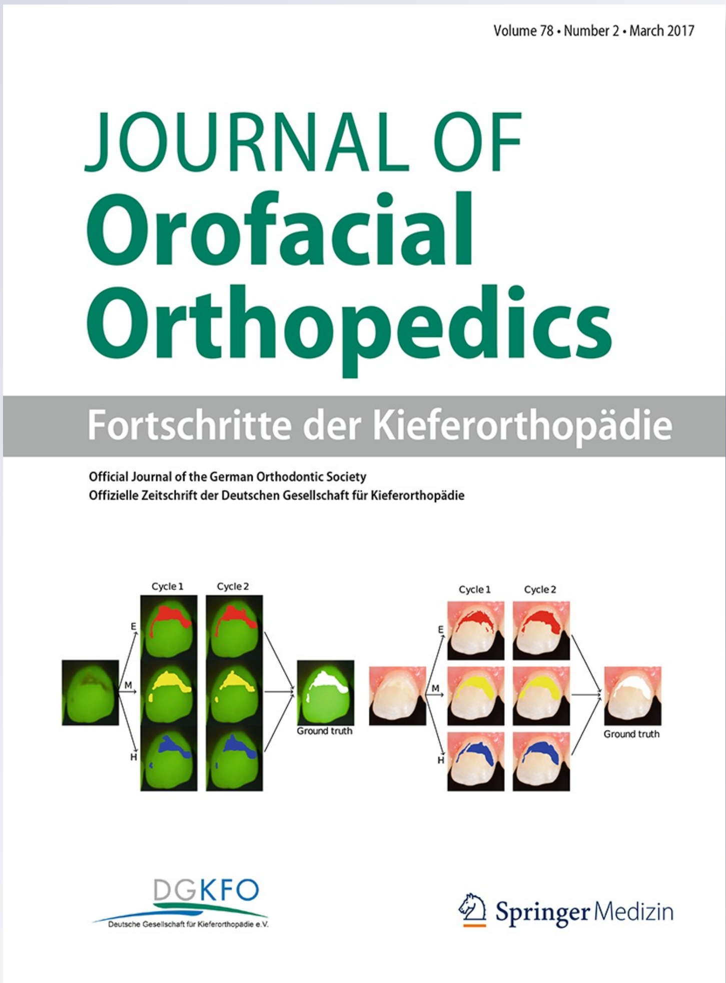
Quantitative light-induced fluorescence images and digital photographs - Reproducibility of manually marked demineralisations

Rosalia Tatano, Eva E. Ehrlich, Benjamin Berkels, Ekaterina Sirazitdinova, Thomas M. Deserno & Ulrike B. Fritz

Journal of Orofacial Orthopedics / Fortschritte der Kieferorthopädie
Official Journal of the German Orthodontic Society / Offizielle Zeitschrift der Deutschen Gesellschaft für Kieferorthopädie

ISSN 1434-5293
Volume 78
Number 2

J Orofac Orthop (2017) 78:137-143
DOI 10.1007/s00056-016-0069-6



Springer

J Orofac Orthop (2017) 78:137–143
DOI 10.1007/s00056-016-0069-6

ORIGINAL ARTICLE

Quantitative light-induced fluorescence images and digital photographs - Reproducibility of manually marked demineralisations

Bildgebung mit quantitativer lichtinduzierter Fluoreszenz und digitalen Aufnahmen - Reproduzierbarkeit manueller Demineralisationsmarkierungen

Rosalia Tatano[1] · Eva E. Ehrlich[2] · Benjamin Berkels[1] · Ekaterina Sirazitdinova[3] · Thomas M. Deserno[3] · Ulrike B. Fritz[2]

Received: 20 April 2016 / Accepted: 30 September 2016 / Published online: 20 February 2017

Abstract

Objective Hard tooth tissue demineralisation is an undesirable side effect of orthodontic treatment with fixed appliances. Whereas both clinically and in digital photographs (DP), demineralisations appear as white spot lesions, WSLs appear as dark areas when quantitative light-induced fluorescence (QLF) imaging is used. This study aims at comparing the reproducibility of the detection of decalcified tooth areas in DP and QLF.

Materials and methods DP and QLF pairs were acquired from 139 teeth of 32 patients after braces removal. Three raters manually marked the decalcified area on both DP and QLF images. The markings were repeated after 2 weeks. A ground truth was estimated for each tooth and modality using the simultaneous truth and performance level estimation (STAPLE) algorithm. The Dice coefficients (DC) of each rater marking to the ground truth were calculated for all teeth and modalities to quantify the spatial agreement. A three-way repeated measures analysis of variance (ANOVA) was used to compare the means of the DCs for both modalities ($p < 0.05$). Intra-observer and intercycle variabilities were assessed comparing the means across the raters and the cycles for both modalities.

