



Automatic Teeth Segmentation in Panoramic X-Ray Images Using a Coupled Shape Model in Combination with a Neural Network

Andreas Wirtz^(✉), Sudesh Ganapati Mirashi, and Stefan Wesarg

Fraunhofer IGD, Fraunhoferstr. 5, 64283 Darmstadt, Germany
{Andreas.Wirtz,Stefan.Wesarg}@igd.fraunhofer.de

Abstract. Dental panoramic radiographs depict the full set of teeth in a single image and are used by dentists as a popular first tool for diagnosis. In order to provide the dentist with automatic diagnostic support, a robust and accurate segmentation of the individual teeth is required. However, poor image quality of panoramic x-ray images like low contrast or noise as well as teeth variations in between patients make this task difficult. In this paper, a fully automatic approach is presented that uses a coupled shape model in conjunction with a neural network to overcome these challenges. The network provides a preliminary segmentation of the teeth region which is used to initialize the coupled shape model in terms of position and scale. Then the 28 individual teeth (excluding wisdom teeth) are segmented and labeled using gradient image features in combination with the model's statistical knowledge about their shape variation and spatial relation. The segmentation quality of the approach is assessed by comparing the generated results to manually created gold-standard segmentations of the individual teeth. Experimental results on a set of 14 test images show average precision and recall values of 0.790 and 0.827, respectively and a DICE overlap of 0.744.

Keywords: Coupled shape model · Automatic segmentation
Panoramic dental x-ray image · Deep learning

1 Introduction

Radiographic images are a common tool used in dentistry for diagnosis. They help the dentist to identify many teeth related problems like caries, infections and bone abnormalities, which would be hard or impossible to detect during visual inspection only. This allows the dentist to choose the optimal treatment plan for the patient. Dental radiographs can be divided into two categories: Intra-oral and extra-oral [13]. Intra-oral images like bitewing, periapical or occlusal are obtained inside the patient's mouth. They only show specific regions of the set of teeth or individual teeth and are mostly used to get more detailed information. Extra-oral images like cephalometric or panoramic, also known as orthopantomographic

images, capture the entire teeth region as well as the surrounding areas and provide fundamental information about the teeth of a patient (cf. Fig. 1).

The analysis of these images is still done manually since automated tools that provide support for this procedure are not available. Therefore, the evaluation of these images and the design of the patient's treatment plan heavily relies on the dentist's experience and visual perception [13]. The lack of automated tools results from difficulties when dealing with dental X-Ray images. These include but are not limited to gaps caused by missing teeth, poor image quality like intensity variation, noise or low contrast, artifacts due to restorations, caries, and variations of the teeth in between patients [1]. An (automatic) segmentation of the individual teeth in a radiographic image is an essential step for providing tools for an automatic analysis of such images [9]. Automatic analysis can not only be helpful for automatic diagnosis in dentistry but could also be used for forensic purposes (i.e. postmortem identification).

Several methods have been proposed in the past to extract information from panoramic radiographs. Lira et al. [6] have used quadtree decomposition, morphological operators and snake models to segment individual teeth. They also employed shape models but use them only for teeth recognition. Amer and Aqel [1] have segmented only the wisdom teeth. They have used otsu-thresholding and morphological dilation to extract the region of interest (ROI) and have then applied masks at the end of the ROI to extract the wisdom teeth. Hasan et al. [3] have used clustering, thresholding and GVF snakes to segment the jaws in panoramic images. Recently, Silva et al. [10] reviewed state of the art methods for teeth segmentation in dental radiographs. They separated relevant works into different categories like threshold-based, region-based or boundary-based methods and also grouped them by the type of image these methods can be applied to. The majority are threshold-based approaches (54%) followed by boundary-based methods (34%). More importantly, the reviewed papers focused mostly on intra-oral X-Ray images (80%). To close this scientific gap, Jader et al. propose a novel data set featuring 1500 panoramic radiographic images. Additionally, they compared 10 different segmentation methods on their data set with the goal to extract the teeth and provide a comprehensive performance assessment. None of the analyzed methods provided satisfactory results and the authors conclude that an adequate method for the segmentation of dental X-Ray images, which can serve as a basis for automatic analysis, is still to be found.

