A Statistical Modeling Approach to Computer-Aided Quantification of Dental Biofilm

Awais Mansoor, *Member, IEEE,* Valery Patsekin, Dale Scherl, J. Paul Robinson, *Member, IEEE,*
and Bartlomiej Rajwa

Abstract—**Biofilm is a formation of microbial material on tooth substrata. Several methods to quantify dental biofilm coverage have recently been reported in the literature, but at best they provide a semi-automated approach to quantification with significant input from a human grader that comes with the grader's bias of what is foreground, background, biofilm, and tooth. Additionally, human assessment indices limit the resolution of the quantification scale; most commercial scales use five levels of quantification for biofilm coverage (0%, 25%, 50%, 75%, and 100%). On the other hand, current state-of-the-art techniques in automatic plaque quantification fail to make their way into practical applications owing to their inability to incorporate human input to handle misclassifications. This paper proposes a new interactive method for biofilm quantification in Quantitative light-induced fluorescence (QLF) images of canine teeth that is independent of the perceptual bias of the grader. The method partitions a QLF image into segments of uniform texture and intensity called *superpixels*; every *superpixel* is statistically modeled as a realization of a single 2D Gaussian Markov random field (GMRF) whose parameters are estimated; the superpixel is then assigned to one of three classes (*background, biofilm, tooth substratum*) based on the training set of data. The quantification results show a high degree of consistency and precision. At the same time, the proposed method gives pathologists full control to post-process the automatic quantification by flipping misclassified superpixels to a different state (background, tooth, biofilm) with a single click, providing greater usability than simply marking the boundaries of biofilm and tooth as done by current state-of-the-art methods.**

Index Terms—**Biofilm, dental plaque, quantitative fluorescence, superpixelization, Gaussian Markov random field.**

I. INTRODUCTION

THE assessment of dental biofilm and dental plaque coverage is a crucial step in many clinical as well as basic biological science applications. For instance, clinicians diagnosing potential gingival conditions or evaluating dental prostheses or dental implants require an accurate estimation of bacterial biofilm coverage. Furthermore, researchers involved in studies of periodontal diseases and periodontal therapies as well as those assessing oral hygiene products often employ information about biofilm depth, area, and distribution.

Awais Mansoor is with the Department of Radiology and Imaging Sciences, National Institutes of Health (NIH). Prior to joining NIH, he was with the Department of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47906 USA. E-mail: awais.mansoor@gmail.com.

Valery Patsekin, and Bartlomiej Rajwa are with Bindley Bioscience Center, Purdue University, West Lafayette, IN 47906 USA. Tel: +1 765 494-0757, Fax: +1 765 494-0517, E-mail: {brajwa}@purdue.edu

J. Paul Robinson is with Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN 47906 USA. Tel: +1 765 494-0757, Fax: +1 765 494-0517. E-mail: wombat@purdue.edu.

Dale Scherl is with Hill's Pet Nutrition. Topeka, KS, USA.

Dental-plaque images have traditionally been collected using reflected light. The quantitative evaluation is then performed manually by outlining the plaque areas using tracing paper, and subsequently quantifying the results by assigning a dental score [1]. In the past, a number of manual dental-plaque scoring indices have been proposed [2], [3], [4], [5], [6]. For instance, the plaque index used by Quigley and Hein, and modified by Turesky, is one of the most frequently employed to grade images. Others in general use are the Ramfjord index, the Turesky index, and the O'Leary index [7].

During the past decade, computer assisted technology have gained a lot of interest in almost every aspect of biology and medicine from accessibility [8] to quantitative measurements [9], [10], [11] to qualitative adjustments [12]. With the introduction of digital imaging, the idea of computer-aided quantification in dentistry and dental research began to gain popularity. However, early computer-aided techniques presented in the literature failed to offer practical benefits to researchers, as almost all reported computer-aided quantification techniques relied on manual segmentation performed by a grader using a software tool. These methods employing commercial photo-editing software are tedious, are time consuming, and do not provide a tangible advantage over manual grading. For instance, in [3], [4] an approximate region of interest (ROI) in a tooth imaged under reflected light is manually selected using Photoshop (Adobe, San Jose, CA) tools; a mean-shift segmentation algorithm is then employed to estimate the percentage of tooth surface covered by dental plaque. Smith et al. [5] use a *pen* and *make path* tools of Photoshop. Subsequently ImagePro Plus (Media Cybernetics, Silver Spring, MD) was used for calibration and calculation of the percentage of total tooth area covered by plaque.

Partly owing to the low quality and low contrast of reflected-light images, the early attempts at computer-aided dental diagnostics did not prove successful. A major breakthrough in automated dental-plaque quantification methods occurred with the introduction to dentistry of fluorescence dyes that stain the bacterial biofilm, providing contrast between the region of interest and background. UV-excited fluorescein produces a yellow plaque signal, which contrasts with the blue tooth enamel signal and with the gingiva, which appears black in images. This makes fluorescein an excellent candidate for automatic quantification of biofilm, and methods based on fluorescein imaging have been proposed in the literature. For instance, Sagel et al. [6] employ UV fluorescence imaging and automatically classify every pixel inside the image by calculating the scalar distance in RGB color space from the

pixel to the median RGB color of each predefined class [6]. The pixel is then assigned to the closest group.

Quantitative light-induced fluorescence (QLF) is yet another dental diagnostic tool for quantitative assessment of caries, plaque, bacterial activity, and staining [13]. The QLF methods are based on the autofluorescence of teeth and plaque-forming bacterial biofilm. The QLF devices marketed by Inspektor Dental Care BV (Amsterdam, The Netherlands) have recently generated considerable interest in the dental research community [14]. The technique uses a small camera that can be easily hand held; the images are free from flash-light specular reflections and major distortions. Additionally, the QLF technique enables very small changes in plaque to be detected, thus increasing accuracy [7]. Going beyond mere detection, QLF can monitor the development or regression of caries and plaque over time. When teeth are excited with blue light, the enamel emits green luminescence, the dental plaque red. This red/orange fluorescence is attributed to metabolic products (mostly porphyrins) from some of the resident bacteria. The intensity of autofluorescence is expected to be proportional to the biofilm depth (Fig. 1). A recent study by Pretty et. al [15] confirmed the potential of digital imaging systems employing polarized white light and QLF for software analysis and assessment by raters for epidemiological work. Owing to these, QLF imaging has of late become a popular modality for plaque quantification, and a number of QLF image–based quantification methods have been proposed [16], [17], [18].
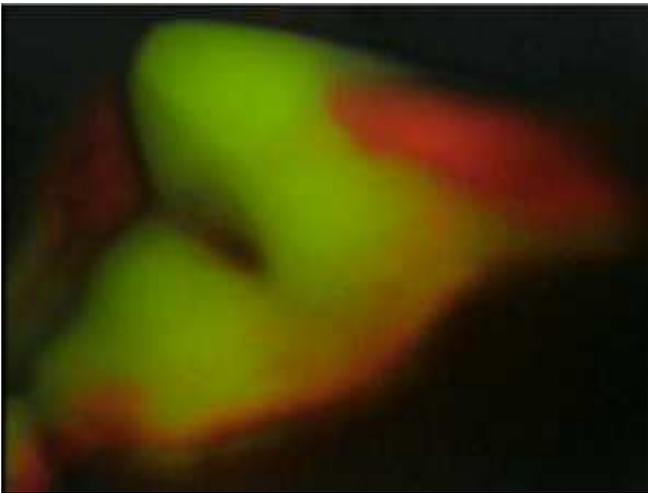

Fig. 1.   Disclosed plaque imaged using QLF instrument.

