King's Research Portal

DOI:
10.1016/j.jdent.2024.104863

Document Version
Publisher's PDF, also known as Version of record

Link to publication record in King's Research Portal

Citation for published version (APA):
Sun, N., Bull, T., Austin, R., Bartlett, D., & O'Toole, S. (2024). Quantifying error introduced by iterative closest point image registration. *Journal of Dentistry*, *142*, Article 104863. https://doi.org/10.1016/j.jdent.2024.104863

Citing this paper
Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights
Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

•Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
•You may not further distribute the material or use it for any profit-making activity or commercial gain
•You may freely distribute the URL identifying the publication in the Research Portal

Take down policy
If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

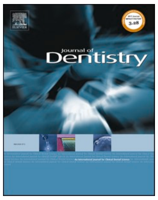
Download date: 20. Oct. 2024

Journal of Dentistry 142 (2024) 104863

Contents lists available at ScienceDirect

Journal of Dentistry

journal homepage: www.elsevier.com/locate/jdent



Quantifying error introduced by iterative closest point image registration

Ningjia Sun [a,*], Thomas Bull [b], Rupert Austin [a], David Bartlett [a], Saoirse O'Toole [a]

[a] *Centre for Clinical, Oral and Translational Sciences, Faculty of Dental, Oral and Craniofacial Sciences, King's College London, Floor 17, Tower Wing, Guy's Hospital, SE1 9RT, UK*

[b] *Mechanical Engineering Department, University of Southampton, 6 University Rd, Southampton SO17 1HE, UK*

ARTICLE INFO

Keywords:
Dental informatics/bioinformatics
Diagnostic systems
Dimensional change
Imaging
Oral diagnosis
Surface metrology software

ABSTRACT

Objectives: The aim of this paper was to quantify the analysis error introduced by iterative closest point (ICP) image registration. We also investigated whether a subsequent subtraction process can reduce process error.

Methods: We tested metrology and two 3D inspection software using calibration standards at 0.39 µm, and 2.64 µm and mathematically perfect defects (softgauges) at 2 and 20 µm, on free form surfaces of increasing complexity and area, both with and without registration. Errors were calculated in percentage relative to the size of the defect being measured. Data were analysed in GraphPad Prism 9, normal and two-way ANOVA with post-hoc Tukey's was applied. Significance was inferred at $p < 0.05$.

Results: Using ICP registration introduced errors from 0 % to 15.63 % of the defect size depending on the surface complexity and size of the defect. Significant differences were observed in analysis measurements between metrology and 3D inspection software and within different 3D inspection software, however, one did not show clear superiority over another. Even in the absence of registration, defects at 0.39 µm, and 2.64 µm produced substantial measurement error (13.39–77.50 % of defect size) when using 3D inspection software. Adding an additional data subtraction process reduced registration error to negligible levels (<1 % independent of surface complexity or area).

Conclusions: Commercial 3D inspection software introduces error during direct measurements below 3 µm. When using an ICP registration, errors over 15 % of the defect size can be introduced regardless of the accuracy of adjacent registration surfaces. Analysis output between software are not consistently repeatable or comparable and do not utilise ISO standards. Subtracting the datasets and analysing the residual difference reduced error to negligible levels.

Clinical significance: This paper quantifies the significant errors and inconsistencies introduced during the registration process even when 3D datasets are true and precise. This may impact on research diagnostics and clinical performance. An additional data processing step of scan subtraction can reduce this error but increases computational complexity.

