Semi-Automatic Teeth Segmentation in Cone-Beam Computed Tomography by Graph-Cut with Statistical Shape Priors

Timothée Evain[1,2](✉) *Xavier Ripoche*[2] *Jamal Atif*[3] *Isabelle Bloch*[1]

[1] LTCI, Télécom ParisTech, Université Paris-Saclay, Paris, France

[2] Carestream Dental, Croissy-Beaubourg, France

[3] Université Paris-Dauphine, PSL Research University, CNRS, UMR 7243 LAMSADE, 75016 Paris, France

ABSTRACT

We propose a new semi-automatic framework for tooth segmentation in Cone-Beam Computed Tomography (CBCT) combining shape priors based on a statistical shape model and graph cut optimization. Poor image quality and similarity between tooth and cortical bone intensities are overcome by strong constraints on the shape and on the targeted area. The segmentation quality was assessed on 64 tooth images for which a reference segmentation was available, with an overall Dice coefficient above 0.95 and a global consistency error less than 0.005.

1. INTRODUCTION

Teeth segmentation in Cone-Beam Computed Tomography (CBCT) is an essential step towards the development of a new era of dental practice, supported by the increasing number of CBCT-equipped dental offices. Analyzing and exploiting 3D acquisitions require a considerable amount of time for practitioners, since they need to review the entire image to detect any potential pathology for diagnosis. Thus there is a need for developing automatic or semi-automatic methods to facilitate and support dental evaluation, diagnosis, surgical and orthodontic planning. Among the different structures in the maxillo-facial area, teeth are the central focus for most of the dental procedures, enhancing the need for their segmentation. Nevertheless, this task is still challenging. First, CBCT images have a poor contrast, with increased noise compared to Multi Slice Spiral (Helical) Computed Tomography (MSCT) due to increased scatter radiations [1]. Secondly, the dentine, which composes most of each tooth, has an intensity similar to the neighboring cortical bone, disturbing the delimitation of teeth area. Another difficulty is caused by metallic elements that are frequent due to dental procedures such as fillings, crown replacements, implants, bridges, or braces. The presence of metal in CT acquisitions produces numerous image artifacts such as high intensity streaks, blackened areas (loss of data) or blur due to the beam hardening and the scatter in the imaging physical process. The last challenging point is the diversity of acquisition protocols, resulting in diverse resolutions (from 75 μm up to 400 μm in each direction) and signal-to-noise ratios. In this context, relying only on intensity for the segmentation does not achieve consistent results. To cope with these difficulties, we propose to integrate shape information and user input into the segmentation process to design an efficient framework to achieve precise results with reasonable computation time. In Section 2, we review related works, before introducing our method in Section 3. Results are presented in Section 4, and conclusion in Section 5.

corresponding author: tevain@telecom-paristech.fr

2. RELATED WORKS

Most of the recent efforts on teeth segmentation rely on level sets methods. In [2], two 3D level sets and a panoramic projection are used in MSCT data. Satisfactory results could be obtained (sensitivity and specificity around 0.88) but due to the aforementioned context and shape variability, relying only on intensity penalizes the method. Moreover, level-sets are computationally expensive, and computation times are not reported. In [3, 4], an iterative segmentation method by 2D level sets including shape priors provided by the previous iteration result is applied to mixed MSCT/CBCT data. Such priors allow achieving state-of-the-art results (reported Dice over 0.95, on molars and incisors only) but lack consistency since they can represent non-tooth shapes depending on the success of the last segmentation, thus making the use of these iterative priors hazardous in presence of metal. Additionally, computation times are reasonable, but not ideal. Graph-cuts (GC) could be used as an optimization of the energy problem as in [5], where its usability is demonstrated, but results suffer from relying only on intensity (average Dice of 0.89). Nonetheless, GC has the advantage of supporting hard constraint inputs with less parameter tuning, which is useful for user interaction. In [6] a statistical shape model (SSM) is used as prior in a probabilistic graph cut framework, but reported results (average Dice of 0.86) suffer from the lack of intensity prior, and computation times are not reported.

Apart from these applications on teeth, we could find interesting related methods applied on different structures, such as in [7] where GC and SSM are combined in a double optimization method used on both CT and MRI. The SSM is formulated as a Point Density Model (PDM), which is classically optimized as in the Active Shape Model (ASM) procedure [8], leading to a distance map shape prior in the GC optimization with an original boundary-based shape prior. Freedman et al. [9] used a deformed template to generate a region-based shape prior into their graph energy for designing a multi-purpose segmentation framework in medical or natural images. In [10] SSM is used in cardiac MRI to generate a distance map highlighting space areas where shape deformations may occur by creating deformed instances of the model. The shape prior then only allows GC segmentation on these areas. These approaches could be adapted to our problem and combined in order to improve the accuracy and the robustness of the teeth segmentation process.