In this paper, a novel method for automatic teeth segmentation in dental panoramic radiographs is presented. A 2-D coupled shape model based on [5] is used to segment and label 28 individual teeth. The 2-D coupled model is composed of a statistical shape model (SSM) for each tooth which is coupled with all other individual models using their spatial relation. This enables a more robust segmentation process using gradient image features (bottom-up) in combination with a priori statistical knowledge about the teeth in order to guide the segmentation process (top-down) [7]. A drawback of statistical models is that they rely on a good initial placement if local search algorithms like active shape models are used [4]. We propose to handle this by using a binary mask of the teeth area that is generated using a deep neural network [8]. The mask is then used for the initialization of the coupled shape model in terms of position and scale.

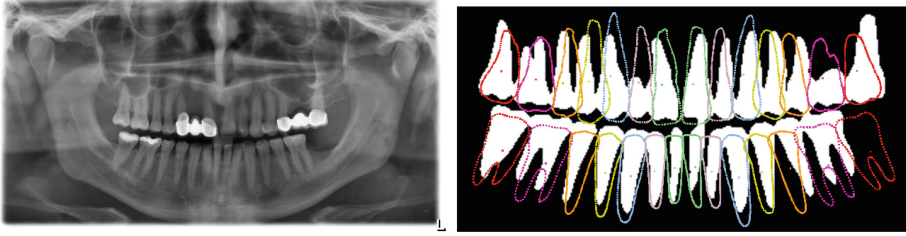


Fig. 1. A more difficult case with bridges and missing teeth (left) and the initial placement of the mean coupled model on the corresponding (cropped) binary mask (right).

2 Methods

2-D Coupled Shape Model. The presented method is based on a 3-D coupled shape model consisting of rigid model items and deformable model items [5]. It has already been successfully applied to CT images in order to segment different structures in the head & neck area [11, 12]. To be able to apply the model to dental X-Ray images, the coupled shape model has been extended to support single 2-D images. Deformable 2-D model items and the corresponding 2-D transformations have been added that represent the contour of objects and their relative transformation to the center of the coupled model. These 2-D model items are represented as statistical shape models and are generated using a point distribution model (PDM) [2] and principal component analysis (PCA).

The contour of an individual item is represented by 100 landmark points and denoted as vector $c = (x_1, y_1, \dots, x_{100}, y_{100})$. Procrustes alignment is used to transform all s training instances of a single item into a common coordinates system. The statistical information of these training instances is then extracted using PCA by computing eigenvectors e_m and their respective eigenvalues λ_m (with $\lambda_i > \lambda_{i+1}$) of the covariance matrix $S = \frac{1}{s-1} \sum_{i=1}^s (c_i - \bar{c})(c_i - \bar{c})^T$, where \bar{c} is the mean shape. For each model item, only the first n eigenvectors required for capturing 95% of the shape variance are kept and the remaining ones are discarded. Every valid shape \tilde{c} can then be approximated by a linear combination of these n principal modes: $\tilde{c} = \bar{c} + \sum_{i=1}^n v_i e_i$.

The coupled shape model is created by combining the relative pose of each model item in relation to the center of mass of the complete model and its shape information. The parameter vector p_j for an individual model item j consists of $4 + n$ entries, the 4 transformation parameters (2 for translation and 1 for rotation and isotropic scaling, respectively) and the n principal modes. By concatenating the parameter vectors p_j of all model items for a training instance k , the configuration vector f_k is generated. Again, PCA is used here to describe the space of all possible configurations over all training instances, which is later used during the adaptation process. Any possible configuration can then be described by a vector b as $\tilde{f} = \bar{f} + A \cdot b + r$, where \bar{f} is the mean configuration and A is the matrix containing all eigenvectors of the covariance matrix of all possible configurations. For more details see [12].