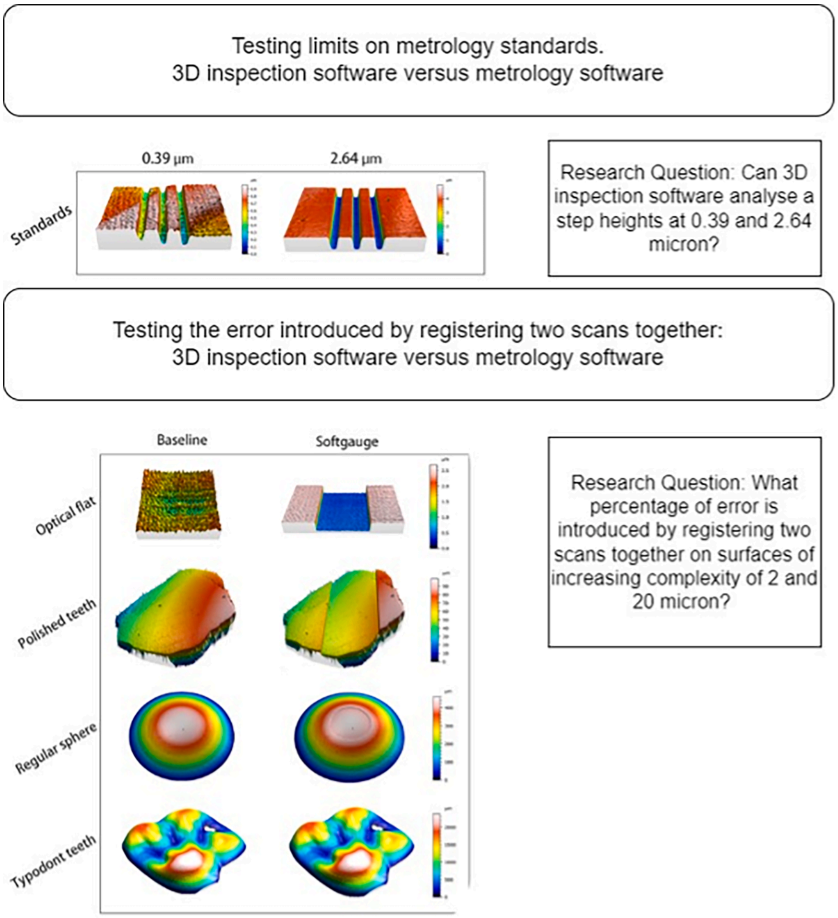
1. Introduction

For decades, in vitro tooth wear research has relied upon precise measurements using profilometry, accurate to nanometers, and the use of metrology grade software. Metrology grade software operates according to ISO standards and often have validated protocols for inspection of metrological change. Measurements are reliable and reproducible. However, the use of 3D inspection software's (Geomagic, Meshmixer, WearCompare etc.) for medical image diagnostics on sequential or multi-modal scans is increasing [1,2]. Most 3D inspection software's are derived from engineering analysis tools designed to evaluate deviation between an explicitly defined geometric reference shape, such as computer models, to the measurement of a manufactured component. However, they can be used for multiple purposes and in multiple ways. Most software do not have validated protocols for inspection of a specific metrological change. More importantly, most 3D inspection software rely upon a form of iterative closet point (ICP) based registration in order to register (superimpose) two surfaces [3]. ICP registration, developed by Besl and McKay [4], is ubiquitously used (at the point of submission, it has been cited 25, 576 times). The algorithm brings two point clouds iteratively to the closest possible proximation based upon mathematically calculating the least square difference

---

[*] Corresponding author.
E-mail address: ningjia.sun@kcl.ac.uk (N. Sun).

https://doi.org/10.1016/j.jdent.2024.104863
Received 12 September 2023; Received in revised form 23 January 2024; Accepted 24 January 2024


Testing limits on metrology standards.
3D inspection software versus metrology software

Research Question: Can 3D inspection software analyse a step heights at 0.39 and 2.64 micron?

Testing the error introduced by registering two scans together:
3D inspection software versus metrology software

Research Question: What percentage of error is introduced by registering two scans together on surfaces of increasing complexity of 2 and 20 micron?

Fig. 1. Experimental design and research questions. First line of images: Last four lines of images: Mathematically created softgauges on free form surfaces of increasing complexity and area (2 µm, 20 µm).

between corresponding points. However, it is highly susceptible to outliers and minor changes in data can dramatically transform the registration. This can make the process unreliable and difficult to reproduce depending on the complexity of the dataset. The use of ICP based registration in both sequential or multi-modal medical scans has been recognised to introduce error, particularly when attempting to quantify biological change [5]. ICP registration is an example of a rigid transformation where a source point cloud of 3D data is kept rigid, i.e. the distance between points in the cloud is preserved, but the cloud is rotated and translated to align to a reference [6]. This is an inherent problem with biological tissues, which deform over time, such that the rigid transformation assumption will always introduce inaccuracy. When a section of the measured point cloud changes at a different rate, i.e. during a pathological process, an ICP registration will attempt to minimise the change at the expense of a holistic scan fit. This has led to a situation where the greater the change and need for an accurate registration, the less accurate an ICP based registration will be [5,7]. Attempts have been made to reduce this error by registering scans on areas which are less likely to have undergone biological change or areas of known relative stability [8,9]. However, it can be problematic identifying these areas, particularly when identifying sub-visual change at the micrometer scale. It has also yet to be investigated if a reference based ICP registration can perform a perfect registration.

Adding an additional computational step of data subtraction, post-ICP registration has been shown to reduce ICP based errors and improve diagnostic potential for several medical image registrations from localisation of neocortical seizures in the brain [10,11] to analysing pulmonary parenchyma in chest CT images [12]. Subtraction reduces errors created by registration as minor differences in the X and Y plane can be compensated for and differences in Z are reduced to a single plane where change and error are more readily identifiable. Subtraction can involve the simple subtraction of X and Y coordinates, or a more computationally heavy subtraction of planes, both of which leave a residual dataset to be analysed. Although the medical examples above have shown reduction in error, inaccuracies in realms of millimeters remain. Dental indications using subtraction post-registration have demonstrated accuracy on 20–160 µm softgauges [13,14]. However, this is still beyond the diagnostic threshold of multiple conditions, including erosive toothwear. Several clinical trials in tooth wear have attempted to quantify tooth wear progression using an ICP based algorithm [15,16–18]. However, largely due to high standard deviations, very few studies have shown differences between high and low risk groups. The only clinical studies which have shown differences between groups are those which have relied upon volume measurements which consider an overall change in the surface [19–22]. The only exception to this is by Bronkhorst et al. who also had high standard deviations (eg wear on the upper central incisor had an average height loss per year of 87.31 µm with a standard deviation of 97.60 µm) but had a significant sample size with three different time points [23].