To the best of our knowledge, no method currently allows performing accurate tooth segmentation in CBCT images in presence of metallic artifacts, with priors ensuring a consistent tooth shape, and in a manner compatible with clinical practice, including easy user input and reasonable computation time. The method proposed in this paper addresses these issues and open problems.

978-1-5090-1172-8/17/$31.00 ©2017 IEEE

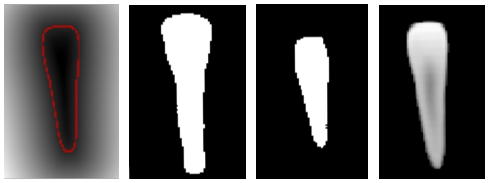

Fig. 1: Statistical model: mean shape distance map ($\overline{\text{M}}$) with isoline 0 in red; two shape variations examples according to the first mode (S_1^+ and S_1^-); intensity mean model (M).

3. PROPOSED METHOD

Our method is based on a graph cut framework with a statistical shape prior enhanced with a user input. The user input indicates a focus region for each tooth, then a mean tooth model is registered on the image, providing the shape prior. Once a preliminary segmentation has been obtained by graph cut optimization with this prior, the result is refined to discard potential missegmentation of the tooth region.

3.1. Statistical Shape Model construction

The shape prior is based on a SSM obtained from 15 full arch CBCT images, which have been manually segmented by experts (dentists and radiologists), each tooth separately. The data are split into 10 different classes depending on their morphological characteristics (e.g. mandibular incisors, maxillar left molars). A SMM is trained for each class following the classical PCA procedure [8], resulting in mean shapes and principal modes in the form of signed distance maps. Then, the manual segmentations are warped by a B-spline registration on the binary mean shape before averaging the intensity, resulting in a mean model with intensity. Each class model is then composed by a mean shape map, principal deformation modes maps (both as signed distances, where negative ones are inside the object) and an intensity mean model (see Figure 1). The intensity probability density functions (PDF) of each class of teeth as well as the background are also computed.

3.2. Initialization

The user is asked to select for each tooth of interest a set of points composed by the "crown point" which is located on the surface of the tooth, approximately in the middle of the edge (for anterior teeth) or in the center of fissures (for posterior teeth), and "root points", which are located inside each root apex. For each selected set of points, the user is asked to assign the corresponding tooth label (according to FDI World Dental Federation notation, described in ISO3950 standard). Using these data, each tooth is pre-processed by the following method:

1. Crop a large region of interest (ROI).
2. Compute the barycenter of root points, then the top-bottom axis with the crown point.
3. Resample the region to align the top-bottom axis with the vertical direction.
4. Crop a precise ROI around the tooth, depending on the tooth label.

This allows us to focus the subsequent processing on the selected tooth, discarding as much as possible information coming from neighboring teeth.

3.3. Registration

The next step is the rigid registration of the intensity mean model of the class corresponding to the tooth label given by the user. The transformation is initialized by aligning the grey levels moments between the intensity mean model and the image. The optimization is carried out through a gradient ascent scheme to maximize the correlation between intensities:

$$Corr(\text{Im,M}) = \frac{\sum_{x\in N}(\text{Im}(x)\cdot\text{M}(x))}{\sqrt{\sum_{x\in N}\text{Im}(x)^2\sum_{x\in N}\text{M}(x)^2}},$$

where Im and M are the image and intensity mean model respectively, x a voxel, N the set of voxel samples (equal to the whole set of voxels in the image in the case of a dense registration). Correlation was experimentally preferred over mutual information for its ability to achieve more precise registration in most cases (i.e. when the grey level distribution of tooth is not damaged by artifacts) for our dataset. Usually, thanks to the initialization, the number of iterations required to reach satisfactory registration is less than a hundred. To attenuate the impact of metallic artifacts, voxels above an empirical threshold (stable for all our data) corresponding to metal range are not included in the correlation computation.

