DENTAL MATERIALS 35 (2019) 495–500

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.intl.elsevierhealth.com/journals/dema

Investigation into the accuracy and measurement methods of sequential 3D dental scan alignment

Saoirse O'Toole[c,*], *Cecilie Osnes*[b,d], *David Bartlett*[a], *Andrew Keeling*[d]

[a] *Centre for Clinical Oral and Translational Sciences, Faculty of Dental, Oral and Craniofacial Sciences, King's College London Dental Institute, Tower Wing, Guy's Hospital, London SE1 9RT, UK*
[b] *Department of Medical Technologies, University of Siena, Siena, Italy*
[c] *Department of Prosthodontics, King's College London Dental Institute, Floor 25, Tower Wing, Guy's Hospital, London SE19RT, UK*
[d] *Department of Restorative Dentistry, Leeds School of Dentistry, Clarendon Way, Leeds LS2 9LU, UK*

ARTICLE INFO

Article history:
Received 16 May 2018
Received in revised form
3 October 2018
Accepted 11 January 2019

Keywords:
Tooth wear
Tooth erosion
Diagnostic imaging
Dental technology

ABSTRACT

Objectives. Alignment procedures have yet to be standardised and may influence the measurement outcome. This investigation assessed the accuracy of commonly used alignment techniques and their impact on measurement metrics.

Methods. Datasets of 10 natural molar teeth were created with a structured-light model-scanner (Rexcan DS2, Europac 3D, Crewe). A 300 μm depth layer was then digitally removed from the occlusal surface creating a defect of known size. The datasets were duplicated, randomly repositioned and re-alignment attempted using a "best-fit" alignment, landmark-based alignment or reference alignment in Geomagic Control (3D Systems, Darmstadt, Germany). The re-alignment accuracy was mathematically assessed using the mean angular and translation differences between the original alignment and the re-aligned datasets. The effect of the re-alignment on conventional measurement metrics was calculated by analysing differences between the known defect size and defect size after re-alignment. Data were analysed in SPSS v24(ANOVA, post hoc Games Howell test, $p < 0.05$).

Results. The mean translation error (SD) was 139 μm (42) using landmark alignment, 130 μm (26) for best-fit and 22 μm (9) for reference alignment ($p < 0.001$). The mean angular error (SD) between the datasets was 2.52 (1.18) degrees for landmark alignment, 0.56 (0.38) degrees for best-fit alignment and 0.26 (0.12) degrees for reference alignment ($p < 0.001$). Using a reference alignment statistically reduced the mean profilometric change, volume change and percentage of surface change errors ($p < 0.001$).

Significance. Reference alignment produced significantly lower alignment errors and truer measurements. Best-fit and landmark-based alignment algorithms significantly underestimated the size of the defect. Challenges remain in identifying reference surfaces in a robust, clinically relevant method.

Crown Copyright © 2019 Published by Elsevier Inc. on behalf of The Academy of Dental Materials. All rights reserved.

* *Corresponding author.*
E-mail address: Saoirse.otoole@kcl.ac.uk (S. O'Toole).
https://doi.org/10.1016/j.dental.2019.01.012
0109-5641/Crown Copyright © 2019 Published by Elsevier Inc. on behalf of The Academy of Dental Materials. All rights reserved.

1. Introduction

Digital 3D scanning, superimposition and comparison, has been used to quantify changes in orthodontics [1,2], periodontics [3,4] and tooth wear [5] measurements, with varying degrees of accuracy. When assessing the accuracy of the 3D comparison process, the majority of literature has focused on the reproducibility error of obtaining the 3D datasets either indirectly via sequential dental models [5–8] or directly using digital scanners. However, the superimposition or alignment of the two datasets is not trivial and is also prone to error [9]. The mathematical complexities of dataset alignment are often hidden from the operator to make software easier to use and may not be immediately obvious to the operator. Errors introduced at this crucial first stage in the digital workflow have rarely been acknowledged in the dental literature [1,10,11] and the complexities in these alignments are under-explored. Comparison is made difficult by the lack of standardisation of measurement metrics. The maximum profilometric change [12], mean profilometric loss [7,13], volume change [6,7] and percentage of surface change [8] have all been used as outcome measures. No single metric has been universally decided upon and it may be that errors in alignment affect measurement metrics differently.