Results ANOVA revealed a statistical significant difference between the modalities [$F(1, 138) = 62.89$, $p < 0.001$]. The standard deviation of the DC for the photographs are lower than those for the QLF images. Intra-observer and intercycle differences are rather small as compared to the intermodality differences.

Conclusions The results indicate a higher spatial reproducibility in identifying a decalcified area on a tooth surface using visual inspection of DP rather than QLF images.

Keywords Demineralisation · QLF · Digitalphotograph · Reproducibility · STAPLE algorithm

Zusammenfassung

Ziel Zahnhartsubstanzschäden in Form von Demineralisationen (WSLs) gelten als unerwünschte Nebenwirkung der kieferorthopädischen Behandlung mit festsitzenden Apparaturen. Sowohl klinisch als auch in digitalen Fotografien (DF) stellen sich WSLs als weißlich-opake, bei der QLF-Diagnostik als dunkle Bereiche dar. Ziel der vorliegenden Studie ist es, die Reproduzierbarkeit von Demineralisationsmarkierungen auf Basis der DF- und QLF-Bildgebung zu vergleichen.

Material und Methoden DP- und QLF-Aufnahmen wurden von 139 Zähnen bei 32 Patienten nach Entfernung einer festsitzenden Apparatur erstellt. Drei Untersucher markierten manuell die Demineralisationen in beiden Modalitäten. Die Markierungen wurden 14 Tage später wiederholt. Für jeden Zahn und jede Modalität wurde mithilfe des STAPLE (Simultaneous Truth and

✉ Rosalia Tatano
tatano@aices.rwth-aachen.de

1 Aachen Institute for Advanced Study in Computational Engineering Science (AICES), RWTH Aachen University, Schinkelstr. 2, 52062 Aachen, Germany

2 Uniklinik RWTH Aachen, Klinik für Kieferorthopädie, Pauwelsstr. 30, 52074 Aachen, Germany

3 Uniklinik RWTH Aachen, Institut für Medizinische Informatik, Pauwelsstr. 30, 52074 Aachen, Germany

Performance Level Estimation)-Algorithmus eine Ground Truth ermittelt. Die örtliche Übereinstimmung der Untersuchermarkierungen zur Ground Truth wurde mit dem Dice-Koeffizient (DC) berechnet. Mittels mehrfaktorieller Varianzanalyse (ANOVA) wurden die Mittelwerte der DC-Werte für beide Modalitäten verglichen ($p < 0{,}05$). Die Intra-Untersucher- und die Inter-Zyklus-Variabilität wurden durch Vergleich der DC-Mittelwerte und Varianzen bewertet.

Ergebnisse ANOVA zeigte einen statistisch signifikanten Unterschied zwischen den Modalitäten [F (1; 138) = 62,89, $p < 0{,}001$]. Die Standardabweichung des DC war bei den digitalen Bildern geringer als bei der QLF-Bildgebung. Die Unterschiede zwischen den Untersuchern und den Zyklen waren vergleichsweise gering.

Schlussfolgerung Die Ergebnisse zeigen eine größere örtliche Reproduzierbarkeit der Demineralisationsmarkierung auf vestibulären Glattflächen auf den Fotographien als auf korrespondierenden QLF-Aufnahmen.

Schlüsselwörter Demineralisation · QLF · Digitale Fotographien · Reproduzierbarkeit · STAPLE-Algorithmus

Introduction

Enamel demineralisation is considered an undesirable, irreversible iatrogenic side effect of orthodontic treatment with fixed appliances [11]. Due to insertion of bands, brackets and wires, the patient's oral hygiene is complicated considerably [24]. Repeated plaque accumulation increases the risk of hard tooth tissue degradation or the incidence of white spot lesions (WSL), the chalky white, opaque enamel areas [16].

Compared to untreated patients, the risk of WSL formation is two to three times higher in patients with fixed orthodontic appliances [19]. The prevalence of WSL in orthodontic patients varies between 4.9 [16] and 84% [23]. In addition to aesthetic impairment [20], WSL is characterized by mineral loss [25] and is regarded as a primary stage of incipient caries (initial lesion).

In addition to clinical inspection and digital photography (DP), various optical methods have been developed to quantify enamel demineralisation. Nondestructive methods for quantifying progression and/or regression of enamel lesions offer the opportunity to monitor changes related to preventive measures over time. With objective quantitative diagnostic methods, the intra- and interexaminer variations would likely decrease and it would be possible to monitor changes of mineral loss or gain and to provide early feedback to the patient [3].

Noninvasive quantitative light-induced fluorescence (QLF) measurements offer further information about the quality of WSL [2], since fluorescence of enamel is directly proportional to its mineral content [13]. Different fluorescence capacities of sound and demineralized enamel allow the identification of a lesion and the assessment of its progression. QLF reports also give information about the demineralization, qualitatively as well as quantitatively [3].