3.4. Shape prior generation

Following the idea from [10], i.e. using the space of allowed shapes from the SSM as segmentation constraints, we generate deformed instances of the model along each principal mode by binarization of the linear combination:

$$S_i^{\pm} = \mathcal{H}(-(\overline{\text{M}} \pm \text{C}\sqrt{\lambda_i}\text{Pm}_i))$$

where $\overline{\text{M}}$ is the mean shape map and Pm_i the i-th principal mode associated with eigenvalue λ_i, $\mathcal{H}$ being the Heaviside function. The constant C controls the amplitude of deformation i.e. the extent of allowed shapes space. These shapes (see Figure 1) are compared with the mean one to retain only areas which have changed, leading to a binary mask where "on" voxels are the ones likely to be part of a deformed shape:

$$\text{Mask}(x) = \begin{cases} 1 & \text{if} \quad \exists\, i : S_i^{\pm}(x) \neq \mathcal{H}(-\overline{\text{M}}(x)), \\ 0 & \text{otherwise.} \end{cases}$$

This mask is applied to the mean shape map resulting in a sparse distance map ϕ:

$$\phi(x) = \begin{cases} \overline{\text{M}}(x) & \text{if} \quad \text{Mask}(x) = 1, \\ +\infty & \text{if} \quad \text{Mask}(x) = 0 \text{ and } \mathcal{H}(-\overline{\text{M}}(x)) = 0, \\ -\infty & \text{if} \quad \text{Mask}(x) = 0 \text{ and } \mathcal{H}(-\overline{\text{M}}(x)) = 1. \end{cases}$$

Finally, the registration transform is applied to ϕ, giving the shape prior used hereafter (see Figure 2).

3.5. Graph Cut segmentation

Our graph cut formulation combines the usual image based terms as in [11] with two shape prior terms relying on the previously computed map, resulting in the following energy:

$$E = \sum_{x\in N}[\alpha R_x(l_x) + \beta D_x(l_x)] + \sum_{x,y\in Neigh}[\gamma B_{x,y} + \delta T_{x,y}],$$

with x, y voxels, N the set of all voxels, $Neigh$ the desired neighborhood (26-connected in our case) and l_x the class label of x (object

"O" and background "B" here). In the following "regional" refers to edges connecting nodes and classes terminals, also known as t-links, while "boundary" refers to edges between nodes, also known as n-links. The image based regional and boundary terms are:

$$R_x(l_x) = \begin{cases} -\ln \mathrm{P}(\mathrm{Im}(x) \mid \mathrm{O}) & \text{if} \quad l_x = \mathrm{O}, \\ -\ln \mathrm{P}(\mathrm{Im}(x) \mid \mathrm{B}) & \text{if} \quad l_x = \mathrm{B}, \end{cases}$$

$$B_{x,y} = \exp\left(-\frac{(\mathrm{Im}(x) - \mathrm{Im}(y))^2}{2\sigma^2}\right)\left(\frac{1}{\mathrm{dist}(x,y)}\right).$$

The first term penalizes the assignation of a voxel to classes which have very different grey levels according to the intensity PDF prior, whereas the second favors cutting the graph along high gradients, which are expected to be edges of objects in the image. The regional shape prior term D_x is inspired by the work in [7] and the boundary one $T_{x,y}$ by [9]:

$$D_x(l_x) = \begin{cases} \frac{1}{1+\exp(\phi(x))} & \text{if} \quad l_x = \mathrm{B}, \\ 1 - \frac{1}{1+\exp(\phi(x))} & \text{if} \quad l_x = \mathrm{O}, \end{cases}$$

$$T_{x,y} = |\phi\left(\frac{x+y}{2}\right)|.$$

The term $D_x(l_x)$ penalizes or favors voxels classification as object based on the distance to the mean shape, while $T_{x,y}$ tends to favor cuts around the boundary of this shape. Voxels out of the allowed segmentation area are considered as hard constraints by setting their t-link value to infinity towards their opposite class (object if outside the mean shape, background if inside). To help the segmentation of roots, shortest paths from root points to crown points are computed based on intensity (to stay inside the tooth). Ideally, we want the paths to run through the root canals, so we enhance them by setting to high intensity values all holes detected inside a thresholded image of dentine. To compute this image, Otsu thresholds [12] are computed for 2 and 3 intensity classes, giving t^2, t^3_{lower} and t^3_{upper} where $\mathrm{t}^3_{lower} \leqslant \mathrm{t}^2 \leqslant \mathrm{t}^3_{upper}$. Usually, t^2-thresholding gives a well-defined tooth shape but with clogged canals, whereas t^3_{upper} one tends to lead to non-closed tooth shape, so we sample values in a $[\mathrm{t}^2, \mathrm{t}^3_{upper}]$ range to find the one giving the largest hole in terms of volume to select it as the final threshold. Once shortest paths have been established, a dilation of the paths is used as object hard constraint into the GC (see Figure 2). The min-cut/max-flow is done with the Kolmogorov open-implementation [13]. The parameters α, β, γ and δ, which are the weights of each term, are set depending on the application.