Each operating system will also have a variation on the ICP performed. This is often not detailed by commercial software due to intellectual property issues. Different measurements may be observed depending on how the software processes meshes and different scan orientation [5]. Although several deep learning techniques have been successfully applied to image registration, success is often denoted as being comparable to an operator-led registration [2]. Given the substantial error within these operator-led approaches, advantages in using deep learning techniques are limited to an increase in speed and decreased operator involvement, not improvement in accuracy. Therefore, additional research is needed to determine basic software limitations and improve on registration accuracy before using more complex, learning algorithms to determine biological change.

Therefore, the aim of this paper is to establish the limitations of analysis, in the absence of registration and post registration. We will test different reference ICP based registration algorithms on free form surfaces of increasing complexity and area, investigate the scale of the error introduced by ICP. Finally, we will investigate whether a subsequent subtraction process can reduce error introduced by the ICP registration to the micrometer level expected for early erosive tooth wear.

The null hypothesis proposed are that

1. 3D inspection software are comparable to metrology software when analysing change in the absence of registration.
2. Using an ICP based registration solely on unaltered reference data will not introduce significant analysis error.
3. Data subtraction post-registration will not reduce error.

2. Materials and methods

2.1. Standard selection and preparation

To assess measurement differences between software, two type A1 step height reference standards were chosen (Taylor Hobson; Ametek, Leicester, UK), with calibrated step artefacts of 0.39 and 2.64 μm. To assess the effect of registration on increasingly complex surfaces, an optical flat, polished dental enamel, a precision ball bearing and a typodont tooth were chosen and pictured in Fig. 1. An optical flat is considered a "perfect plane" and is a transparent ceramic, with a flatness tolerance of 25 nm [24]. To obtain polished teeth samples, extracted caries-free human molar teeth were collected from the Oral Surgery Department at Guy's Hospital under the consent of patients (IRAS 252, 842 REC Ref 18/WM/0351). They were disinfected and stored in deionized water. Buccal surfaces of molars were sectioned with a water-cooled circular diamond blade (XL 12,205; Benetec, London, UK) and embedded in bis-acryl composite (3 M; Protemp4, Neuss, Germany). A polishing process was applied by using the water-cooled rotation polishing unit (Struers; Laboforce-100TM, Marne, France) followed by silicon carbide discs (Struers ApS; Versocit, Ballerup, Denmark). An outer layer of 400 μm of enamel was removed by the sequential polishing process of 500, 1200, 2000 and 4000 grit for 25, 30, 60 and 120 s respectively [25], which controlled the flatness tolerance of all samples within ± 0.4 μm. A precision stainless steel ball bearing with a 20 mm diameter with a confirmed ± 2.5 μm tolerance on roundness and *a* ± 12.5 μm tolerance on specified diameter. An upper right second molar plastic typodont tooth model (Frasaco GmbH; Tettnang, Germany) to be used as a complex free form surface, with no published tolerance. These surfaces can be visualised in Fig. 1.

2.2. Scanning

All surfaces were measured with a non-contacting profilometer using a red-light 655 nm wavelength confocal laser with 2 μm spot-diameter and 600 μm vertical gauge (TaiCaan Technologies; Southampton, UK), The profilometer was kept in a temperature-controlled room to minimise thermal variation [26]. Data were collected in a raster scanning pattern at 10 μm scanning intervals. The optical flat standard and polished teeth were measured over a 2 × 2 mm^2 area, the ball bearing over a 7 × 7 mm^2 area and the typodont tooth over a 10 × 10 mm^2 area. Each sample was measured three times. An example of each scan type is shown in Fig. 1.

2.3. Software comparison

Softgauges were chosen as the software measurement standard. These mathematically created standards contain specified defects facilitating accurate comparison of data from different software and techniques [27]. To create the softgauges, the measured data is used as a reference point cloud. The reference data is exported into Excel (Microsoft; Redmond, Wash., USA) and a vertical step is mathematically introduced by subtracting either 2 μm or 20 μm from the Z height values of 500 central coordinates within each scan. The provides a pair of data sets comprising a reference measurement and complimentary softgauge.

Two different software were chosen to compare the ICP based registrations. The first was Geomagic (Geomagic Control; Darmstadt, Germany), a professional engineering software for 3D inspection of digital scans. The second was WearCompare (Wearcompare; Leeds, UK), a simple, custom-built open-source software for the detection of wear on 3D scans. To assess the additional computational step of subtraction, Mountains Map (MountainsMap; Besançon, France), an image analysis and surface metrology software platform to study surface texture and form in 3D at the microscopic scale was chosen.

2.3.1. Calibration standard and softgauge analysis without registration

Geomagic Protocol: For analysis of the standard, a 2D section was created through the centre of the scan and a single line difference in Z between the reference area and step height determined.

Mountains Map: For analysis of the calibration standard, a step height in Z was determined by applying ISO5436–1:2000 step height measurement method on the profiles.

2.3.2. Softgauge analysis with registration

Geomagic Protocol: The reference data and softgauge were registered using the Geomagic specific "Best Fit" registration on 1000 of the unaltered reference data points. To calculate the impact of the best fit registration a 2D section was obtained through the centre of the mathematically introduced step and the mean mesh difference in Z between the baseline and altered scan determined.