Several publications have focused on WSL diagnosis using clinical photos [4, 5, 31] and QLF-based evaluation [6, 12, 18, 21]. Both diagnostic measures showed good reproducibility in terms of intra- and interexaminer reliability [6, 26, 28].

However, WSL diagnosis is predominantly based on the visual accessibility of the lesion in the picture, which is called segmentation. The reliability of humans (or computer algorithms) performing image segmentation is difficult to quantify because the true area that needs to be marked, also referred to as ground truth or gold standard [7], is usually unknown [29]. In this context, the simultaneous truth and performance level estimation (STAPLE) algorithm has been introduced to manage the multi-observer variability and iteratively determining an estimated ground truth that has been used already successfully in many applications of medical image segmentation [1, 9, 15].

The aim of the present study is to compare the spatial reproducibility of the detection of decalcified enamel based on digital photographs and QLF. As gold standard, we use the ground truth delivered from the STAPLE algorithm. Our hypothesis is that the more appropriate imaging modality supports a more reliable reading by the human expert.

Materials and methods

A total of 32 patients (17 female, 15 male) were examined from November 2013 to November 2014 after orthodontic treatment with a fixed appliance. All patients exhibited white spot lesions to various extents and severity at one or more teeth after debonding. Mean age of the patients was 17 years (range 13–38 years). The average duration of treatment time was 26 months (range 12–45 months). Due to their good accessibility, upper and lower incisors and canines were recorded. DP and QLF images were taken (Fig. 1). A total of 139 pairs of images were included in the study.

Debonding

The fixed appliances were debonded using Weingart pliers. Carbide burs were utilized for adhesive removal. Finally, the teeth were polished using polishing brush and paste. An average of 4.9 months (range 0–31.5 months) passed between the debonding and the acquisition of the images.

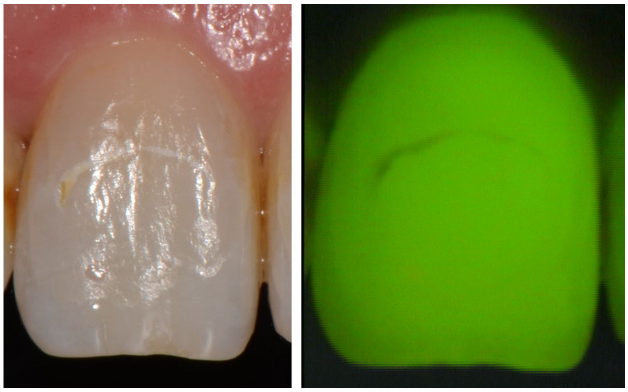


Fig. 1 DP/QLF images of a central incisor
Abb. 1 Exemplarisches DP/QLF-Bildpaar eines mittleren Frontzahnes

Images acquisition

A digital single lens reflex (DSLR) camera (D7000, Nikon, Japan) was used to acquire the digital photographs under standardized conditions. A macro lens (Nikkor 105 mm, 1:2.8, Nikon, Japan) with a ring flash (Sigma EM-140 DG, Kawasaki, Japan) was used. The camera settings were manually specified: shutter speed 1/250 s, aperture f/29, film speed ISO 100, and flash setting $\frac{1}{4}$s. After careful cleaning with polishing cup and paste (Zircate Prophy, Dentsply Sirona International, York, PA, USA) and air drying, the teeth were photographed with the camera at right angles to its longitudinal axis. A black contraster was positioned behind the teeth to achieve a better visualization of the tooth contour.

The QLF images were recorded using the Inspector Pro system, composed of QLF (WV-KS 152 QLF-clin, Panasonic, Japan) and Inspector Pro Software Version 2.0.0.49 (Inspector Research System BV, The Netherlands). This system integrates the camera in a toothbrush handle to record the light reflected from the surface of the tooth that is exposed to a xenon lamp (13 mW/cm^2) of wavelength 370 $\pm$ 80 nm. Furthermore, the system applies a filter to detect the emitted fluorescence light of wavelength 520 nm.

To avoid additional exposures and to ensure the hygienic conditions on the patient, we applied an auxiliary attachment to the toothbrush handle (ambient light shield). Before each measurement, the tooth surface was again cleaned and dried. The examination room was darkened. The handle was carefully positioned such that the incident light meets the tooth surface at a right angle.

Manual marking

All DP/QLF pairs were analyzed in randomized order by three raters, indicated with E, M, and H, applying manual markings to identify the decalcified area on the examined tooth surface (Cycle 1). To assess intrarater reliability in both modalities, the markings were repeated (Cycle 2). To eliminate any memory bias, at least 2 weeks were allowed to pass between both cycles, and the order of Cycle 2 was randomized again. Thus, for each tooth and each modality, six markings of the decalcified area are available.