With the advent of fluorescence-imaging techniques in dentistry several attempts have been made to automate the quantification process, and a number of statistical and machine-learning approaches have been tested for the purpose of automation in an attempt to minimize the subjectivity of manual scoring. For instance, [19] used a cellular neural network to search for the threshold between biofilm and clean areas of the tooth; [3], [20], [21] quantified plaque by using *fuzzy c-means* clustering. However, the implementation of functional and robust automated segmentation in the context of dental images turned out to be an unexpectedly difficult problem. The demarcation of a tooth of interest from neighboring teeth

fluorescing almost identically, as well as from gum tissue, still needs to be investigated further, and current state-of-the-art mathematical models fail to accurately mimic human expert assessments. The problem is significantly worse in veterinary dental imaging owing to the lower image quality caused by difficulties associated with animal compliance.

Ambitious attempts to replace image graders or qualified pathologists with computer systems usually fail owing to lack of algorithm robustness, but also because of tradition and praxis. On the other hand, digital image-analysis platforms designed for general use do not provide practitioners with tools that rely on and supplement their expertise. Consequently, current state-of-the-art automatic plaque-quantification approaches have not gained much popularity owing to the fact that they allow only the choice between accepting or rejecting the automated quantification; even in cases where the result is off by only a little, the user typically has no option of correcting the misclassification but must reject it altogether and resort to a painstaking method of manual demarcation.

The method outlined in this paper employs a combination of manual and automated segmentation, providing what can be considered essentially a computer-aided segmentation and quantification. In contrast to a completely automated quantification approach, the described method provides an initial estimation of the background, the teeth, and the plaque-covered area based on the prior model of expert classification embedded into the method. This first approximation that happens to be very accurate for most cases, can be subsequently modified by a grader or dental research practitioner in an interactive manner if deemed necessary. Additionally, the interaction does not require tedious manual re-segmentation. Instead, the operator points to entire computer-suggested regions to assign them to one of the pre-defined categories (biofilm, tooth substrata, background). This approach allows for much faster and more streamlined interaction between operator and computer system. The availability of the initial assessments provided by the algorithm allows for rapid quantification with a minimal number of significant corrections. The described technique has been practically tested and validated for quantification of dental film on canine teeth.

The approach presented in this paper, which combines automated and interactive functionality, represents a broader solution to image-analysis problems in clinical and research pathology, and is especially applicable to challenging, difficult-to-segment biomedical images. The described technique is also amenable to further extension employing machine-learning methods.

The basic principle of the utilized algorithm relies on its ability to divide images into statistically and texturally uniform regions called superpixels, which can be subsequently compared using a measure of dissimilarity [22]. The block diagram of the algorithm is shown in Fig. 2. The method works in two stages: (i) *initial segmentation*, and (ii) *interactive correction*. During the initial segmentation stage the superpixelization process creates a set of statistically uniform regions, which are classified into three categories (background, clean area (tooth substratum), and biofilm-covered area) using statistical texture estimation based on GMRF. The initial guess,

imitating the expert quantification, is based on a calibration that employs a number of manually scored clean and covered areas from various teeth. Owing to the calibration this initial segmentation process offers reasonably high concordance with manual grading. However, in the interactive corrections stage, the segmentation can be further modified by operators by pointing to a questionable region and clicking the mouse. This methodology offers much higher efficiency than a requirement for manual drawing of outlines or a requirement to redraw contested areas. Moreover, as shown later in the results section, the correction needs to be applied only to one image in a series, the subsequent images may be re-segmented automatically using the corrected mask.
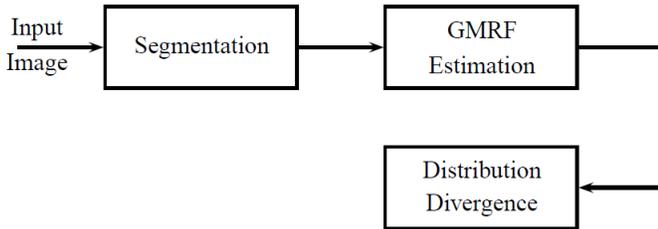

Fig. 2. A block diagram summarizing the proposed algorithm for quantification of dental biofilm on canine teeth.

II. BACKGROUND

As mentioned in the introduction, general-use photo-editing packages fail to provide intuitive assistance to biomedical image analysts such as pathologists or dental image graders. The published procedures often need a precise outline to be drawn around tooth and/or biofilm. However, with the advancement of human-computer interaction techniques, it becomes clear that creating a link between human grader experience and automated system should lead to improvement in computer-aided dental diagnostics. The method proposed here presents a tool that quantifies biofilm using an interactive segmentation interface that allows practitioners to quickly refine the initial assessment suggested by the algorithm.

The principle of the proposed automatic quantification method is based on the fact that a uniform region inside an image can be modeled via a Gaussian Markov random field (GMRF), a statistical image-modeling technique that has long been used for statistical image segmentation [23] and classification [24]. The statistical modeling serves two purposes in terms of dental biofilm quantification: it will help isolate the tooth inside the image and it will segment the plaque region inside the tooth. The first step is to segment the image into sub-regions uniform enough that they can be modeled using a single GMRF; therefore we start with a brief introduction to the concepts of the oversegmentation method proposed by Malik et al. [25], [26].

A. Oversegmentation (Superpixelization)

The idea of oversegmentation by dividing an image into what is known in the literature as *superpixels* was proposed by Ren and Malik [27]. Their motivation was to come up with

a unit of *visual* representation having uniform textural and statistical properties, since digital pixels are units of *digital* representation. Since the inception, the idea has been used in various applications of image processing [28], [29], [30], machine learning [31], [32], and computer vision [33]. Our algorithm uses this method as a preprocessing step to GMRF estimation. The superpixels are texturally and statistically more meaningful than regular pixels in that each superpixel forms a perceptually consistent unit, i.e., all pixels inside a superpixel are assumed to be consistent in their statistical properties and texture [25], [29], [34]. In our method, the superpixel segmentation map not only helps segregate biofilm from the tooth subtrata, but also eliminates the need for any independent background subtraction procedure.