WearCompare Protocol: The reference data set and softgauge were exported into WearCompare. Measurement data were registered using an initial global based registration followed by an ICP registration on manually selected unaltered regions. The mean mesh difference in Z between the baseline and altered scan was determined from the centre of the artificially created defect.

Geomagic/Mountains Protocol: For the subtraction analysis, after scans were registered in Geomagic they were exported into Mountains Map. The root mean square deviation in the X and Y dimension was minimised and the two scans subtracted. This resulted in a residual dataset to which an ISO5436–1:2000 step height measurement method was applied.

Each set of three scans were analysed separately and the average value calculated for analysis give. All errors were reported in relative terms to the size of the defect being measured. Data were analysed in GraphPad Prism 9 (GraphPad; Boston, MA, USA). The normality of data sets was inspected visually using histograms and boxplots in addition to Shapiro-Wilk tests. All data sets were normally distributed. A two-way ANOVA was applied using software and surface as the independent variables and percentage error as the dependent variable. The interactions were significant therefore a post hoc Tukey's analysis was applied. Statistical significance was inferred at $p < 0.05$ for all analysis.

3. Results

3.1. Errors introduced by 3D inspection software in the absence of registration

A single measurement of a defect on a 3D inspection software (Geomagic) to a metrological software (Mountains Map) were

Table 1
Percentage error (%) and standard deviation associated with no registration (left columns) and post registration (right columns) at different known lesion depths. Note consistently lower error levels when subtraction was added as a final data processing step (final column).

Calibration Standards					
	Percentage Error With No Registration (%)		Percentage Error Post-Registration (%)		
	Mountains	Geomagic	Geomagic	WearCompare	Geomagic/Mountains (Registration/Subtraction)
	% (SD)	% (SD)	% (SD)	% (SD)	% (SD)
0.39 μm	11.88 (9.37)	13.39 (18.22)	37.26 (29.71)	41.71 (30.46)	12.22 (7.82)
2.64 μm	1.40 (0.92)	3.84 (2.63)	2.96 (2.08)	10.73 (6.58)	1.33 (1.10)
Soft Gauges					
	Optical Flat		Optical Flat		
2.00 μm	2.56 (1.45)	5.00 (4.63)	0.63 (1.77)	0.00 (0.00)	0.00 (0.00)
20.00 μm	0.27 (0.15)	0.44 (0.32)	0.00 (0.00)	0.00 (0.00)	0.13 (0.08)
	Polished Teeth		Polished Teeth		
2.00 μm	4.94 (2.31)	77.50 (12.54)	5.63 (6.23)	1.25 (2.32)	0.00 (0.00)
20.00 μm	1.03 (0.69)	3.06 (1.32)	1.38 (0.64)	0.50 (0.46)	0.41 (0.43)
			Ball Bearing		
2.00 μm			15.63 (9.43)	0.00 (0.00)	0.00 (0.00)
20.00 μm			6.81 (2.10)	1.56 (0.18)	0.10 (0.08)
			Typodont Teeth		
2.00 μm			8.75 (6.41)	13.13 (3.72)	1.06 (0.32)
20.00 μm			5.34 (1.83)	9.69 (1.25)	0.14 (0.12)

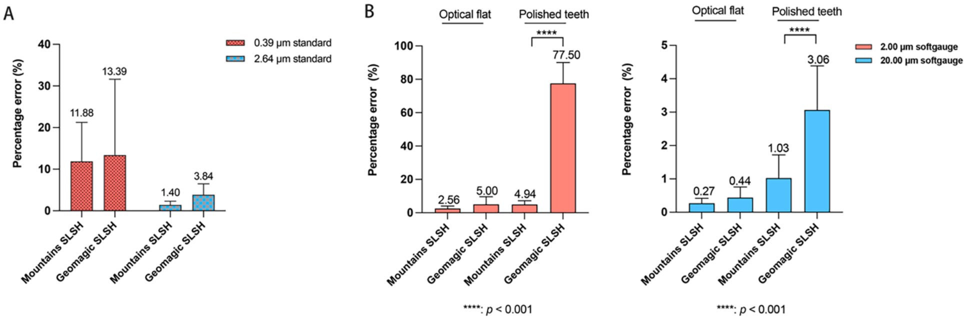


Fig. 2. Percentage error comparing Single Line Step Height (SLSH) analysis in the metrology software (Mountains) and the 3D inspection software (Geomagic) in the absence of any registration. This analysis cannot be performed in WearCompare. A, percentage error of SLSH analysis on calibrations (0.39 μm, 2.64 μm). B, Percentage error of SLSH analysis on softgauges (2 μm, 20 μm) created on optical flat and polished teeth (Statistically significant differences indicated with asterisk ($P<0.001$)).

compared. Calculating step height standards, in the absence of any registration, introduced an error of 11.88 % SD for the metrological software and 13.39 % for the 3D inspection software for the standard step height of 0.39 μm. When calculating the standard step height of 2.64 μm, errors reduced to 1.40 % (SD = 0.92) for the metrological software and 3.84 % (SD = 2.63) for the 3D inspection software (Table 1). There was no statistical difference in error rate between step heights or the software when measuring at this level.

