
Zhao Y-J, Xiong Y-X, Wang Y: Three-dimensional accuracy of facial scan for facial deformities in clinics: a new evaluation method for facial scanner accuracy. PLoS One 12: e0169402, 2017
Zhou Z, Li P, Ren J, Guo J, Huang Y, Tian W, et al: Virtual facial reconstruction based on accurate registration and fusion of 3D facial and MSCT scans. J Orofac Orthop 77: 104–111, 2016
Zinser MJ, Mischkowski RA, Sailer HF, Zöller JE: Computer-assisted orthognathic surgery: feasibility study using multiple CAD/CAM surgical splints. Oral Surg Oral Med Oral Pathol Oral Radiol 113: 673–687, 2012