Traditionally, three different types of scan alignment have been used landmark-based alignment, best-fit alignment and then a reference best-fit. A "landmark based alignment" is performed by the operator manually selecting common landmarks or common points on each dataset which are then aligned by the software. Landmark alignment is relatively straight-forward and widely used in medical applications where precision at micron level is not required. However, this method is highly subjective and dependent on the skill and comprehension of the alignment by the operator. In situations when the initial alignment guess is poor or a manual error in a landmark is made, the alignment process will only be partially complete [1].

A standard "best-fit alignment" uses an iterative closest point (ICP) algorithm to align scans, with each software using a slightly different algorithm and do not involve operator-based decisions. The alignment is performed by minimising the mesh distance error between each corresponding data point. By the very nature of the iterative algorithm's termination criteria, alignment will minimise mesh distance error and spread errors evenly over positive and negative deviations. If there is a large defect, the algorithm will attempt to minimise the absolute distance between the two datasets, regardless of the clinical outcome. This may explain erroneous results in tooth wear progression analysis where the tooth appears to have grown over the measurement period [6].

To circumvent this error, researchers have attempted to align on surface areas which have experienced change below a predefined threshold [8] and a recent systematic review on wear measurements has recognised this as a superior approach [14]. A "reference best-fit alignment" aligns datasets by restricting alignment to operator-identified sections of the dataset which are least likely to have undergone change [6,15,16]. This avoids the error of minimising the defect of interest to be measured but introduces an operator error when selecting sections of the dataset.

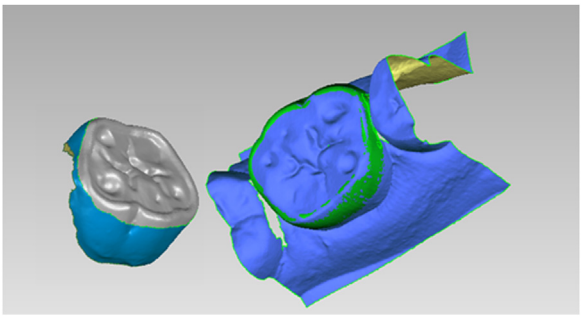

Fig. 1 – Image of the four surfaces used for analysis. The original surface (purple) is overlaid by the digitally eroded surface in green in perfect alignment allowing true quantification of the defect. The same surface was then duplicated and displaced to test re-alignment. For reference best-fit alignment, the area used for alignment is shown in blue. The area excluded from alignment in grey. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Quantification of the error of each alignment method has yet to be performed. We are also unaware of what impact an alignment error will have on the measurement outcome. The aim of this investigation was to assess the accuracy of each alignment technique and the subsequent influence on different measurement metrics. The following null hypotheses were formulated: There is no difference in the position or angulation of the dataset following alignment between the different techniques. Secondly, there will be no difference in the maximum profilometric gain, maximum profilometric loss, mean profilometric gain, mean profilometric loss, percentage of surface change and volume change measurements of a known defect between the different alignment techniques.

2. Method

Ten randomly chosen lower molar teeth were scanned in a dental model scanner (Rexcan DS2, Europac 3D, Crewe) with a stated accuracy of <10 μm. Duplicates of the datasets were created and an arbitrary defect was created by digitally removing a 300 μm layer from the occlusal surface using Meshlab [17], leaving a 1 mm intact perimeter. To repair the gaps created in the dataset during this process, a Meshlab Poisson Reconstruction Filter (subdivision level = 11) was used. The "eroded step" can be seen in the grey data set in Fig. 1. This produced two scans in perfect alignment, with the latter exhibiting a known defect which was quantified in Geomagic Control (version 2.0) with the following metrics: the maximum profilometric gain (μm) and maximum profilometric loss (μm) defined as the maximum difference (positive and negative) in the Z axis within the analysed area; the mean profilometric gain (μm) and mean profilometric loss (μm) defined as the average difference (positive and negative) in the Z axis within the analysed