Ground truth estimation

A ground truth marking was constructed by applying the simultaneous truth and performance level estimation (STAPLE) algorithm to all six markings for each modality and each tooth. The STAPLE algorithm is an expectation–maximization algorithm that computes a probabilistic estimate of the hidden, implicit ground truth from a set of manual markings [29]. Figure 2 exemplifies the six markings and the ground truth obtained from the STAPLE. The estimated ground truth is then considered as gold standard of the decalcified area and used to evaluate the performance of each rater.

Measurement of agreement

The agreement between the segmentation of a rater and the ground truth is measured by the Dice coefficient (DC) [10] that assesses the spatial overlap of two sets A and B

$$\mathrm{DC}\ (A,B) = \frac{2|A \cap B|}{|A| + |B|}$$

where $|\cdot|$ indicates the number of elements in the sets. The higher the DC, the more the sets overlap; with DC $= 0$ and DC $= 1$ indicating total separation and complete overlap, respectively.

For each manual marking of all the 139 teeth, the DC was calculated as a measure of the rater's performances and used to compute the reproducibility of the markings for both modalities as well as the intrarater and interrater agreements.

Statistical analysis

The present study aims at comparing the reproducibility of the demineralisation detection on digital photographs and QLF images taking into account its spatial location. To this end, descriptive statistics (mean and standard deviation) of the Dice coefficients (DC) for both photographs and QLF was calculated. A three-way repeated measures analysis of variance (ANOVA) was used to compare the means of the DCs for the two modalities. The following three within-subjects factors were taken into consideration: modality (DP and QLF), rater (E, M and H) and cycle (1 and 2).

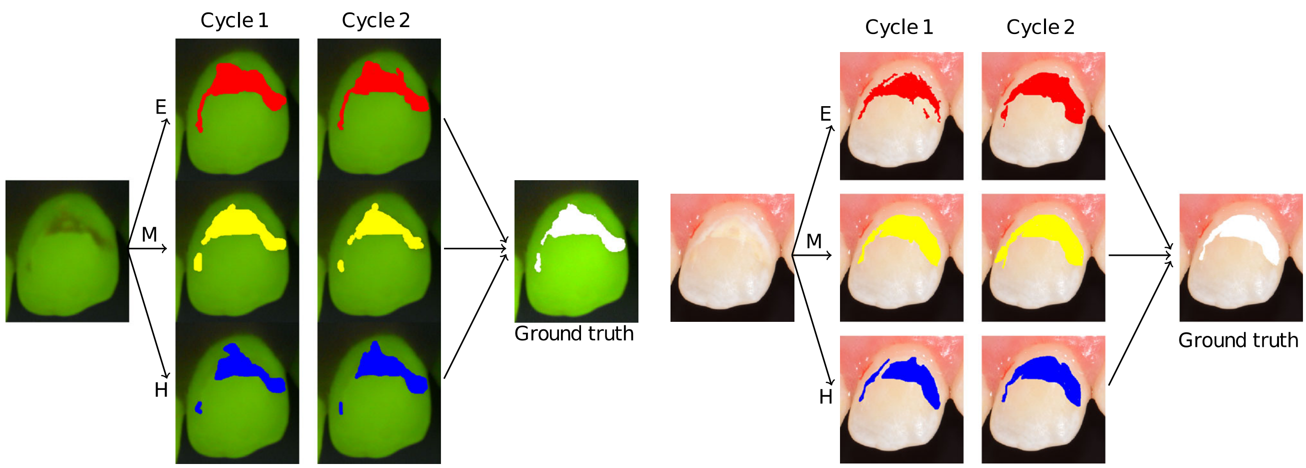


Fig. 2 Markings of the decalcified area from three raters in two cycles and corresponding ground truth computed with STAPLE for a QLF image (*left*) and a digital photo (*right*) of a tooth

Abb. 2 Demineralisationsmarkierungen von 3 Untersuchern aus 2 Zyklen und die zugehörige Ground Truth, berechnet mit STAPLE, für ein QLF-Bild (*links*) und eine digitale Fotografie (*rechts*) eines Zahns

For each modality, the measurement relative to a rater (R) at a cycle (C) is indicated as tuple (R,C). For example, (E,1) indicates the marking of rater E at cycle 1.

Intrarater reliability in both digital photographs and QLF is assessed comparing the means of the DCs. All the tests are performed using IBM SPSS 24 with significance level $\alpha = 0.05$.

Results

Examples of good and poor agreements in the markings for photographs and QLF are depicted in Figs. 3 and 4, respectively.

The mean and the standard deviation (SD) of the 139 Dice coefficients are 0.7964 (range 0.7200–0.8311) for DP and 0.6979 (range 0.6478–0.7390) for QLF (Table 1). Figure 5 clearly shows the overall superiority of DP-based dots and intervals (red) as compared to the QLF-based ones (blue), indicating the better reliability of digital photographs over QLF. The three-way repeated measures ANOVA revealed a statistical significant difference between the modalities ($F(1, 138) = 62.890, p < 0.001$). The cycle differences between the raters are given in Table 2. Intrarater and intercycle variability is rather small as compared to the intermodality variations.