3.6. Post processing

Segmentation errors may occur during the previous step due to the lack of gradient between neighboring crowns or blurred edges. A watershed is performed on the inside of the signed distance map of the segmentation and only the central component is kept, which discards touching crowns leaks. Finally a morphological opening is performed with a spherical element of variable size (largest on crown area and smaller toward roots) to avoid potential residues.

4. EXPERIMENTS

Experiments were conducted on clinical CBCT data obtained from partners dental offices on a dual-core CPU (4 threads @ 2.9GHz). Average computation times (from initialization to final segmentation, excluding user input) range from 12 seconds for anterior teeth to 30 seconds for posterior ones, while memory consumption tops at 2 GB. The database was made of 64 teeth images with an isotropic

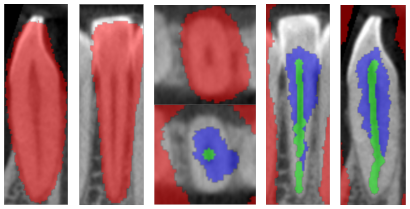

Fig. 2: Mean model (in red) registration (on 3 orthogonal views), and constraint map applied to graph cut (same views): background seeds (red, outside of allowed deformation area) and object seeds (green: shortest path, blue is outside of allowed deformation area).

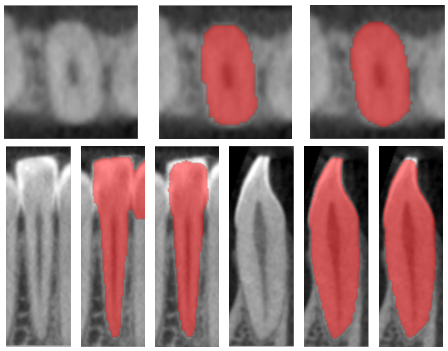

Fig. 3: Segmentation results on a mandibular incisor (tooth 31) along axial, coronal and sagittal axes. For each view, the second image is the graph cut output, and the third one is the result after the post processing.

resolution of 150 μm (450K to 1.5M voxels depending on tooth, all tooth classes were included). Note that patients were biting a gum-shield since occlusal surface is barely visible in occlusion case, and maxillar-mandibular teeth discrimination is a challenging task, which is out of the scope of this paper. Implementation was done in C++ based on the ITK library (www.itk.org). In the following, the number of iterations for the registration is set to 30, with a 0.001 minimal gradient change, deformation constant C = 3, and parameters are set to $\sigma = 100$, $\alpha = 1$, $\beta = 5$, $\gamma = 15$ and $\delta = 1$.

Figure 3 shows typical results of the method on an anterior tooth. The missegmentation problem mentioned in Section 3.6 is visible on the coronal slices, and is corrected by the post-processing stage. These final results slightly underestimate the true tooth volume due to the morphological opening, but this choice was made to be sure to exclude neighboring teeth parts. Figure 4 reports the comparison

	GC output	Post-process output
Sensitivity	0.969 ± 0.023	0.972 ± 0.023
Specificity	0.998 ± 0.001	0.999 ± 0.001
Accuracy	0.998 ± 0.001	0.998 ± 0.001
Dice	0.953 ± 0.024	0.958 ± 0.023
GCE [14]	0.0041 ± 0.0026	0.0038 ± 0.0021

Table 1: Overall averages of sensitivity, specificity, accuracy, Dice and global consistency error, as defined in [14].

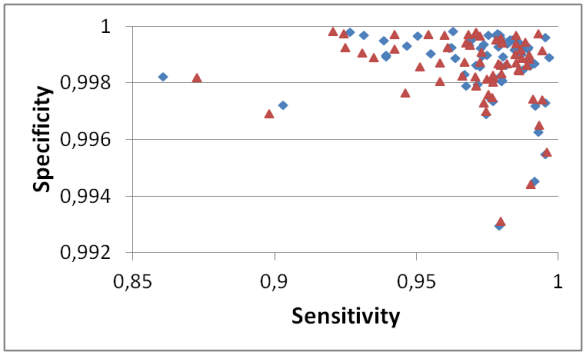

Fig. 4: Segmentation comparison with ground truth over the 64 samples. Red: GC output, blue: final result.