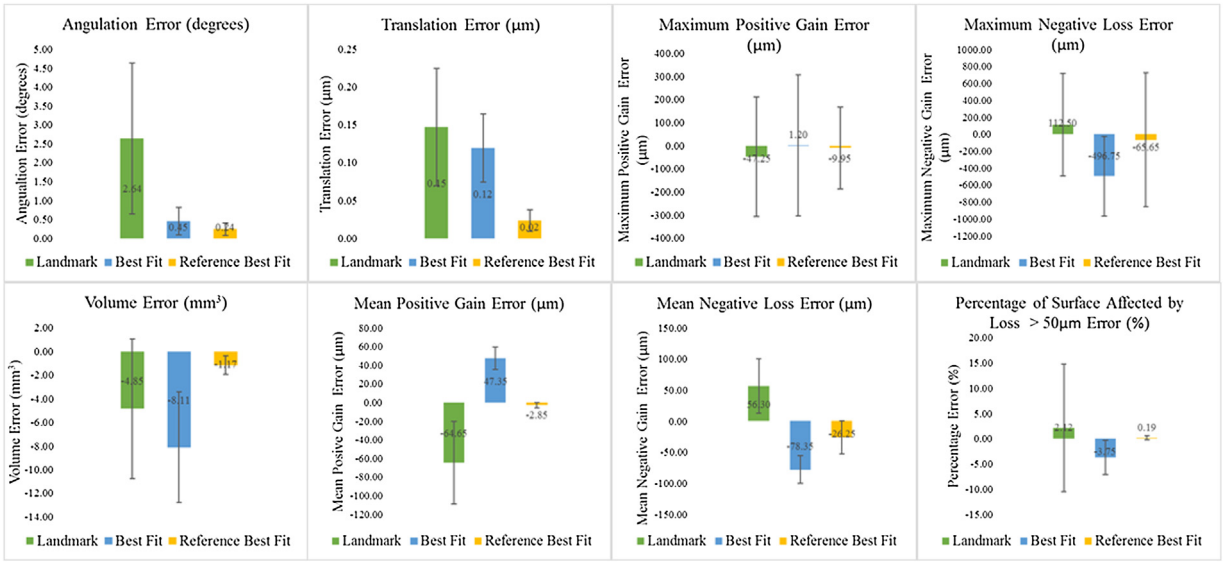

Fig. 2 – Figure showing errors introduced after the alignment. The angular error and translation errors (mathematical differences between the true alignment and re-alignment) are shown in the top right corner. The remaining graphs represent the resultant measurement errors (differences from known true value of the defect).

area; the percentage of the surface with profilometric loss >50 μm and the volume change of the defect (mm^3).

Datasets with the defect were randomly repositioned from their perfect alignment using custom software written in C++. Each dataset underwent a random rotation about an axis by an amount varying 0–360°, and a displacement along X, Y and Z between −10 and +10 mm.

Re-alignment with the original, unaltered dataset was then performed using one of three methods in Geomagic Control. For the landmark based alignment algorithm, ten convenient carefully chosen and corresponding landmarks on each dataset were selected by a single operator. The software then aligned the datasets by superimposing these landmarks. The best-fit algorithm was performed using the entire dataset by aligning 1000 randomly selected data points, which was then refined with an alignment on 5000 data points. For the reference alignment, the area of the defect and surrounding occlusal surface we wished to measure were manually selected by the operator and deleted from the dataset (visible in light blue in Fig. 1). The best-fit alignment process as described above was performed using this reduced dataset, which was assessed by the operator as having not experienced change. The transformation matrix was then applied to the complete displaced dataset to realign it with the same orientation.