Discussion

Visual inspection and digital photographs are the most common method used in the diagnosis of caries [14]. QLF is a modern, noninvasive method that exploits the fluorescence of the tooth structure to qualitatively and

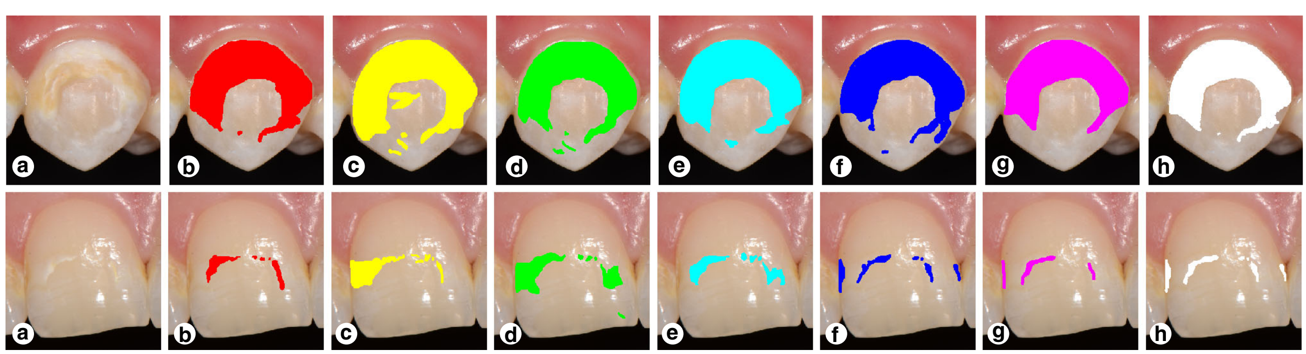


Fig. 3 The second highest (*top*) and the lowest (*bottom*) average Dice coefficients (*DC*) for digital photographs. From *left* to *right*: **a** digital photo, **b** marking by rater E cycle 1, **c** marking by rater E cycle 2, **d** marking by rater M cycle 1, **e** marking by rater M cycle 2, **f** marking by rater H cycle 1, **g** marking by rater H cycle 2 and **h** estimated ground truth. The average DCs in these cases are 0.9483 and 0.5928, respectively

Abb. 3 Zweithöchster (*oben*) und niedrigster (*unten*) mittlerer Dice-Koeffizient (*DC*) der digitalen Photographien. Von *links* nach *rechts*: **a** digitale Fotografie, **b** Markierung von Untersucher E, Zyklus 1, **c** Markierung von Untersucher E, Zyklus 2, **d** Markierung von Untersucher M, Zyklus 1, **e** Markierung von Untersucher M, Zyklus 2, **f** Markierung von Untersucher H, Zyklus 1, **g** Markierung von Untersucher H, Zyklus 2 und **h** geschätzte Ground Truth. Die mittleren Dice-Koeffizienten in diesen Fällen betragen 0,9483 bzw. 0,5928

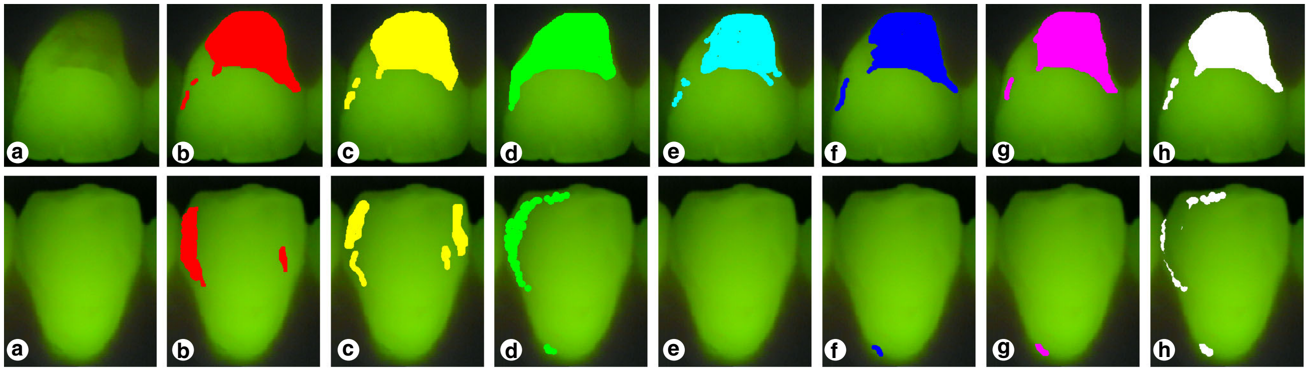


Fig. 4 The highest (*top*) and the second lowest (*bottom*) average DCs for quantitative light-induced fluorescence (QLF). From *left* to *right*: **a** QLF image, **b** marking by rater E cycle 1, **c** marking by rater E cycle 2, **d** marking by rater M cycle 1, **e** marking by rater M cycle 2, **f** marking by rater H cycle 1, **g** marking by rater H cycle 2 and **h** estimated ground truth. The average DCs in these case are 0.9227 and 0.1753, respectively