When quantifying change between reference and softgauge data at a 2 μm level on flat surfaces, all software introduced error. This was particularly significant for the 3D inspection software where an average error of 77.50 % (SD = 12.54) was produced, $p < 0.001$. When quantifying change on flat surfaces with a 20 μm defect, percentage error significantly reduced. Quantifying step height on the 3D inspection software produced significantly greater error (3.06 % (SD = 1.32) than the metrological software (1.03 % SD = 0.69, $p < 0.001$). These results are shown in Fig. 2.

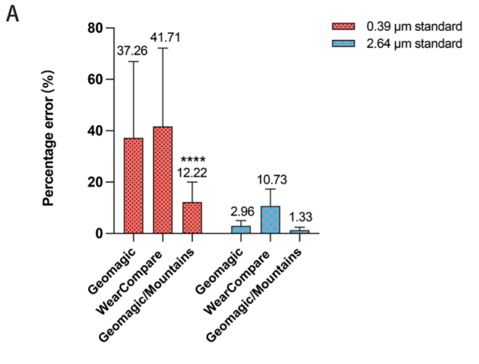
3.2. Errors introduced by 3D inspection software with registration

Here we compared two different 3D inspection software (Geomagic and WearCompare) with a metrological software (Mountains Map) when registering two scans together. Post registration, analysis errors increased when analysing the 0.39 μm calibration standard to 37.26 % (SD = 29.71) for Geomagic and 41.71 % (SD = 30.46) for WearCompare. Performing subtraction on the datasets and analysing the residual reduced the error to 12.22 % (SD = 7.81) $p < 0.001$.

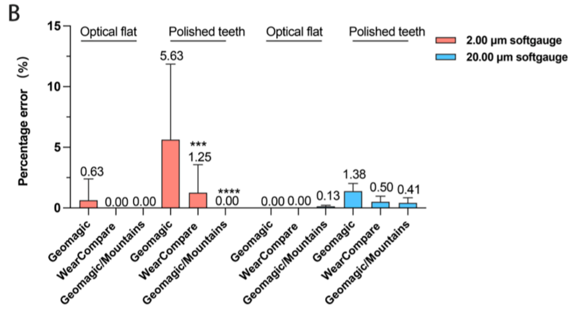
When scans were registered analysing the 2.64 μm calibration standard depth, observed errors were 2.96 % (SD = 2.08) for Geomagic, 10.73 % (SD = 6.58) for WearCompare and 1.33 % (SD = 1.10) for subtraction analysis.

When analysing known step heights on softgauges, errors were < 1% on an optical flat at both the 2 μm and 20 μm level. On polished teeth with the 2 μm softgauge, analysis with Geomagic introduced 5.63 % (SD = 6.23) error which was statistically different to analysis with WearCompare (1.25 % SD = 2.32). Performing subtraction after registration reduced the error to negligible levels (0.00 %). With the larger 20 μm softgauge on polished teeth, errors were reduced to circa 1 % with no differences between the software.

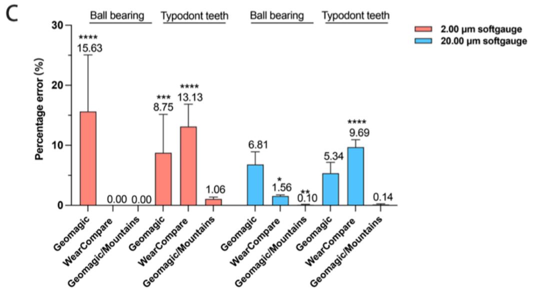
On the ball bearing surface, analysis with Geomagic introduced the greatest error, up to 15.63 % (SD = 9.43) with the 2 μm softgauge and 6.81 % (SD = 2.10) error with the 20 μm softgauge. Adding an additional subtraction step in Mountains map reduced error to negligible levels (<0.1 % on a 20 μm surface) which were not statistically different



****: $p < 0.001$ compared with Geomagic, WearCompare



***: $p < 0.01$ compared with Geomagic, ****: $p < 0.001$ compared with Geomagic



Ball bearing 2 µm (****: $p < 0.001$ compared with WearCompare and Geomagic/Mountains)
Typodont teeth 2 µm (***: $p < 0.01$ compared with Geomagic/Mountains; ****: $p < 0.001$ compared with Geomagic/Mountains)
Ball bearing 20 µm (*: $p < 0.05$ compared with Geomagic; **: $p < 0.01$ compared with Geomagic)
Typodont teeth 20 µm (****: $p < 0.001$ compared with Geomagic/Mountains)

Fig. 3. Percentage error on known softgauge depths post registration (Geomagic, WearCompare) and post-registration/subtraction (Geomagic/Mountains). A, Percentage error of analysis on calibrations (0.39 µm, 2.64 µm) Statistically significant differences indicated with asterisk (*P*<. 001 compared with Geomagic, WearCompare). B, Percentage error of analysis on softgauges (2 µm, 20 µm) created on optical flat and polished teeth. Statistically significant differences indicated with asterisk (*** *P*<. 0.01 compared with Geomagic; **** *P*<. 0.001 compared with Geomagic). C, Percentage error of analysis on softgauges (2 µm, 20 µm) created on ball bearing and typodont teeth. Statistically significant differences indicated with asterisk (Ball bearing 2 µm **** *P*<. 0.001 compared with WearCompare and Geomagic/Mountains; Typodont teeth 2 µm *** *P*<. 0.01 compared with Geomagic/Mountains, **** *P*<. 0.001 compared with Geomagic/Mountains; Ball bearing 20 µm (* *P*<. 0.05 compared with Geomagic, ** *P*<. 0.01 compared with Geomagic; Typodont teeth 20 µm (**** *P*<. 0.001 compared with Geomagic/Mountains).