Transformation matrices for all alignments were used to robustly calculate the deviation from the known perfect alignment. Custom software was written using Singular Value Decomposition in the Point Cloud Library (www.pointclouds.org) to calculate the absolute separation in microns of the geometric centre of the dataset (translation error) and the absolute difference in angulation/rotation in degrees (angular error) between the original dataset and re-aligned dataset. The impact of the alignment errors on measurement outcome was assessed by obtaining the mean profilometric gain, mean profilometric loss and volume change measurements using the realigned dataset and subtracting it from the known true defect size.

Data were analysed in SPSS version 24 and initially tested for normality using histograms, boxplots and Shapiro–Wilk's test. The variance between groups was statistically significant using Levene's test. The data were normally distributed therefore differences were assessed using a one-way ANOVA with post hoc Games Howell test for multiple comparisons.

3. Results

The mean translation error or separation between the original dataset and re-aligned data set was 139 μm (SD 42) using landmark alignment, 130 μm (SD 26) for best-fit and 22 μm (SD 9) for reference best-fit alignment. Reference best-fit had statistically significant reduced translation error compared to both landmark and best-fit alignment ($p < 0.001$).

The landmark alignment resulted in a mean angular error of 2.52° (SD 1.18), 0.56° (SD 0.38), for best-fit alignment and 0.26° (SD 0.12) for reference best-fit alignment. Reference best-fit alignment had reduced angular error compared to landmark alignment ($p < 0.001$) but not best-fit alignment ($p = 0.094$).

All mathematical and measurement errors are shown in Fig. 2. Negative results indicate an underestimation of the defect size while positive results indicate an overestimate of the defect size. The mean profilometric gain error was −60.4 μm (SD 23.2) for landmark alignment, −47.3 μm (SD 7.0) for standard best fit alignment and 1.2 (SD 5.3) μm for reference best fit alignment. These differences were statistically significant ($p < 0.001$). The mean profilometric loss error was 63.4 μm (SD 33.3) for landmark alignment, 81.2 μm (SD 14.8) for standard best-fit alignment and 26.4 μm (SD 14.7) using refer-