Abb. 4 Höchster (*oben*) und niedrigster (*unten*) mittlerer Dice-Koeffizienten der QLF Bilder. Von links nach rechts: **a** QLF Bild, **b** Markierung von Untersucher E, Zyklus 1, **c** Markierung von Untersucher E, Zyklus 2, **d** Markierung von Untersucher M, Zyklus 1, **e** Markierung von Untersucher M, Zyklus 2, **f** Markierung von Untersucher H, Zyklus 1, **g** Markierung von Untersucher H, Zyklus 2 und **h** geschätzte Ground Truth. Die mittleren Dice-Koeffizienten in diesen Fällen betragen 0,9227 und 0,1753

Tab. 1 Mean and standard deviation (SD) of the 139 Dice coefficients of each marking and the ground truth for digital photographs (DP) and quantitative light-induced fluorescence (QLF)

Tab. 1 Mittelwert und Standardabweichung (SD) der 139 Dice-Koeffizienten aus jeder Markierung mit der Ground Truth für digitale Fotografien (DP) und quantitative lichtinduzierte Fluoreszenz (QLF)

(Rater, Cycle)	Descriptive statistics			
	DP		QLF	
	Mean	SD	Mean	SD
(E,1)	0.8206	0.0998	0.6856	0.2293
(E,2)	0.8130	0.0975	0.6981	0.2165
(M,1)	0.8194	0.1087	0.7216	0.2132
(M,2)	0.8311	0.1062	0.6478	0.2422
(H,1)	0.7200	0.1553	0.7390	0.2174
(H,2)	0.7745	0.1311	0.6955	0.2308

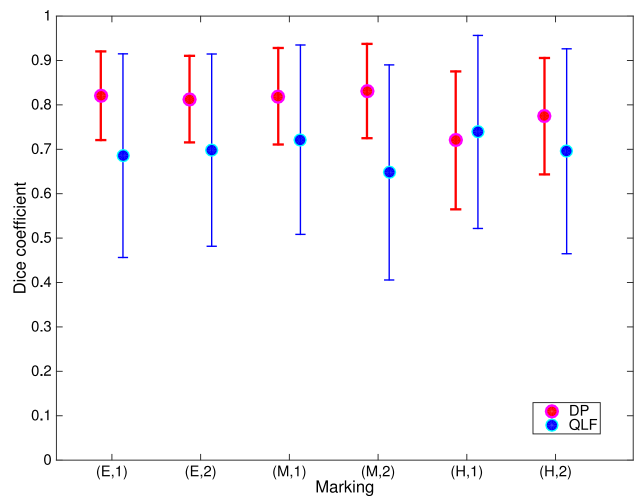


Fig. 5 *Error-bar* plot of the Dice coefficients (DC) for all markings and both modalities. The *dot* represents the mean of the DCs, while the *error bars* indicate the standard deviation

Abb. 5 Fehlerbalkendiagramm der Dice-Koeffizienten (DC) für alle Markierungen und beide Modalitäten. Der Punkt zeigt jeweils das Mittel der Dice-Koeffizienten, während der Fehlerbalken die Standardabweichung wiedergibt

quantitatively assess demineralisation [6]. The fluorescence of the tooth correlates with the mineral content of the tooth structure and changes in the course of demineralisation [17]. Both methods provide accurate diagnostics in terms of intra- and interexaminer reliability [6, 28, 31].

However, reproducibility of demineralisation detection as well as the monitoring of its progression and/or regression plays a crucial role for dentists and orthodontists. This requires accurate and reliable segmentation of the lesion area.

Aim of the present study is to compare the spatial reproducibility of demineralisation represented in photographs and QLF images, where the ground truth was established using the STAPLE algorithm and considered as gold standard for quantitative analysis.

In 2004, Warfield et al. [29] introduced STAPLE which computes the ground truth by estimating an optimal combination of a set of independent manual or automatic segmentations. The obtained ground truth allows to compare the performance of the raters [29].

Meanwhile STAPLE has been established in various medical disciplines like radiology, gynecology or neurology using image-guided diagnoses on, for instance, positron emission tomography (PET) [9], digital cervicographic images (cervigrams) [15] and magnetic resonance (MR) [1], STAPLE has not been applied in dentistry.