The *graph cut* method for superpixelization [25], [34] uses the graph-theory concept of iterative binary partitioning: each pixel is modeled as a vertex of the graph, and a weighted edge between two pixels shows the degree of similarity (or dissimilarity) between them. Within this model, the image-segmentation problem is reduced to partitioning a connected graph into several disjoint subgraphs called *graph cuts*, and the nodes of the graphs are the entries that are partitioned. In the case of images, the pixels form the nodes of the graph, and the edges between neighboring nodes correspond to the *strength* with which two pixels belong to the same segment. The criterion for partitioning is to minimize the weights of the connections within each group and maximize the weights among different groups. The input arguments required by the algorithm to produce an image segmentation map are the image and the number of groups (*superpixels*). Fig. 3 shows images segmented into 200 and 1000 superpixels. The typical number of superpixels per image depends upon the size of the image and the textural details inside the image and determined empirically by the user, we have not found any evidence of the dependence of the number of superpixels on patients in our experiments. The selection to the number of superpixels need to be made once only for a typical dataset. If the superpixels are too small, the second step of manually relabeling them takes a lot of time, on the contrary if they are too large the precision of the estimated parameters of the random field in the next stage is compromised. Empirically in our experiments and as shown later in Fig. 3, for QLF images of size 256×256, partition to 200 superpixels is found sufficient for dental quantification purposes.

B. GMRF Modeling of QLF Images

After obtaining superpixels for an image, the proposed method models every superpixel as a realization of GMRF. GMRF model is used to capture the statistical properties of the neighborhood around a pixel. A 2D Gaussian random field (GRF) is a random vector following a multivariate normal distribution; putting an additional constraint of conditional independence makes it *Markovian*, which essentially means that the statistics of a pixel inside an image are independent of those of the pixels outside a predefined neighborhood, hence the term Gauss Markov random field. Mathematically, for two

indices i and j inside the images,

$$x_i \perp x_j | \mathbf{x}_{-ij}$$

where x_i and x_j are conditionally independent of each other given the neighborhood $\mathbf{x}_{-ij}$, ($\mathbf{x}_{-ij}$ denotes all the elements inside the neighborhood but not including i and j). Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph where $\mathcal{V} = 1, 2, \ldots$ are vertices of the graph and $\mathcal{E} = (i, j)$ are the edge pair of the graph; then a random vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ is called the GMRF with respect to the graph $\mathcal{G}$ with mean μ and covariance matrix Σ if its density has the form

$$\Pi(\mathbf{x}) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)\right)$$
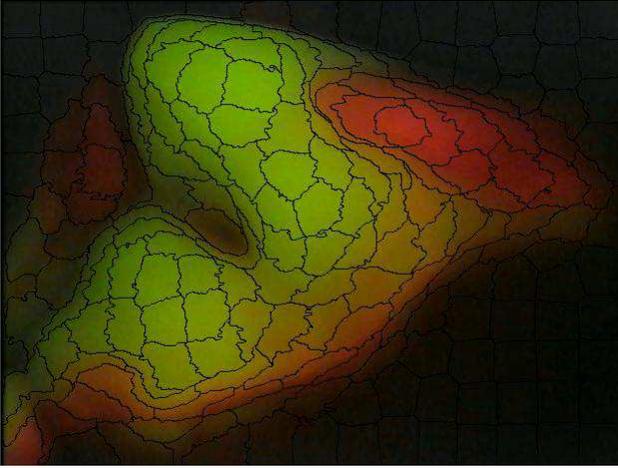
(a)
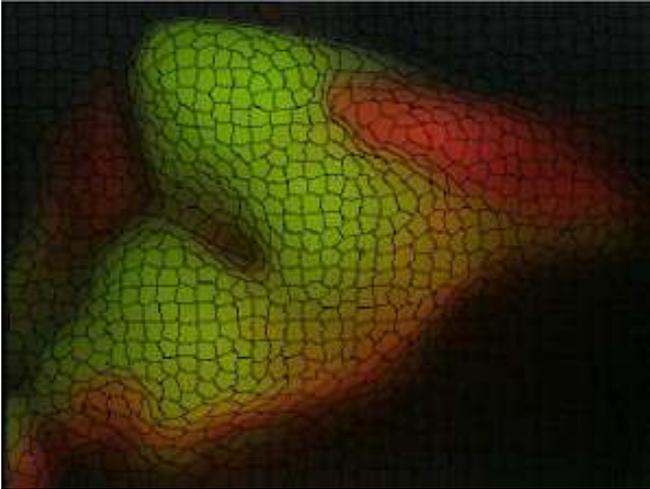
(b)

Fig. 3. An example of superpixels obtained through the *normalized cut* algorithm [25]. (a) is segmented into 200 superpixels; (b) is segmented into 1000 superpixels.

C. Parameter Estimation of GMRF

The modeling of 2D homogeneous texture is connected to a 2D symmetric autoregressive process driven by a white noise process $E_{\mathbf{s}}$, $\mathbf{s} \in \mathbb{R}^2$ [35], [36]. For any random field, a 2D symmetric autoregressive process driven by a white noise can be written as

$$A_{\mathbf{s}} = B_{\mathbf{s}}\theta_{\mathbf{s}} + E_{\mathbf{s}} \qquad (1)$$

where $\theta_{\mathbf{s}}$ are symmetric autoregressive parameters calculated over a predefined neighborhood, and $A_{\mathbf{s}}$ is a discrete-valued random field defined on a 2D lattice. $B_{\mathbf{s}}$ is the neighborhood matrix such that every element $b_{\mathbf{s}} = [a_{\mathbf{s}+\mathbf{r}} + a_{\mathbf{s}-\mathbf{r},\mathbf{r}}]$, η is a predefined neighborhood of radius $\mathbf{r}$. Owing to the Markovian assumption and the superpixel segmentation the parameters of the random field are independent of the pixels outside the predefined neighborhood η. To estimate the parameters of the Gauss-Markov model we used a least-squares (LS) approach on a predefined neighborhood $A_{\mathbf{s}}, \mathbf{s} \in \eta$. The least-squares minimization to (1) with respect to $\theta_{\mathbf{s}}$ is

$$\theta_{\mathbf{s}} = \underset{\theta_{\mathbf{s}}}{\arg\min} E_{\mathbf{s}}^T E_{\mathbf{s}} = \left(B_{\mathbf{s}}^T B_{\mathbf{s}}\right)^{-1} B_{\mathbf{s}}^T A_{\mathbf{s}}' \qquad (2)$$

where $A_{\mathbf{s}}' = A_{\mathbf{s}} - \hat{\mu}_{\mathbf{s}}$ and $\hat{\mu}_{\mathbf{s}} = (N \times M)^{-1} \sum_{\mathbf{s} \in \omega_{\mathbf{r}}} a_{\mathbf{r}}$, $N \times M$ are the dimensions of the neighborhood η (one can use arbitrary dimensions but, in practice, $M = N$). The variance estimate $\hat{\sigma}_{\mathbf{s}}^2$ can be calculated in a similar manner:

$$\hat{\sigma}_{\mathbf{s}}^2 = \frac{1}{NM} \sum_{\mathbf{s} \in \eta_{\mathbf{r}}} \left(a_{\mathbf{s}} - \mathbf{b}_{\mathbf{s}}^T \hat{\theta}_{\mathbf{s}}\right)^2$$

$$\hat{\Sigma}_{\mathbf{s}} = \frac{1}{NM} \left(A_{\mathbf{s}} - B_{\mathbf{s}}\hat{\theta}_{\mathbf{s}}\right)^T \left(A_{\mathbf{s}} - B_{\mathbf{s}}\hat{\theta}_{\mathbf{s}}\right) \qquad (3)$$

An inscribed bounding-box is used inside every superpixel for texture estimation. The least-squares estimate provides a feature vector $f_{\mathbf{s}} = \left[\theta_{\mathbf{s}}, \hat{\Sigma}_{\mathbf{s}}, \hat{\mu}_{\mathbf{s}}\right]$. It is important to note here that the method assumes homogeneity of texture inside a predefined neighborhood, achieved through superpixelization.