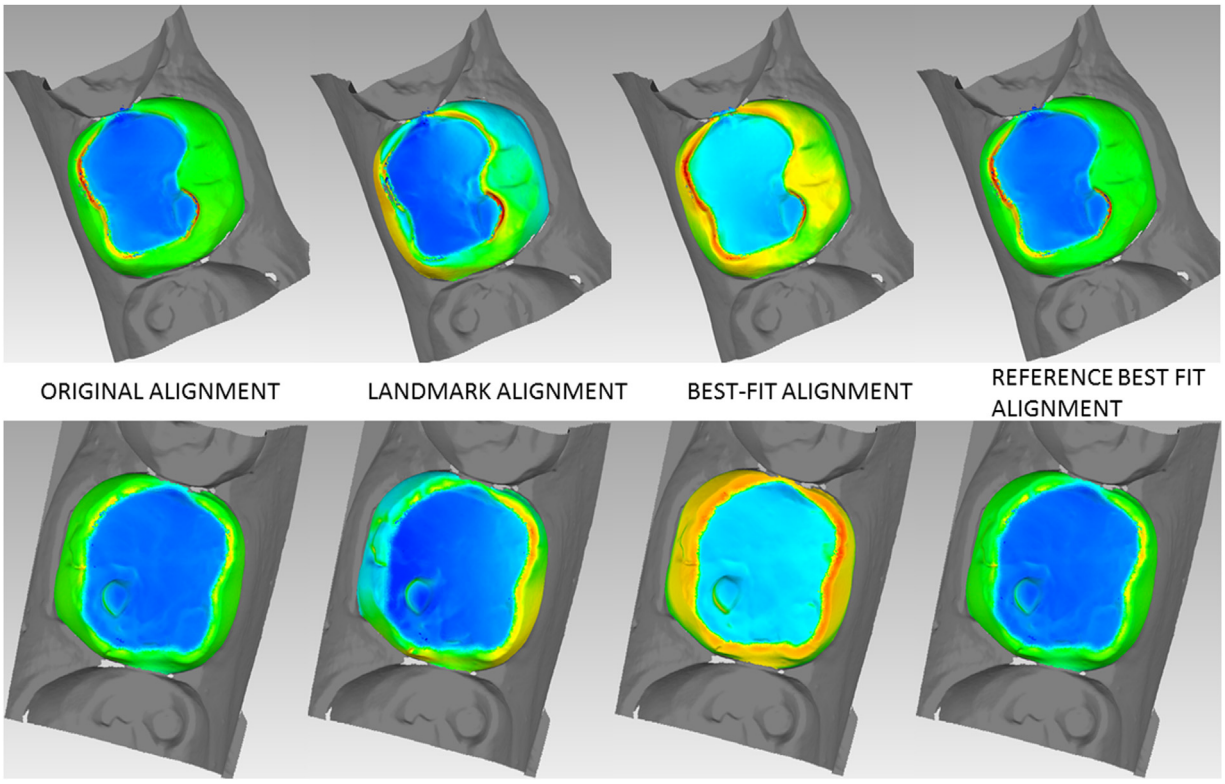

Fig. 3 – Representative images showing defects after alignment. Areas in green depict no change, deepening areas of blue indicate deepening areas of loss and deepening areas of red indicate areas of gain. (For interpretation of the references to colour mentioned in the text, the reader is referred to the web version of this article.)

ence best fit alignment. Statistical differences were observed between reference best fit and standard best fit alignment ($p < 0.001$) and between landmark alignment and standard best fit alignment ($p = 0.006$). The volume change error was $-6.2\,mm^3$ (SD 5.3) for landmark alignment, $-9.0\,mm^3$ (SD 2.9) for best fit alignment and $-1.1\,mm^3$ (SD 0.5) for reference best fit alignment. Reference best fit alignment had statistically significant reduced measurement error compared to both landmark alignments ($p = 0.028$) and reference best fit alignment ($p < 0.001$). There were also statistical differences between landmark alignment and best fit alignment ($p = 0.009$). Maximum profilometric errors were statistically similar for all methods of alignment with large standard deviations within groups.

The best-fit alignment resulted in the greatest underestimation of volume change and mean profilometric loss compared to reference best-fit alignment ($p < 0.001$) and landmark based alignment ($p < 0.05$). Best-fit alignment was the only alignment which resulted in overestimation of mean profilometric gain ($p < 0.001$).

Representative colour maps from two examples of the alignments are shown in Fig. 3. The scale ranges from +0.4 mm to −0.4 mm. The true defect colour maps (first column) show a clear negative deviation occlusally as expected (blue), with unchanged buccal and lingual tooth tissue (green). A slight peripheral swelling (yellow/red) can be seen on the circumference of the occlusal table due to the Poisson surfacing algorithm closing the mesh defect caused during the creation process. The second column shows the colour maps after landmark alignment displaying tilts in the data set. The third column displays colour maps after a best-fit alignment. The defect appears to have been pulled occlusally decreasing the size of the defect. A profilometric gain (gain in tooth structure) in yellow on the buccal and lingual surfaces is also visible. The fourth column shows the colour maps after reference best-fit alignment. The appearance is very similar to the true defect, with the full degree of loss recorded.

4. Discussion

Perfect re-alignment will be difficult to obtain with digital comparison software. However, the technique is developing, and many researchers are quoting outcomes for measuring change. The technique using reference best-fit alignment significantly improved the alignment accuracy and decreased the measurement error. In contrast, landmark-based alignment and standard best-fit alignment resulted in statistically significant increased alignment errors. This resulted in significant underestimation of the defect size and errors consistent with profilometric gain over the data set.

Landmark alignment had the largest angular error which also resulted in the largest mean profilometric gain error. The large standard deviations reflect the difficulty of manually selecting convenient landmarks accurately at a micron level. This method creates greater inconsistencies in the data and poor inter-examiner reliability. In contrast, the best-fit align-

ment resulted in better angulation of the datasets but the greatest underestimation of the defect as the software minimised the difference between the two datasets. There was also a substantial amount of surface "gain" as the errors were spread evenly across the data set. This may explain why clinical investigations using this method of alignment have failed to show significant differences in the tooth wear progression over time between groups of different risk levels [8,13] compared to those who have used reference best-fit alignment [6].