to WearCompare. In contrast, on a typodont surface, both WearCompare and Geomagic introduced up to 13.13 % error (SD=3.72) on the 2 µm softgauge and 9.69 % (SD = 1.25) error on the 20 µm softgauge. Performing surface subtraction after registration reduced error to <1.06 % (SD=0.32).

These results can be visualised in Fig. 3.

4. Discussion

When measuring a 2.00 µm defect, independent of registration, a 3D inspection software produced significantly greater errors and greater standard deviations compared to metrology-based ISO standard measurement software. Although the error was less when quantifying a larger 20.00 µm defect, significant differences were still observed and a 3 % error rate was observed. Therefore the first null hypothesis was rejected. Errors were consistently introduced by registration in 3D inspection software, despite a well-defined lesion and a mathematically perfect reference area. The largest errors were observed when analysing the test extremes. Error rates were high when measuring 0.39 µm standards and 2.00 µm softgauge and indicated that the step heights were outside the limits of analysis with an ICP registration. Large errors were also seen when analysing 20 µm defects on typodont teeth. This demonstrates the potential for inconsistency and inaccuracy of the registration procedure. Error on in vivo biological data is likely to be significantly greater than this when accurate surfaces for registration are not available and defects are less well defined.

Despite Geomagic and WearCompare both using an ICP based registration statistically different results were obtained. Within each software, considerable standard deviations were also observed when analysing the same surfaces demonstrating how an ICP registration can produce different results when a repeat registration is performed. This would indicate that data analysed by different 3D inspection software may give incomparable results and secondly how the same software can produce different results even when fed the same data multiple times. This is in addition to changes caused by software updates which may remain unknown to the operator. Due to the black box nature of the fitting, the operator is unlikely to know when the registration is accurate and when it is flawed. A manual operator can visualise and inspect the registration. However, an algorithm cannot. This demonstrates further problems when applying machine learning techniques to ICP based registrations.

Authors have presented data obtained from these software in different ways. Michou et al. did not attempt quantification below 50 µm and removed positive errors by simply eliminating them from the dataset [7]. Rodriguez et al. also removed any positive errors from the dataset before presenting data. Bronkhurst et al. observed the measurement error using registration on intraoral scanners to be 62 µm [28]. Others using desktop scanning of 3D casts have observed errors of 33 µm [29]. The unreliability with the registration process contributes to the substantial standard deviations seen in many clinical trials which makes evaluation of statistical differences difficult. One clinical trial observed error of 10 µm associated with the registration of scans for each tooth. This is likely to significantly increase when analysis is done cross arch [30].

Adding an additional data processing step of subtraction post-registration consistently reduced error to negligible levels therefore the third null hypothesis was accepted. Subtraction can be a relatively straightforward process, superimposing the grids and subtracting X1-X2, Y1-Y2 etc. to give a difference in Z. However, this assumes that all data has been captured and orientated in the same location and plane. A more computationally heavy step is to resample all the data and subtract according to normals to a plane. However, subtraction will only perform well in an optimally calibrated and orientated environment. Variables such as increasing distance between vertices, polygon size, overlapping areas to be registered, the registration and the orientation of the scans will all impact on the accuracy of the subtraction step. A further limitation to the subtraction step is that data subtraction needs to be in one plane only. Clinically this means that it is limited to a single surface measurement with no undercuts. Therefore segmentation of the 3D dataset will be necessary to perform this analysis to get the required improvements in accuracy. There is ongoing work in several institutions and commercial companies to auto segment 3D intraoral scans and this field is rapidly evolving. With increased computing power and AI tools facilitating auto segmentation and registration, registration errors <1 % may be possible for intraoral scans in the near future.

There are several limitations to this study. As outlined above, the test dataset was a mathematically calibrated dataset chosen to assess the limitations of an ICP based algorithm. It is likely that a biological dataset will introduce error much greater than observed in this study. We chose two different ICP based registration software and it highly likely that other software could produce different results. There is no standard methodology and none conform to ISO standard methodologies. Each software will have different methods of processing the data which are not observable to the operator. Until we can understand each step in data processing, we will be unable to ascertain what proportion of the measurement given is error or an actual change between two sequential scans. Lastly, the datasets are relatively small and subtraction can be computationally heavy. This may produce complications when scaling up to larger scan files.

Considering these limitations, solely using an ICP based registration process is flawed, particularly on complex surfaces. An additional subtraction step is computationally heavy but significantly improves on accuracy. Subtraction is also reliant on scan orientation and further work is required to automate biological scan orientation. Additional research on non-ICP registration approaches or methods to limit the susceptibility of the ICP based process to outliers is also required to progress the medical image registration field.