Tab. 2 Absolute mean and standard deviation (SD) of the differences in Dice coefficients between the markings at the two for digital photographs (DP) and quantitative light-induced fluorescence (QLF)
Tab. 2 Mittelwert und Standardabweichung (SD) des Betrags der Dice-Koeffizienten aus den Markierungen der 2 Zyklen für digitale Fotografien (DP) und quantitative lichtinduzierte Fluoreszenz (QLF)

Pair	DP		QLF	
	Mean	SD	Mean	SD
(E,1)–(E,2)	0.0075	0.1048	0.0123	0.1415
(M,1)–(M,2)	0.0117	0.0815	0.0738	0.2024
(H,1)–(H,2)	0.0544	0.1601	0.0435	0.1762

Due to the methodical differences, the presented results are hardly comparable to previous studies on the topic. In our study, exclusively affected teeth were recorded, while Heinrich-Weltzien et al. [18] scored presence or absence of initial caries lesions using both visual inspection and QLF. A total of 87.2% of all smooth surfaces were scored as sound or initially carious when assessed by visual inspection and QLF in combination: 4.9% were detected by visual inspection alone and 7.9% by QLF alone. Cochrane et al. [8] recoded that digital photographs and QLF provided data that is correlated moderately with transverse microradiography (TMR) as the gold standard. The authors concluded that QLF was a preferable technique for assessing mineral content changes when compared with photographs due to the absence of glare in the QLF images, the quantification software and the lower variability.

In previous studies [6, 26, 28], the intra- and interrater reliability of photographs and QLF was assessed comparing the area (in mm^2) of the white spot lesions individuated by the raters and statistically analyzed e.g. using the intraclass correlation coefficient (ICC) [22, 27]. However, comparing the area exclusively takes into account the size of the lesion, but disregards the lesion's location and shape. Contrarily, we have compared manual segmentations with STAPLE-generated ground truth. The performances of the raters have been assessed quantitatively by the spatial overlap of the markings and the ground truth using the Dice coefficient. Taking into account spatial correspondences makes a vast difference and gains new insights about the modalities used.

The ICC is affected by data variability. Disregarding other factors, low values of between-subjects variability lower the value of the ICC even though the scores of the subjects present small differences [30]. Therefore, it has only values when applied to the specific population. Furthermore, the ICC is only applicable to values that are not inherently quantifying the quality. For instance, the ICC applied to the lesion volume marked by experts can be used to assess the interrater reliability in determining the lesion volume, since the lesion volume itself does not indicate the reliability. Contrarily, the DC itself quantifies the rater reliability. Thus, the ICC is inapplicable to the DCs.

The results of the present study based on descriptive statistics of the DCs indicate that the detection of demineralised areas in digital photographs is more reproducible than the detection of damaged areas in QLF images. The DC means of the three raters for the photographs are higher than those for the QLF (Table 1), indicating that the raters performed better in marking a decalcified area in digital photos of the teeth than in identifying it in QLF images. The standard deviation of the DC for the photographs are lower than those for the QLF (Fig. 5), suggesting that the QLF measurement were more variable than the digital photographs measurement and that the ratings were then more consistent for photographs than for QLF.

Demineralisation diagnosed based on digital photographs and QLF are two valuable methods providing information for detection, monitoring of its progression and/or regression. While our study has assessed digital photographs versus QLF, combining both modalities might be superior to each of it individually. This will be the focus of future research.

Conclusions

The assessment of demineralisation in digital photographs is more accurate than in QLF imaging with respect to the reproducibility of lesion's positions and shapes. Hence, visual diagnosis certainly remains a good practice in white spot lesion detection, due to its simple and quick applicability for trained clinicians.

Acknowledgements The authors would like to thank Andras Keszei, Uniklinik RWTH Aachen, Germany, for his guidance in the statistical analysis. The authors at AICES RWTH Aachen University were funded in part by the Excellence Initiative of the German Federal and State Governments.

Compliance with ethical standards

Conflict of interest The authors declare that they do not have any conflict of interest.

Informed consent Informed consent was obtained from all individuals participating in the trial.

References

1. Akhondi-Asl A, Warfield SK (2012) Estimation of the prior distribution of ground truth in the STAPLE algorithm: an empirical Bayesan approach. Med Image Comput Comput Assist Interv 15:593–600
2. Alammari MR, Smith PW, de Josselin de Jong E, Higham SM (2003) Quantitative light-induced fluorescence (QLF): a tool for

early occlusal dental caries detection and support decision making in vivo. J Dent 41:127–132