Reference best-fit alignment significantly reduced the error for each of the measurement metrics. The absolute mathematical translation errors were 6 times smaller in the reference alignment group, while the angular (rotational) errors were half those of the full alignment group. The profilometric gain, profilometric loss and volume change error was also significantly reduced and not statistically different from the true defect. The reference surfaces used for this study represented a small section of the dataset (blue area in Fig. 2). Intuitively, one might think that a reduced dataset would adversely influence the alignment algorithm. However, as only one true fit exists for the minimal dataset, the alignment process was unaffected. Challenges remain in identifying these surfaces in an objective, robust method while minimising operator error. It may be that a best fit alignment could be used to identify areas of change before a more accurate selective surface alignment could be performed. A suggested approach may be to initially manually select reference areas which are unlikely to have experienced change based upon clinical knowledge. Alignment could then be restricted to sections on the dataset which have not experienced beyond a predefined threshold. A measurement process error of 15 microns have been observed by our group [5] and others [15,16] and thus a threshold of 20-25 microns as identified by a recent systematic review would seem reasonable [14]. Markers such as gingival margins, adjacent teeth movement and soft tissues are susceptible to change outside of this threshold and thus cannot be assumed to be reliable references.

Maximum positive and negative change measurements led to large standard deviations with no statistically significant differences between alignment methods and a lack of clinically relevant information. As the metric is reliant upon a single data point, any outliers present will corrupt the data. For this reason, maximum profilometric changes should be used with caution unless complete trueness of the dataset can be assumed, which is rare in clinically collected data. The mean positive and negative profilometric changes are well-accepted measures and yet resulted in large underestimations or overestimations of change depending the type of alignment used. When an average value for the entire dataset is taken, it may not always reflect the change being clearly visible in the colour maps. In contrast, volume change measurements are not reduced or averaged when substantial sections of the dataset have not experienced change. This may explain why authors using volumetric analysis to investigate wear progression observed differences in groups [6,7], while those relying on mean profile differences did not [13,18]. If the size of the surface area to be analysed is standardised (for example a 4×4 mm section of the dataset) then this may facilitate more accurate comparison of profilometric change measurements.

Previous work by our group [13] and others [12,19] have used best-fit alignments to measure changes in tooth tissue over time or assess the reproducibility of dental materials and tools. This data would suggest that we may be underestimating changes which have occurred. This study is limited in that we have only assessed the performance of Geomagic. Other software packages may result in a more accurate alignment and this will be a focus of future work. The findings from this paper highlight the importance of understanding the datasets which are being compared and choosing the most suitable measurement metric to ensure accurate, clinically relevant conclusions are reached.

5. Conclusions

Reference best-fit alignment resulted in significantly lower alignment errors and truer measurements. Both standard best-fit and landmark-based alignment algorithms significantly underestimated the size of the defect. Aligning data using a best-fit algorithm on selected surfaces which are unaffected by change can significantly improve the measurement accuracy. However challenges remain in identifying these surfaces with a clinically relevant, robust method and require validation with clinical, longitudinal data.

Funding

This research was supported by King's College London and the University of Leeds. There are no conflicts of interest to be declared.

REFERENCES

[1] Becker K, Wilmes B, Grandjean C, Drescher D. Impact of manual control point selection accuracy on automated surface matching of digital dental models. Clin Oral Investig 2018;22:801–10, http://dx.doi.org/10.1007/s00784-017-2155-6.

[2] Camardella LT, Breuning H, Vilella ODV. Are there differences between comparison methods used to evaluate the accuracy and reliability of digital models? Dental Press J Orthod 2017;22:65–74, http://dx.doi.org/10.1590/2177-6709.22.1.065-074.oar.