The coupled shape model consists of 28 individual teeth and is trained based on 10 manually annotated panoramic X-Ray images. Wisdom teeth are not included in the model at the moment due to the limited amount of training data available. In order to adapt the model and segment the teeth in the panoramic image, a robust initialization of the coupled model onto the image is required.

Binary Mask Generation. The binary mask of the teeth area, which is used for the initialization process, is generated by a modified neural network (u-net model) proposed by Ronneberger et al. [8]. Simpler, threshold-based methods proved ineffective in producing reliable results for the initialization of the coupled model. The network was originally used to segment neuronal structures in microscopy images, but has also been successfully applied to bitewing radiographs [13]. The advantage of this kind of neural network is that it can be trained with a limited number of training data and still produce accurate segmentation results. This is achieved by relying on data augmentation to extend the training set and a specially tuned network architecture. For more details see [8]. The modified U-Net model used in this work has the same internal architecture as the original one, but was modified to work on input images with a resolution of 608×320 pixels and only uses half the amount of channels on each layer (i.e. 32 on the first layer instead of 64). With this design, the generated binary masks were robust and detailed enough to be used for the initialization process. Figure 1 shows a panoramic image and the corresponding (cropped) binary mask.

The network was trained on the same 10 training instances that were used for the coupled shape model. Image augmentation was applied to increase the number of training images to about 4000.

Model Initialization. The coupled shape model needs to be initialized on the input image in terms of position and scale. Both values are computed on the binary mask of the teeth area. For the position, instead of using a simple center of mass, which is easily affected by missing or broken teeth, the presence or absence of wisdom teeth and unusual teeth configurations, the x- and y-coordinates are computed separately. First, the bounding box of the binary mask is computed. The y-coordinate is then determined using a horizontal projection to identify the location of the ‘valley’ between upper and lower jaw. The x-coordinate is calculated by taking the average of the x-centroid of the bounding box area and its inverse. The scale value is initially approximated by the ratio between the sizes of the bounding box of the binary mask and the bounding box of the mean model. This value is then refined by placing the model according to previously computed position and maximizing the overlap between mask and model. Having both the values for position and scale, the coupled model can be initialized onto the input image. An example of an initialization is shown in Fig. 1.

Model Adaptation. After the initialization process, the model is adapted to the input image. The adaptation is done by minimizing an energy functional E . It depends on two parameters: (a) the transformation t describing the global position of the model in terms of translation and rotation and (b) the configuration

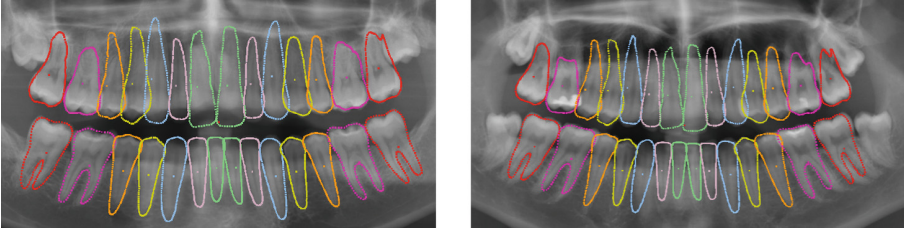


Fig. 2. Two examples of segmentation results.

vector f describing the configuration of the coupled model. The functional is given by $E(f, t) = E_{ext}(f, t) + \lambda E_{int}(f)$. The external term E_{ext} is responsible for ensuring that the contour of model items moves in the direction of strong image features (gradient features) whereas the internal energy term E_{int} restricts the model to stay within or close to the learned configuration space. The optimization process is done using a gradient descent optimizer. The transformation parameters t are optimized first, and then the configuration and transformation parameters f, t are optimized jointly.