3. Angmar-Månsson B, Al-Khateeb S, Tranaeus S (1996) Monitoring the caries process. Optical methods for clinical diagnosis and quantification of enamel caries. Eur J Oral Sci 104(4):480–485
4. Benson P, Pender N, Higham S, Edgar WM (1998) Morphometric assessment of enamel demineralisation from photographs. J Dent 26(8):669–677
5. Benson P, Pender N, Higham S (2000) Enamel demineralisation assessed by computerised image analysis of clinical photographs. J Dent 28(5):319–326
6. Benson P, Pender N, Higham S (2003) Quantifying enamel demineralization from teeth with orthodontic brackets—a comparison of two methods. Part 1: repeatability and agreement. Eur J Orthod 25(2):149–158
7. Cardoso JR, Pereira LM, Iversen MD, Ramos AL (2014) What is gold standard and what is ground truth? Dental Press J Orthod 19(5):27–30
8. Cochrane NJ, Walker GD, Manton DJ, Reynolds EC (2012) Comparison of quantitative light-induced fluorescence, digital photography and transverse microradiography for quantification of enamel remineralization. Aust Dent J 57(3):271–276
9. Dewalle-Vignion AS, Betrouni N, Baillet C, Vermandel M (2015) Is STAPLE algorithm confident to assess segmentation methods in PET imaging? Phys Med Biol 60(24):9473–9491
10. Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26(3):297–302. http://www.jstor.org/stable/1932409
11. Enaia M, Bock N, Ruf S (2011) White-spot lesions during multibracket appliance treatment: a challenge for clinical excellence. Am J Orthod Dentofac Orthop 140(1):e17–e24
12. Ferreira Zandoná A, Santiago E, Eckert G, Fontana M, Ando M, Zero DT (2010) Use of ICDAS combined with quantitative light-induced fluorescence as a caries detection method. Caries Res 44(3):317–322
13. Gmur R, Giertsen E, van der Veen M, de Josselin de Jong E, ten Cate J, Guggenheim B (2006) In vitro quantitative light-induced fluorescence to measure changes in enamel mineralization. Clin Oral Investig 10(3):187–195
14. Gomez J (2015) Detection and diagnosis of the early caries lesion. BMC Oral Health 15(Suppl 1):S3
15. Gordon S, Lotenberg S, Long R, Antani S, Jeronimo J, Greenspan H (2009) Evaluation of uterine cervix segmentations using ground truth from multiple experts. Comput Med Imaging Graph 33(3):205–216
16. Gorelick L, Geiger A, Gwinnett A (1982) Incidence of white spot formation after bonding and banding. Am J Orthod 81(2):93–98
17. Heinrich-Weltzien R, Kühnisch J, van der Veen M, de Josselin de Jong E, Stösser L (2003) Quantitative light-induced fluorescence (QLF)—a potential method for the dental practitioner. Quintessence Int 34(3):181–188
18. Heinrich-Weltzien R, Kühnisch J, Ifland A, Tranaeus S, Angmar-Månsson B, Stösser L (2005) Detection of initial caries lesions on smooth surfaces by quantitative light-induced fluorescence and visual examination: an in vivo comparison. Eur J Oral Sci 113(6):494–498
19. Klukowska M, Bader A, Erbe C, Bellamy P, White DJ, Anastasia MK, Wehrbein H (2011) Plaque levels of patients with fixed orthodontic appliances measured by digital plaque image analysis. Am J Orthod Dentofac Orthop 139(5):e463–e470
20. Knösel M, Attin R, Becker K, Attin T (2007) External bleaching effect on the color and luminosity of inactive white-spot lesions after orthodontic appliances. Angle Orthod 77(4):646–652
21. Kühnisch J, Heinrich-Weltzien R (2004) Quantitative light-induced fluorescence (QLF)—a literature review. Int J Comput Dent 7(4):325–338
22. McGraw KO, Wong SP (1996) Forming inferences about some intraclass correlation coefficients. Psychol Methods 1:30–46
23. Mizrahi E (1982) Enamel demineralization following orthodontic treatment. Am J Orthod 82(1):62–67
24. Ousehal L, Lazrak L, Es-Said R, Hamdoune H, Elquars F, Khadija A (2011) Evaluation of dental plaque control in patients wearing fixed orthodontic appliances: a clinical study. Int Orthod 9(1):140–155
25. Palamara J, Phakey P, Rachinger W, Orams H (1986) Ultrastructure of the intact surface zone of white spot and brown spot carious lesions in human enamel. J Oral Pathol 15(1):28–35
26. Pretty IA, Hall AF, Smith PW, Edgar WM, Higham SM (2002) The intra- and inter-examiner reliability of quantitative light-induced fluorescence (QLF) analyses. Br Dent J 193(2):105–109
27. Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. Psychol Bull 36:420–428
28. Tranaeus S, Shi XQ, Lindgren LE, Trollas K, Angmar-Månsson B (2002) In vivo repeatability and reproducibility of the quantitative light-induced fluorescence method. Caries Res 36(1):3–9
29. Warfield SK, Zou KH, Wells WM (2004) Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging 23(7):903–921
30. Weir JP (2005) Quantifying test–retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 19(1):231–240
31. Willmot DR, Benson PE, Pender N, Brook AH (2000) Reproducibility of quantitative measurement of white enamel demineralisation by image analysis. Caries Res 34(2):175–181