D. KL Divergence

After estimating the parameters of the GMRF, we use Kullback-Leibler (KL) divergence to express the degree of similarity between two probability distributions. Let $\mathcal{C}_{1\Lambda}(n)$ and $\mathcal{C}_{2\Lambda}(n)$ be the sets 2D random-field estimates inside two channels $\mathcal{C}_1$ and $\mathcal{C}_2$ respectively; then the KL divergence between them is defined as

$$KL(\mathcal{C}_{1\Lambda} || \mathcal{C}_{2\Lambda}) = \sum_{n=1}^{\mathcal{N}} \mathcal{C}_{1\Lambda}(n) \log \frac{\mathcal{C}_{1\Lambda}(n)}{\mathcal{C}_{2\Lambda}(n)} \qquad (4)$$

where $n \in \mathcal{N}$ is the superpixel index, $\mathcal{N}$ is the total number of superpixels.

Since KL divergence is not symmetric, we make Eq. (4) symmetric by calculating both $KL(\mathcal{C}_{1\Lambda} || \mathcal{C}_{2\Lambda})$ and $KL(\mathcal{C}_{2\Lambda} || \mathcal{C}_{1\Lambda})$. The symmetric variant of the KL_S is defined as

$$KL_S(\mathcal{C}_{1\Lambda}, \mathcal{C}_{2\Lambda}) = KL(\mathcal{C}_{1\Lambda} || \mathcal{C}_{2\Lambda}) + KL(\mathcal{C}_{2\Lambda} || \mathcal{C}_{1\Lambda}) \qquad (5)$$

From [37], the KL divergence for multivariate normal densities $\mathcal{C}_{1\Lambda}(\mu_{\mathcal{C}1}, \Sigma_{\mathcal{C}1})$ and $\mathcal{C}_{2\Lambda}(\mu_{\mathcal{C}2}, \Sigma_{\mathcal{C}2})$ forming a Markov random field (i.e. GMRF) is

$$KL(\mathcal{C}_{1\Lambda} || \mathcal{C}_{2\Lambda}) = 0.5 \log \frac{|\Sigma_{\mathcal{C}1}|}{|\Sigma_{\mathcal{C}2}|} + 0.5 Tr\left(\Sigma_{\mathcal{C}1}^{-1} \Sigma_{\mathcal{C}2}\right) \qquad (6)$$

$$+ 0.5(\mu_{\mathcal{C}1} - \mu_{\mathcal{C}2})^T \Sigma_{\mathcal{C}1}^{-1}(\mu_{\mathcal{C}2} - \mu_{\mathcal{C}1}) - \frac{d}{2}$$

where the operator $|.|$ denotes the determinant of a matrix and d is the dimension of the covariance matrix.

III. AUTOMATIC QUANTIFICATION ALGORITHM FOR BIOFILM COVERAGE

In the previous section we presented the theoretical basis for our algorithm for quantification of biofilm on canine teeth. This algorithm is summarized in the block diagram shown in Fig. 4. The algorithm has two stages: the *calibration* stage and the *quantification* stage. The proposed method begins by converting the RGB image into HSI (*hue, saturation, intensity*) domain and computing superpixels (oversegmentation) inside the *intensity* channel of a QLF image. The superpixel map obtained is mapped over the green channel, followed by the estimation of statistical parameters of every superpixel individually in both channels. The idea is to assume every superpixel to be a realization of a unique GMRF, estimate the parameters of the random field in two channels, and compare their statistics using a divergence measure. As explained before, in QLF images the biofilm areas of the tooth are captured primarily in the red channel and the clean areas (tooth substratum) in green. However, the tooth substratum masked by biofilm does not register inside the green channel; therefore, we used the intensity channel and the green channel for random-field estimation and divergence estimation. The task is the isolation of three classes inside the image: biofilm, tooth substratum, and background. This is challenging owing to the presence of multiple neighboring teeth in the background, ambient illumination, and imaging artifacts near the gums. Most biofilm quantification methods proposed in the literature require background subtraction as a preprocessing step. However, our method does not need any independent background subtraction technique to isolate a tooth from the image, owing to the statistical modeling approach presented in this paper that can separate background, tooth substrata, and the biofilm based solely on the parametric differences of their estimated Markov random field model. The algorithm takes a tree-like approach to classify the image areas as *biofilm*, *tooth substratum*, and *background* by first separating the *background* from the target dental field, and then segmenting the *tooth substratum* from the *biofilm*. The extent of divergence between superpixels from the green channel and the total intensity channel will determine if a particular superpixel belongs to biofilm or not. The absence of biofilm corresponds to both channels being identical in terms of intensity and texture, and therefore to a low statistical divergence value. In the presence of biofilm, superpixels will produce a high statistical divergence. The KL divergence values for the background and the clean tooth substratum will be close, so we added another feature for classification: the mean intensity of the superpixel. After classifying biofilm superpixels, the algorithm performs binary classification for the rest of the superpixels based on two features: KL divergence and mean illumination intensity. Fig. 5 shows a single tooth isolated by superpixelization and the tooth-isolation procedure of our algorithm. To account for multiple teeth inside an image, it is further assumed that the tooth of interest lies close to the geometric center of the image.

A. Calibration Stage

Expert opinions are usually used as the ground-truth in medical and biological imaging, the calibration stage in our method is a way of incorporating the expert-knowledge into the system. Once the expert determines the boundary between the biofilm and the tooth substrata; our method by extracting the statistical differences of the two class, imitates the human expert for images acquired in similar settings. The block diagram in Fig. 4 outlines the calibration procedure. The aim of the calibration stage is to establish the KL-divergence threshold between tooth substratum and biofilm. To define clean areas we utilize images which are deemed clean by two expert graders. The experts used in our experiments for manual scoring possess several years of industrial experience in acquisition, calibration, and evaluation of QLF images.

After the *tooth-isolation* step that separates tooth from background based on two features, i.e. KL-divergence, and the mean intensity of superpixel, the calibration sets the KL threshold equal to the 95th percentile of KL divergences among superpixels inside the clean teeth. The percentile is chosen empirically to accommodate human bias and hardware and capturing related issues such as inappropriate lighting. Since the KL divergences are calculated between the green channel and the intensity channel, the KL-divergence values increase with the increase in the red hue inside the superpixels. It is up to the human expert to decide how low the red signal level must be in order to declare the substratum region clean. The ground truth from the expert segmentations are created using the STAPLE algorithm [38].