5. Conclusion

Commercial 3D inspection software which do not conform to ISO standards can introduce error during simple direct measurements below 3 μm. When using an ICP registration, errors over 15 % of the defect size can be introduced regardless of the accuracy of adjacent registration surfaces. Different 3D inspection software will produce different errors on different samples which increases inconsistency in measurements. The greater the complexity of the surface, the harder it is to measure change. Error can be reduced by data set subtraction but this requires validation on clinical data and increased computational power.

Funding

This work was supported by the King's-China Scholarship Council [grant number 202106230065].

CRediT authorship contribution statement

Ningjia Sun: Investigation, Formal analysis, Data curation, Conceptualization. **Thomas Bull:** Software, Methodology, Data curation. **Rupert Austin:** Software, Methodology, Investigation, Data curation, Conceptualization. **David Bartlett:** Supervision, Project administration, Funding acquisition. **Saoirse O'Toole:** Writing – review & editing, Writing – original draft, Investigation.

Declaration of competing interest

The authors declare no potential conflicts of interest with respect to the authorship and/or publication of this article.

References

[1] K.T. Islam, S. Wijewickrema, S. O'Leary, A deep learning based framework for the registration of three dimensional multi-modal medical images of the head, Sci. Rep. 11 (2021) 1–13, https://doi.org/10.1038/s41598-021-81044-7.

[2] Y. Fu, Y. Lei, T. Wang, W.J. Curran, T. Liu, X. Yang, Deep learning in medical image registration: a review, Phys. Med. Biol. 65 (2020) 20TR01, https://doi.org/10.1088/1361-6560/ab843e.

[3] H.Y. Baluwala, L. Risser, J.A. Schnabel, K.A. Saddi, Toward physiologically motivated registration of diagnostic CT and PET/CT of lung volumes, Med. Phys. 40 (2013) 021903, https://doi.org/10.1118/1.4771682.

[4] P.J. Besl, N.D. McKay, Method For Registration of 3-D Shapes, Sensor Fusion IV: Control Paradigms and Data Structures, Spie, Boston, 1992, https://doi.org/10.1117/12.57955.

[5] S. O'Toole, C. Osnes, D.W. Bartlett, A. Keeling, Investigation into the validity of WearCompare, a purpose-built software to quantify erosive tooth wear progression, Dent. Mater. 35 (2019) 1408–1414, https://doi.org/10.1016/j.dental.2019.07.023.

[6] W.R. Crum, T. Hartkens, D. Hill, Non-rigid image registration: theory and practice, Br. J. Radiol. 77 (2004) S140–S153, https://doi.org/10.1259/bjr/25329214.

[7] S. Michou, C. Vannahme, K.R. Ekstrand, A.R. Benetti, Detecting early erosive tooth wear using an intraoral scanner system, J. Dent. 100 (2020) 103445, https://doi.org/10.1016/j.jdent.2020.103445.

[8] S. O'Toole, P. Charalambous, A. Almatrafi, S. Mukar, S. Elsharkawy, D.W. Bartlett, Progress and limitations of current surface registration methods when measuring natural enamel wear, J. Dent. 112 (2021) 103738, https://doi.org/10.1016/j.jdent.2021.103738.

[9] S. O'Toole, C. Osnes, D.W. Bartlett, A. Keeling, Investigation into the accuracy and measurement methods of sequential 3D dental scan alignment, Dent. Mater. 35 (2019) 495–500, https://doi.org/10.1016/j.dental.2019.01.012.

[10] P.J. Lewis, A. Siegel, A.M. Siegel, C. Studholme, J. Sojkova, D.W. Roberts, V. M. Thadani, K.L. Gilbert, T.M. Darcey, P.D. Williamson, Does performing image registration and subtraction in ictal brain SPECT help localize neocortical seizures? J. Nucl. Med. 41 (2000) 1619–1626.

[11] T.J. O'Brien, E.L. So, G.D. Cascino, M.F. Hauser, W.R. Marsh, F.B. Meyer, F. W. Sharbrough, B.P. Mullan, Subtraction SPECT coregistered to MRI in focal malformations of cortical development: localization of the epileptogenic zone in epilepsy surgery candidates, Epilepsia 45 (2004) 367–376, https://doi.org/10.1111/j.0013-9580.2004.54703.x.

[12] D. Grob, L. Oostveen, J. Rühaak, S. Heldmann, B. Mohr, K. Michielsen, S. Dorn, M. Prokop, M. Kachelrieβ, M. Brink, Accuracy of registration algorithms in subtraction CT of the lungs: a digital phantom study, Med. Phys. 46 (2019) 2264–2274, https://doi.org/10.1002/mp.13496.

[13] P. Charalambous, S. O'Toole, T. Bull, D.W. Bartlett, R.S. Austin, The measurement threshold and limitations of an intra-oral scanner on polished human enamel, Dent. Mater. 37 (2021) 648–654, https://doi.org/10.1016/j.dental.2021.01.006.

[14] P. Charalambous, S. O'Toole, R.S. Austin, D.W. Bartlett, The threshold of an intra oral scanner to measure lesion depth on natural unpolished teeth, Dent. Mater. 38 (2022) 1354–1361, https://doi.org/10.1016/j.dental.2022.06.022.