The process of adapting the model to the image is separated into multiple steps. Initially, the model is adapted to the binary mask to ensure a good placement of all teeth before using the intensity image. Starting with only the 4 incisors, the set of model items (teeth) which are adapted to the image is gradually extended. Model items that are not adapted during an adaptation step are only changed passively through the learned (spatial) configuration. The gradual extension of the set of adapted items is done because the mean model is initialized according to its center of mass. Therefore, the teeth close to the center (the incisors) show good overlap with the input image while the teeth farther away from the center (e.g. molars) might not match as good, depending on the structure of the teeth in the input image. By adapting the outer teeth at a later time, they have already been (passively) moved closer to their correct position and more reliable image features will be found once they are adapted. The final adaption is then performed on the intensity image and the segmentation result of each individual tooth is stored as a binary image.

3 Experiments and Results

The presented fully automatic segmentation method has been evaluated on a separate test set of 14 manually annotated panoramic radiographic images (referred to as gold-standard segmentations), which were not part of the training set. The images each have a resolution of 2440×1280 pixels. The test set includes images of a variety of cases with several difficulties like completely- and partially missing teeth, artificial teeth, fillings and bridges. It also covers patients that have all, none or only some wisdom teeth.

First, the results of the automatic initialization have been assessed visually to determine if the model was positioned and scaled correctly. The placement was

Table 1. The average, minimum and maximum values for different metrics used for comparing the segmentation results to gold-standard segmentations.

	Precision	Recall	Accuracy	Specificity	F-Score	DICE
Avg.	0.790	0.827	0.818	0.799	0.803	0.744
Min.	0.404	0.440	0.582	0.647	0.417	0.381
Max.	0.883	0.914	0.887	0.862	0.886	0.824

considered as correct, if the incisor teeth of the mean model were overlapping with the corresponding structure in the binary mask. The position of the model after initialization was correct for 13 out of the 14 test cases. The incorrect placement was caused by an asymmetry in the binary mask.

Therefore, the centroid computed on the binary mask was shifted too far to one side causing a wrong overlap. The model was unable to recover from this incorrect initialization during the adaptation process, resulting in a DICE overlap of only 0.38 (cf. Table 1). Scale estimation was accurate enough for all cases to enable a working adaptation. However, in some cases the initial size of the model was too large, so that molar teeth were in between two teeth in the input image, causing an incorrect segmentation of these teeth as a result. Central teeth like incisors or canine were still segmented correctly.

Exemplary segmentation results are depicted in Fig. 2. The final segmentations are compared to gold-standard segmentations and evaluated in terms of precision, recall, accuracy, specificity, f-score and dice overlap. In order to receive meaningful results for specificity and accuracy, the evaluation is done on the minimum bounding box that covers both automatic- and gold-standard segmentation. For a single test instance, first the values for each tooth present in the test instance are computed. Then, the average over all these teeth is computed to get the result for that test instance. Finally, the average is computed over all 14 test instances. Table 1 shows the average values as well as minimum and maximum values for each category.

4 Discussion

The proposed approach combines a coupled shape model with a neural network to robustly segment teeth in panoramic radiographic images. Most state of the art methods rely only on information extracted from the image itself, which is directly influenced by image quality. The coupled shape model, however, employs a priori knowledge about the shape and spatial configuration of the individual teeth to guide the search for suitable image features which helps to handle the poor image quality of dental X-Ray images. The neural network provides a robust binary mask for calculating the necessary information for the initialization of the coupled model. Additionally, the mask is useful for an initial adaptation to ensure a good placement of individual teeth. This way, the parameters of the final adaptation on the intensity image can be chosen specifically to detect the correct tooth contour.