B. Quantification Stage

The *quantification stage* follows the same steps as the *calibration stage*. After the *tooth-isolation* procedure, the quantification stage compares the KL divergence of isolated-tooth superpixels with the KL divergence threshold of tooth substratum and plaque calculated in the calibration stage. The *biofilm quantification index* (BQI) defines the percent area of tooth covered with plaque; mathematically

$$BQI = \frac{\sum_{t \in plaque} |I_{sp}(t)|}{\sum_{t \in tooth} |I_{sp}(t)|} \tag{7}$$

where the $|.|$ operator gives the area of an individual superpixel.

IV. IMAGING PLATFORM

Images of canine teeth were captured using a quantitative light-induced fluorescence (QLF) system from Inspektor Research Systems. This system is equipped with a Philips MPXL RP50-2P xenon arc lamp, Hoya HA30 infrared band-pass filter, Hoya B370 blue bandwidth filter, Lumatec S380 liquid light guide, Hoya Y52 yellow high-pass filter, Sony DXC-LS1P CCD camera system, Integral Technologies FlashPoint 3DX Lite PCI framegrabber, and Inspektor Research Systems software version 2.0.38. Image capture was perform on conscious dogs, with the room lights dimmed in order to reduce

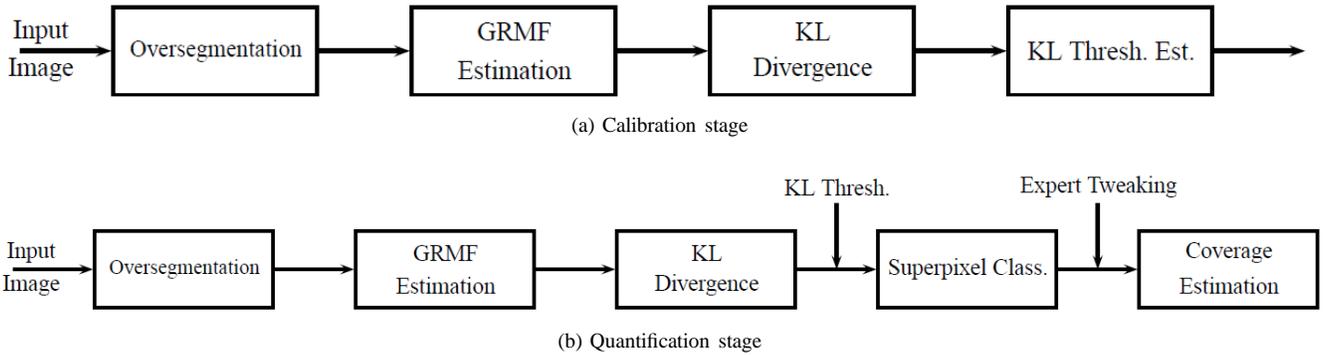(a) Calibration stage

(b) Quantification stage

Fig. 4. A block diagram summarizing the proposed algorithm for quantification of dental biofilm. The diagram shows both stages of the proposed algorithm: (a) *the calibration stage* and (b) *the quantification stage*.

interference from visible light. The camera uses a fiber-optic cable in combination with a light shield and mirror that allows it to be positioned in the mouth with an orientation that best facilitates image capture of each tooth. QLF imaging uses a high-intensity blue light (405 ± 20 nm) that is shined onto the tooth surface and causes clean, undamaged enamel to fluoresce green. Red/orange fluorescence can also be seen on the tooth surface from within the oral biofilm.

V. RESULTS AND DISCUSSION

To evaluate the performance of our approach, we tested the quantification of dental plaque on canine teeth using QLF images. In order to evaluate the quality of the initial starting segmentation provided by the algorithm, another test is created where the automatic quantification results were compared to the rater scoring at 8 grades. The human experts assigned teeth imaged using QLF to 8 classes based on the amount of plaque covering the teeth (0%, 12.5%, 25%, 37.5%, 50%, 62.5%, 75%, and > 87.5%).

The biofilm quantification index computed during the first phase of computer-aided biofilm quantification closely follows manual scoring results. The algorithm has been evaluated on 470 images. The results are highly consistent with the manual gradings. Fig. 5 shows an initial-guess quantification on a set of dental images for two teeth taken over a period of 6 months. To evaluate the agreement between the quantification performed manually by human graders and the initial auto-mated segmentation, we employed *Cohen's kappa coefficient*, which measures the agreement between two raters (ground truth and the proposed method) who are classifying N terms into C mutually exclusive classes. The kappa coefficient κ is defined as

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \qquad (8)$$

where $\Pr(a)$ is the relative observed agreement among raters, and $\Pr(e)$ is random agreement among raters. If there is a perfect agreement among raters then $\kappa = 1$, and if there is no agreement except by chance (defined by $\Pr(e)$) then $\kappa = 0$. We applied κ statistics to our data set ($N = 470$,

		Predicted Class							
		1	2	3	4	5	6	7	8
Actual Class	1	64	9	0	0	0	0	0	0
	2	8	57	7	0	0	0	0	0
	3	0	6	51	4	0	0	0	0
	4	0	0	7	56	9	0	0	0
	5	0	0	2	5	48	9	0	0
	6	0	0	0	0	10	52	0	0
	7	0	0	0	0	0	9	49	5
	8	0	0	0	0	0	0	0	0

TABLE I
THE CONFUSION MATRIX FOR THE CANINE TEETH WITH 8 ACTUAL CLASSES PROVIDED BY THE HUMAN GRADER. IN ORDER TO HAVE A COMPARATIVE STUDY, AUTOMATIC BIOFILM QUANTIFICATION RESULTS FROM THE ALGORITHM ARE MAPPED AS 8 *predicted* CLASSES.