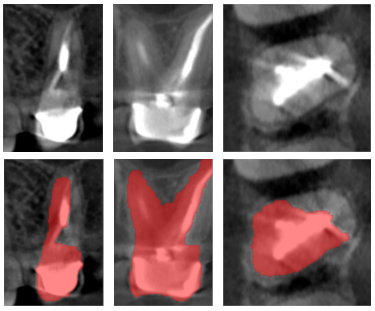

Fig. 5: Graph-cut segmentation output on maxillar molar with full metallic crown and fillings.

with the manual delineations done by experts for each sample, while the average values are shown in Table 1. Although the global results are satisfactory, there are still complex cases where the algorithm stumbles, as in Figure 5, where metallic artifacts reduce the registration consistency and induce false edges. In such cases, increasing the α weight and using a more stringent segmentation area by reducing the allowed deformation of the model (Section 3.4) increase dental parts recovering, at the cost of an ill-defined contour.

5. CONCLUSION

We presented a new framework for the segmentation of teeth in dental CBCT images, based on shape priors and user inputs to be able to cope with modality-inherent and context-based problems such as similarity with the immediate surroundings or weakly defined edges. Experiments demonstrate state-of-the-art quality [3] of the results, with an average Dice above 0.95, independently of the tooth class. However, the method can still be improved, first by better handling artifacts due to metallic objects, which penalize registration as well as segmentation. Enhancing the images beforehand with a Metallic Artifact Reduction (MAR) method should be considered for improving the detection and segmentation quality. Also, the framework should take advantages of considering multiple segmentations concurrently (of several teeth) to improve neighboring teeth delineation. Secondly, computation times are reasonable, but longer than [3]. The most consuming task is the GC part, so code optimization and volume sub-sampling are currently investigated.

6. REFERENCES

[1] W. De Vos, J. Casselman, and G. R. J. Swennen, "Cone-beam computerized tomography (CBCT) imaging of the oral and maxillofacial region: A systematic review of the literature," *International Journal of Oral and Maxillofacial Surgery*, vol. 38, pp. 609–625, 2009.

[2] M. Hosntalab, R. Aghaeizadeh Zoroofi, A. Abbaspour Tehrani-Fard, and G. Shirani, "Segmentation of teeth in CT volumetric dataset by panoramic projection and variational level set," *International Journal of Computer Assisted Radiology and Surgery*, vol. 3, no. 3, pp. 257–265, 2008.

[3] H. Gao and O. Chae, "Individual tooth segmentation from CT images using level set method with shape and intensity prior," *Pattern Recognition*, vol. 43, no. 7, pp. 2406–2417, 2010.

[4] D. X. Ji, S. H. Ong, K. Foong, and C. Weng, "A level-set based approach for anterior teeth segmentation in cone beam computed tomography images," *Computers in Biology and Medicine*, vol. 50, pp. 116–128, jul 2014.

[5] L. T. Hiew, S. H. Ong, K. Foong, and C. Weng, "Tooth Segmentation From Cone-Beam CT Using Graph Cut," in *Asia-Pacific Signal and Information Processing Association (APSIPA)*, 2010, pp. 272–275.

[6] J. Keustermans, D. Vandermeulen, and P. Suetens, "Integrating Statistical Shape Models into a Graph Cut Framework for Tooth Segmentation," in *Machine Learning in Medical Imaging - MLMI12*, 2012, vol. LNCS 7588, pp. 242–249.

[7] X. Chen, J. K. Udupa, A. Alavi, and D. A. Torigian, "GC-ASM: Synergistic integration of graph-cut and active shape model strategies for medical image segmentation," *Computer Vision and Image Understanding*, vol. 117, no. 5, pp. 513–524, 2013.

[8] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active Shape Models-Their Training and Application," *Computer Vision and Image Understanding*, vol. 61, pp. 38–59, 1995.

[9] D. Freedman and T. Zhang, "Interactive graph cut based segmentation with shape priors," in *Computer Vision and Pattern Recognition, CVPR*, 2005, vol. I, pp. 755–762.

[10] D. Grosgeorge, C. Petitjean, J. N. Dacher, and S. Ruan, "Graph cut segmentation with a statistical shape model in cardiac MRI," *Computer Vision and Image Understanding*, vol. 117, no. 9, pp. 1027–1035, 2013.

[11] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient N-D image segmentation," *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109–131, 2006.

[12] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 20, no. 1, pp. 62–66, 1975.

[13] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124–1137, 2004.

[14] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool.," *BMC Medical Imaging*, vol. 15, pp. 1–29, 2015.