[15] J. Rodriguez, R.S. Austin, D.W. Bartlett, In vivo measurements of tooth wear over 12 months, Caries Res. 46 (2012) 9–15, https://doi.org/10.1159/000334786.

[16] R.G. Chadwick, H.L. Mitchell, S.L. Manton, S. Ward, S. Ogston, R. Brown, Maxillary incisor palatal erosion: no correlation with dietary variables? J. Clin. Pediatr. Dent. 29 (2005) 157–164, https://doi.org/10.17796/jcpd.29.2.g413861067g2g884.

[17] M.A. Schlenz, M.B. Schlenz, B. Wöstmann, A.S. Glatt, C. Ganss, Intraoral scanner-based monitoring of tooth wear in young adults: 24-month results, Clin. Oral Investig. 27 (2023) 2775–2785, https://doi.org/10.1007/s00784-023-04858-x.

[18] H. Bronkhorst, E. Bronkhorst, S. Kalaykova, T.P. Cenci, M.C. Huysmans, B. Loomans, Inter-and intra-variability in tooth wear progression at surface, tooth and patient level over a period of three years: a cohort study: inter-and intra-variation in tooth wear progression, J. Dent. 138 (2023) 104693, https://doi.org/10.1016/j.jdent.2023.104693.

[19] M.R. Pintado, G.C. Anderson, R. DeLong, W.H. Douglas, Variation in tooth wear in young adults over a two-year period, J. Prosthet. Dent. 77 (1997) 313–320, https://doi.org/10.1016/s0022-3913(97)70189-6.

[20] D. Antbirojn, M.R. Pintado, A. Versluis, C. Dunn, R. Delong, Quantitative analysis of tooth surface loss associated with gastroesophageal reflux disease: a longitudinal clinical study, J. Am. Dent. Assoc. 143 (2012) 278–285, https://doi.org/10.14219/jada.archive.2012.0153.

[21] F. Marro, S. O'Toole, E. Bernabé, D. Bartlett, V. Aránguiz, Associated risk factors with quantitative erosive tooth wear progression, J. Dent. 123 (2022) 104179, https://doi.org/10.1016/j.jdent.2022.104179.

[22] S. O'Toole, T. Newton, R. Moazzez, A. Hasan, D. Bartlett, Randomised controlled clinical trial investigating the impact of implementation planning on behaviour related to the diet, Sci. Rep. 8 (2018) 8024, https://doi.org/10.1038/s41598-018-26418-0.

[23] H. Bronkhorst, E. Bronkhorst, S. Kalaykova, W.V.D. Meer, M.C. Huysmans, B. Loomans, J. Vis. Exp. 12 (2022) 185, 10.3791/63680.

[24] R.S. Austin, K.S. Stenhagen, L.H. Hove, S. Dunne, R. Moazzez, D.W. Bartlett, A. B. Tveit, A qualitative and quantitative investigation into the effect of fluoride formulations on enamel erosion and erosion–abrasion in vitro, J. Dent. 39 (2011) 648–655, https://doi.org/10.1016/j.jdent.2011.07.006.

[25] P. Mylonas, R.S. Austin, R. Moazzez, A. Joiner, D.W. Bartlett, In vitro evaluation of the early erosive lesion in polished and natural human enamel, Dent. Mater. 34 (2018) 1391–1400, https://doi.org/10.1016/j.dental.2018.06.018.

[26] P. Mylonas, T. Bull, R. Moazzez, A. Joiner, D.W. Bartlett, Detection threshold of non-contacting laser profilometry and influence of thermal variation on characterisation of early surface form and textural changes in natural human enamel, Dent. Mater. 35 (2019) e140–e152, https://doi.org/10.1016/j.dental.2019.04.003.

[27] A. Chiboub, Y. Arezki, A. Vissiere, C.M. Souzani, N. Anwer, B. Alzahrani, M. L. Bouazizi, H. Nouira, Generation of reference softgauges for minimum zone fitting algorithms: case of aspherical and freeform surfaces, Nanomaterial 11 (2021) 33862021, https://doi.org/10.3390/nano11123386.

[28] H. Bronkhorst, E. Bronkhorst, S. Kalaykova, W.V.D. Meer, M.C. Huysmans, B. Loomans, Precision of *in vivo* quantitative tooth wear measurement using intra-

oral scans, J. Vis. Exp. 12 (2022) 185, https://doi.org/10.3791/63680. PMID: 35913137.

[29] K.E. Ahmed, J. Whitters, X. Ju, S.G. Pierce, C.N. MacLeod, C.A. Murray, A proposed methodology to assess the accuracy of 3D scanners and casts and monitor tooth wear progression in patients, Int. J. Prosthodont. 29 (2016) 514–521, https://doi.org/10.11607/ijp.4685.

[30] M.A. Schlenz, M.B. Schlenz, B. Wöstmann, A. Jungert, C. Ganss, Intraoral scanner-based monitoring of tooth wear in young adults: 12-month results, Clin. Oral Investig. 1 (2022) 1869–1878, https://doi.org/10.1007/s00784-021-04162-6.