$C = 8$), and obtained $\Pr(a) = 0.6827$, $\Pr(e) = 0.1389$, and $\kappa = 0.6315$. According to [39], values < 0 indicate complete disagreement, 0 to 0.20 slight agreement, 0.21 to 0.40 fair agreement, 0.41 to 0.60 moderate agreement, 0.61 to 0.80 substantial agreement, and 0.81 to 1.0 perfect agreement. Table I shows the confusion matrix for actual (*ground truth*) and predicted (*automatic quantification algorithm*) classes. Furthermore, we created 8 one-versus-all classifications by categorizing images as either belonging to their true class or not. For an n-way classifier, a one-versus-all classifier can be constructed automatically from an $n \times n$ confusion matrix by treating the image to be classified as belonging to the class if the class is the result of classifying it correctly. For most cases the proposed algorithm provides a fairly accurate assessment of the biofilm; however, there could be instances where the au-tomatic quantification fails to perform optimally and therefore needs the additional integrated manual supervision provided by the second stage of our method. We found two image types that produce consistent misclassifications: the first includes images with multiple teeth that are fluorescing with similar intensity, thus making it difficult to define a clear boundary between them; the second includes images containing reddish artifacts appearing at the boundary between tooth and gum. Owing to relatively simple texture in QLF dental scans, the
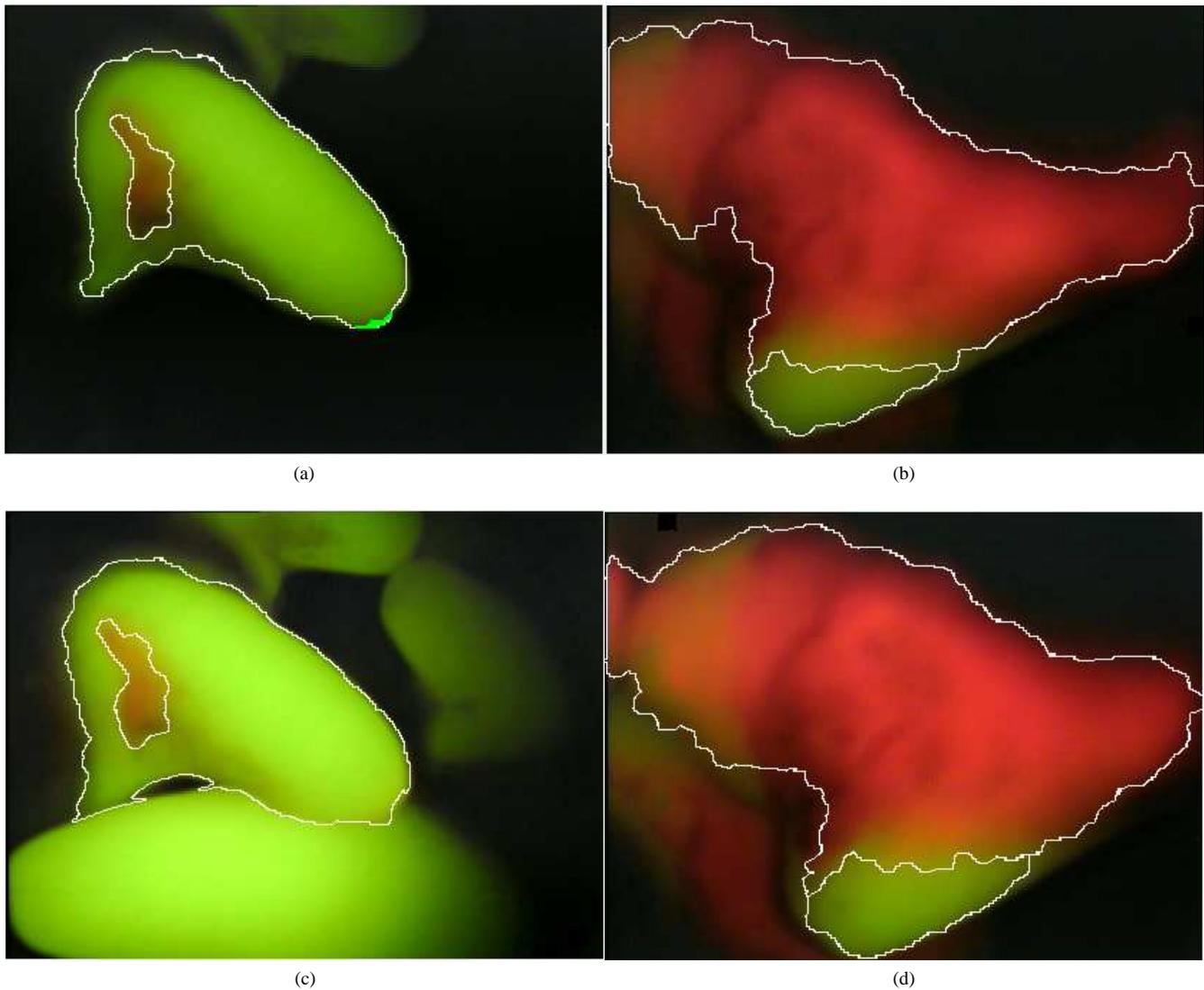
Fig. 5. The quantification maps produced by the proposed automatic quantification algorithm for QLF images (pre-expert tweaking). Two teeth from the same subject are scanned over a period of 6-months to quantify plaque over time. (a) shows a premolar canine tooth at month 0, (c) shows the same tooth after 6 months. (b), and (d) show a molar teeth of a canine imaged at the start of the study (month 0) and month 6 respectively.

number of superpixels does not appear to have much impact on the performance of the method over a wide range. The adequate number of superpixels depends on the ratio of size of the image to size of captured tooth of interest inside the image as well as the amount of fluorescent from neighboring teeth.

The images with misclassified superpixels constitute about 1% of our testing set. We observed that it takes a human grader interacting with initial unsupervised guesses 5 seconds per image on average to fix misclassifications. Screen shots of the software are shown in Fig. 6 (*left* fully automatic first stage , and *right* second stage optional correction). The software can be downloaded from here (https://www.nitrc.org/projects/biofilmquant). The current version of the program has the capability to process an entire jaw set in batch mode along with the longitudinal analysis (Fig. 7). The longitudinal analysis of the proposed technique provide a tool to help understand the progression of biofilm as well

as the monitoring of the effectiveness of clinical procedure over time. The software also has the capabilities to output the label image, the total biofilm coverage, and the mean biofilm coverage over time.

Our tests demonstrated that the utilized *click-switch* approach of superpixel re-classification is much more intuitive and accurate than drawing boundaries around tooth and plaque. The Fitt's Law providing an empirical model of human muscle movement (primarily used in human-computer interaction) explains the speed-accuracy trade-off characteristics: the faster we move, the less precise our movements are; the stricter the constraints are, the slower we move. According to Fitt's Law, the *click-switch* approach is expected to be faster and more accurate than drawing boundaries by order of logarithm.

VI. Conclusion

This paper presented a semi-automated method for biofilm quantification in canine teeth that is independent of grader
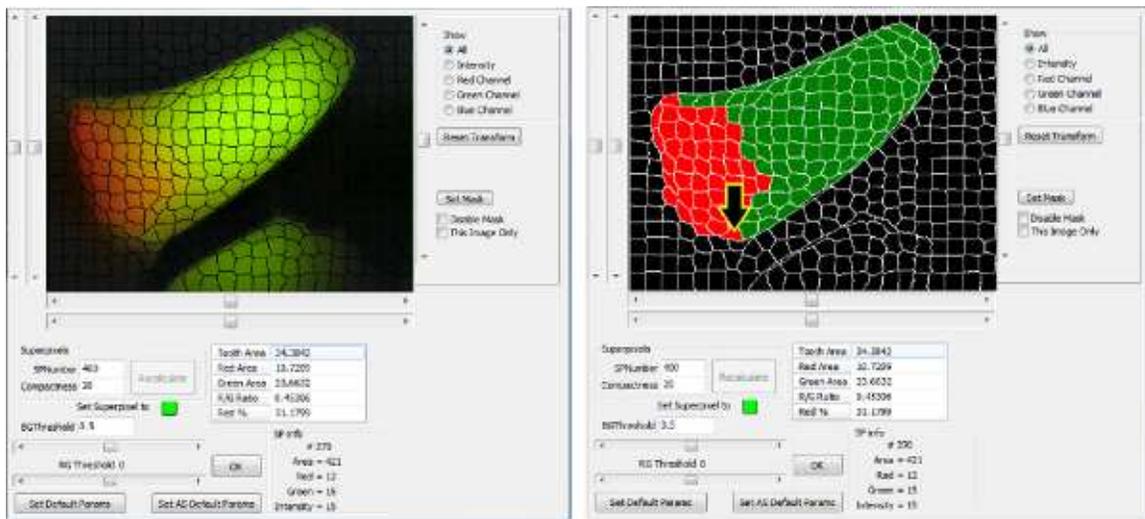
Fig. 6. Screenshot of the software implementing the algorithm, showing the quantified tooth before manual adjustment. The initial automatic quantification (*left*), if deemed suboptimal (red arrow), can be tweaked by the expert by a simple *click-and-flip* approach to a different class (*biofilm*, *tooth subtrata*, *background*)(*right*).