[3] Clozza E, Biasotto M, Cavalli F, Moimas L, Di Lenarda R. Three-dimensional evaluation of bone changes following ridge preservation procedures. Int J Oral Maxillofac Implants 2012;27:770–5.

[4] Szathvary I, Caneva M, Bressan E, Botticelli D, Meneghello R. A volumetric 3-D digital analysis of dimensional changes to the alveolar process at implants placed immediately into extraction sockets. J Oral Sci Rehabil 2015;1:63–9.

[5] Rodriguez JM, Austin RS, Bartlett DW. A method to evaluate profilometric tooth wear measurements. Dent Mater 2012;28:245–51, http://dx.doi.org/10.1016/j.dental.2011.10.002.

[6] Tantbirojn D, Pintado MR, Versluis A, Dunn C, Delong R. Quantitative analysis of tooth surface loss associated with gastroesophageal reflux disease. J Am Dent Assoc 2012;143:278–85, http://dx.doi.org/10.14219/jada.archive.2012.0153.

[7] Pintado MR, Anderson GC, DeLong R, Douglas WH. Variation in tooth wear in young adults over a two-year period. J Prosthet Dent 1997;77:313–20.
[8] Chadwick RG, Mitchell HL, Manton SL, Ward S, Ogston S, Brown R. Maxillary incisor palatal erosion: no correlation with dietary variables? J Clin Pediatr Dent 2005;29:157–64.
[9] Besl P, McKay N. A method for registration of 3-D shapes. IEEE Trans Pattern Anal Mach Intell 1992;14:239–56, http://dx.doi.org/10.1109/34.121791.
[10] Mitchell HL, Chadwick RG. Mathematical shape matching as a tool in tooth wear assessment—development and conduct. J Oral Rehabil 1998;25:921–8.
[11] Keeling AJ, Brunton PA, Holt RJ. An in vitro study into the accuracy of a novel method for recording the mandibular transverse horizontal axis. J Dent 2014;42:122–8, http://dx.doi.org/10.1016/j.jdent.2013.11.019.
[12] Ahmed K, Whitter J, Ju X, Pierce SG, MacLeod C, Murray C. A proposed methodology to assess the accuracy of 3D scanners, casts, and monitor tooth wear progression in patients. Int J Prosthodont 2016;29:514–21, http://dx.doi.org/10.11607/ijp.4685.
[13] Rodriguez JM, Austin RS, Bartlett DW. In vivo measurements of tooth wear over 12 months. Caries Res 2012;46:9–15, http://dx.doi.org/10.1159/000334786.
[14] Wulfman C, Koenig V, Mainjot AK. Wear measurement of dental tissues and materials in clinical studies: a systematic review. Dent Mater 2018;34(June 6):825–50, http://dx.doi.org/10.1016/j.dental.2018.03.002.
[15] Heintze SD, Zellweger G, Sbicego S, Rousson V, Noz-Viveros CM, Stober T, et al. Wear of two denture teeth materials in vivo—2-year results. Dent Mater 2013;29:e191–204, http://dx.doi.org/10.1016/j.dental.2013.04.012.
[16] Stober T, Bermejo JL, Rammelsberg P, Schmitter M. Enamel wear caused by monolithic zirconia crowns after 6 months of clinical use. J Oral Rehabil 2014;41:314–22, http://dx.doi.org/10.1111/joor.12139.
[17] Cignoni P, Callieri M, Corsini M. Meshlab: an open-source mesh processing tool; 2008 http://www.academia.edu/download/41716786/MeshLab_.
[18] Chadwick RG, Mitchell HL, Ward S. Rationalization of quantitative tooth surface loss data for epidemiological research. J Oral Rehabil 2004;31:42–6.
[19] Marghalani A, Weber HP, Finkelman M, Kudara Y, El Rafie K, Papaspyridakos P. Digital versus conventional implant impressions for partially edentulous arches: an evaluation of accuracy. J Prosthet Dent 2016;2017:574–9, http://dx.doi.org/10.1016/j.prosdent.2017.07.002.