and instrument bias. The method takes a leap from traditional techniques of manual demarcation of tooth and plaque boundaries; at the same time it avoids simplistic automation approaches that allow pathologists only to completely accept or reject the quantification these methods produce. The core of the approach is a Gauss Markov random-field statistical model for QLF images. The results demonstrate that biofilm quantification using the statistical model is robust, reliable, and reproducible. The segmentation of QLF images into superpixels makes the classification and the subsequent quantification process accurate and fast. The method provides an initial quantification guess that is found to match human expert judgment in 99% of the tested images. The misclassified images can be manually corrected by an expert using a single-click switching of superpixels among three classes (background, tooth substratum, and dental plaque). The approach could be extended further to other quantification methods such as dental fluorosis by incorporating more features. Dental fluorosis is caused by excessive exposure to high concentration of fluoride. Fluorosis often appears as tiny white streaks or specks in the enamel of the tooth, and in extreme forms tooth appearance is marred by discoloration and brown markings. Current method of quantification is 6-class rater-based H.T. Dean's fluorosis index [40] proposed in 1942. More accurate quantification methods such as proposed here in this paper can help early diagnosis and better treatment. Furthermore, using expert corrections as a feedback with machine-learning methods for adjusting the cut-off threshold can help refine the initial segmentation.

REFERENCES

[1] S. P. Ramfjord, "Indices for prevalence and incidence of periodontal disease," *Journal of Periodontology*, vol. 30, pp. 51–59, 1959.

[2] J. Kang and Z. Ji, "Dental plaque quantification using mean-shift-based image segmentation," in *Computer Communication Control and Automation (3CA), 2010 International Symposium on*, vol. 1, May 2010, pp. 470 –473.

[3] J.-Y. Kang, L.-Q. Min, Q.-X. Luan, X. Li, and J.-Z. Liu, "Dental plaque quantification using FCM-based classification in HSI color space," in *ICWAPR '07. International Conference on Wavelet Analysis and Pattern Recognition, 2007.*, vol. 1, Nov 2007, pp. 78 –81.

[4] P. Hennet, E. Servet, H. Salesse, and Y. Soulard, "Evaluation of the Logan & Boyce plaque index for the study of dental plaque accumulation in dogs," *Research in Veterinary Science*, vol. 80, no. 2, pp. 175 – 180, 2006.

[5] R. N. Smith, A. H. Brook, and C. Elcock, "The quantification of dental plaque using an image analysis system: reliability and validation," *Journal of Clinical Periodontology*, vol. 28, no. 12, pp. 1158–1162, 2001.

[6] P. A. Sagel, P. G. Lapujade, J. M. Miller, and R. J. Sunberg, "Objective quantification of plaque using digital image analysis," *Monographs in Oral Science*, vol. 17, pp. 130–43, 2000.

[7] I. Pretty, W. Edgar, P. Smith, and S. Higham, "Quantification of dental plaque in the research environment," *Journal of Dentistry*, vol. 33, no. 3, pp. 193 – 207, 2005, symposium in honour of the work of Professor W. M. Edgar.

[8] A. Mansoor, W. M. Ahmed, A. Samarapungavan, J. Cirillo, D. Schwarte, J. P. Robinson, and B. S. Duerstock, "Accessscope project: Accessible light microscope for users with upper limb mobility or visual impairments," *Disability & Rehabilitation: Assistive Technology*, vol. 5, no. 2, pp. 143–152, 2010.

[9] B. Foster, U. Bagci, A. Mansoor, Z. Xu, and D. J. Mollura, "A review on segmentation of positron emission tomography images," *Computers in biology and medicine*, vol. 50, pp. 76–96, 2014.

[10] Z. Xu, U. Bagci, B. Foster, A. Mansoor, and D. J. Mollura, "Spatially constrained random walk approach for accurate estimation of airway wall surfaces," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*. Springer, 2013, pp. 559–566.

[11] A. Sandouk, U. Bagci, Z. Xu, A. Mansoor, B. Foster, and D. Mollura, "Accurate quantification of brown adipose tissue through pet-guided ct image segmentation," *Society of Nuclear Medicine Annual Meeting Abstracts*, vol. 54, no. Supplement 2, p. 318, 2013.

[12] A. Mansoor, U. Bagci, and D. Mollura, "Noise adaptive multi-resolution technique to accurately denoise pet, mri-pet, and pet-ct images," *Society of Nuclear Medicine Annual Meeting Abstracts*, vol. 55, no. Supplement 1, p. 2050, 2014.

[13] I. Pretty, A. Hall, P. Smith, W. Edgar, and S. Higham, "The intra- and inter-examiner reliability of quantitative light-induced fluorescence (QLF) analyses," *British Dental Journal*, vol. 193, no. 2, pp. 105–109, 2002.

[14] B. T. Amaechi and S. M. Higham, "Quantitative light-induced fluorescence: a potential tool for general dental assessment," *Journal of Biomedical Optics*, vol. 7, no. 1, pp. 7–13, 2002.

[15] I. Pretty, M. McGrady, C. Zakian, R. Ellwood, A. Taylor, M. O. Sharif, T. Iafolla, E. A. Martinez-Mier, P. Srisilapanan, N. Korwanich *et al.*,
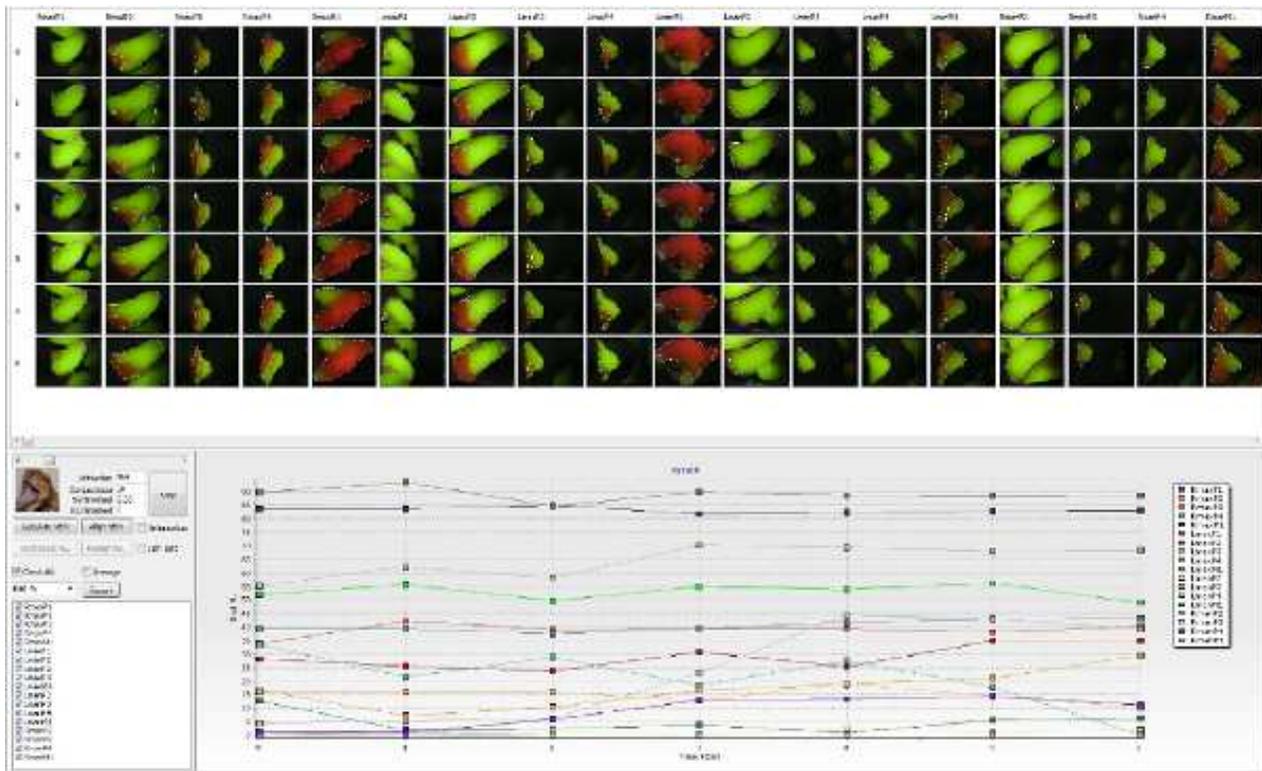
Fig. 7. The *batch-mode* of the proposed method showing the entire jaw-set being processed at once. The expert correction if needed due to non-uniformities in the acquisition procedure needs to adjusted once only on a single representative image for a given capture settings.

"Quantitative light fluorescence (QLF) and polarized white light (PWL) assessments of dental fluorosis in an epidemiological setting," *BMC Public Health*, vol. 12, no. 1, p. 366, 2012.

[16] I. A. Pretty, W. M. Edgar, and S. M. Higham, "Detection of in vitro demineralization of primary teeth using quantitative light-induced fluorescence (QLF)," *International Journal of Paediatrics Dentistry*, vol. 12, no. 3, pp. 158–167, 2002.

[17] S. Tranus, S. Al-Khateeb, S. Bjrkman, S. Twetman, and B. Angmar-Mnsson, "Application of quantitative light-induced fluorescence to monitor incipient lesions in caries-active children. a comparative study of remineralisation by fluoride varnish and professional cleaning," *European Journal of Oral Sciences*, vol. 109, no. 2, pp. 71–75, 2001.

[18] I. A. Pretty, G. S. Ingram, E. A. Agalamanyi, W. M. Edgar, and S. M. Higham, "The use of fluorescein-enhanced quantitative light-induced fluorescence to monitor de- and re-mineralization of in vitro root caries," *Journal of Oral Rehabilitation*, vol. 30, no. 12, pp. 1151–1156, 2003.

[19] J. Kang, X. Li, Q. Luan, J. Liu, and L. Min, "Dental plaque quantification using cellular neural network-based image segmentation," in *Intelligent Computing in Signal Processing and Pattern Recognition*, ser. Lecture Notes in Control and Information Sciences, D. S. Huang, K. Li, and G. Irwin, Eds. Springer Berlin / Heidelberg, 2006, vol. 345, pp. 797–802.

[20] J. Kang, Z. Ji, and C. Gong, "Segmentation and quantification of dental plaque using modified kernelized fuzzy C-means clustering algorithm," in *Control and Decision Conference (CDC), 2010 Chinese*, May 2010, pp. 788 –791.

[21] K. Jiayin and J. Zhicheng, "Dental plaque segmentation and quantification using histogram-aided fuzzy C-means algorithm," in *Control Conference (CC), 2010 29th Chinese*, Jul 2010, pp. 3068 –3071.

[22] A. Mansoor, "Statistical analysis and modeling of biological fluorescence images. methods and applications," Ph.D. dissertation, Purdue University, 2012.

[23] S. Krishnamachari and R. Chellappa, "Multiresolution Gauss-Markov random field models for texture segmentation," *Image Processing, IEEE Transactions on*, vol. 6, no. 2, pp. 251–267, 1997.

[24] R. Chellappa and S. Chatterjee, "Classification of textures using Gaussian Markov random fields," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP) '96.*, vol. 33, no. 4, August 1985, pp. 959–963.

[25] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 888–905, Aug 2000.

[26] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications*. CRC, 1984.

[27] X. Ren and J. Malik, "Learning a classification model for segmentation," *IEEE International Conference on Computer Vision*, vol. 1, p. 10, 2003.

[28] K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. M. Blei, and M. I. Jordan, "Matching words and pictures," *Journal of Machine Learning Research*, vol. 3, pp. 1107–1135, Mar 2003.

[29] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, pp. 167–181, 2004.

[30] Y. Deng and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 800–810, 2001.

[31] A. Blum and S. Chawla, "Learning from labeled and unlabeled data using graph mincuts," in *Proceedings of the Eighteenth International Conference on Machine Learning*, ser. ICML '01. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001, pp. 19–26.

[32] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, vol. 10, pp. 207–244, Jun 2009.

[33] C. Stauffer, W. Eric, and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 747–757, 2000.

[34] J. Shi, S. Belongie, T. Leung, and J. Malik, "Image and video segmentation: The normalized cut framework," in *IEEE International Conference on Image Processing*, 1998, pp. 943–947.

[35] X. Descombes, M. Sigelle, and F. Preteux, "Estimating Gaussian Markov random field parameters in a nonstationary framework: Application to remote sensing imaging," *IEEE Transactions on Image Processing*, vol. 8, no. 4, pp. 490–503, 1999.

[36] E. Ranguelova and A. Quinn, "Analysis and synthesis of three-dimensional Gaussian Markov random fields," in *Proceedings of IEEE International Conference on Image Processing, 1999.*, vol. 3, 1999, pp. 430–434.

[37] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. Wiley, Mar 1982.

[38] S. K. Warfield, K. H. Zou, and W. M. Wells, "Validation of image segmentation and expert quality with an expectation-maximization algorithm," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2002*. Springer, 2002, pp. 298–306.

[39] K. L. Gwet, *Handbook of Inter-Rater Reliability*. Gaithersburg, Maryland: Advanced Analytics LLC., Jun 2010.

[40] *Fluoridation facts*. American Dental Association, Council on Access, Prevention and Interprofessional Relations, 1993.