Fig. 3. Difficult cases: bridge and missing teeth (left), broken tooth (middle) and failed segmentation in case of a missing premolar tooth and no gap in this place (right)

Wisdom teeth have not been incorporated into the model at the moment. Their position and shape can vary highly from patient to patient and not all patients have them. There was simply not enough training data available to train a reliable shape model and get a meaningful estimate of the space of possible positions. Wisdom teeth will be added at a later point when sufficient training data is available. The approach is able to handle missing teeth, if the space originally occupied by the missing tooth is still visible so that the mean shape model of that tooth can be placed into the gap and subsequent teeth can be positioned correctly. In case the gap is too small or no longer present, subsequent teeth are labeled incorrectly. Partially missing or broken teeth are labeled correctly. However, due to their unnatural shape, segmentation accuracy is lower since the shape adaptation is limited to valid shapes given by the shape model. Post-processing steps would be required to tackle this problem. Fillings can influence the segmentation accuracy since the denser material appears brighter in X-Ray images and results in stronger gradient features. The contour might be placed incorrectly in cases where fillings do not match the true contour of the teeth. Figure 3 depicts some examples for the aforementioned cases.

Overall, the presented approach performs significantly better than any of the state of the art methods reviewed in [10]. None of the reviewed methods managed to achieve good results in both precision and recall. Watershed-based methods provided a high recall value of 0.816, but only a low precision of 0.478. The opposite is the case for splitting/merging methods with a precision of 0.816 and recall of only 0.081. The best methods according to the f-score value were local-threshold methods with a precision of 0.513 and recall of 0.826. The approach presented in this paper achieves an average precision of 0.790 together with a recall of 0.827.

5 Conclusion

In this paper an automatic approach for teeth segmentation in panoramic radiographic images was presented. It performs better than current state of the art methods and is able to handle difficulties like missing or broken teeth, filling and bridges. On a set of 14 test images the approach achieves an average DICE overlap of 0.744 and precision and recall values of 0.790 and 0.827, respectively.

Future work includes increasing the segmentation accuracy and robustness, extending the coupled shape model by using a larger set of training images and potentially including wisdom teeth into the model. Since the framework itself is generic, it can also be applied to other types of dental X-Ray images.

Acknowledgements. We thank Dr. Werner Betz, Goethe University Frankfurt for providing the panoramic radiographic images used in this work.

References

1. Amer, Y.Y., Aqel, M.J.: An efficient segmentation algorithm for panoramic dental images. *Procedia Comput. Sci.* **65**, 718–725 (2015)
2. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Training models of shape from sets of examples. In: Hogg, D., Boyle, R. (eds.) *BMVC92*, pp. 9–18. Springer, London (1992). https://doi.org/10.1007/978-1-4471-3201-1_2
3. Hasan, M.M., Ismail, W., Hassan, R., Yoshitaka, A.: Automatic segmentation of jaw from panoramic dental x-ray images using GVF snakes. In: *2016 World Automation Congress (WAC)*, pp. 1–6, July 2016
4. Heimann, T., Meinzer, H.P.: Statistical shape models for 3D medical image segmentation: a review. *Med. Image Anal.* **13**(4), 543–563 (2009)
5. Jung, F., Steger, S., Knapp, O., Noll, M., Wesarg, S.: COSMO - coupled shape model for radiation therapy planning of head and neck cancer. In: Linguraru, M. (ed.) *CLIP 2014. LNCS*, vol. 8680, pp. 25–32. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-13909-8_4
6. Lira, P.H., Giralaldi, G.A., Neves, L.A.: Panoramic dental x-ray image segmentation and feature extraction. In: *Proceedings of V Workshop of Computing Vision, Sao Paulo, Brazil* (2009)
7. McInerney, T., Terzopoulos, D.: Deformable models in medical image analysis: a survey. *Med. Image Anal.* **1**(2), 91–108 (1996)
8. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015. LNCS*, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
9. Said, E.H., Nassar, D.E.M., Fahmy, G., Ammar, H.H.: Teeth segmentation in digitized dental x-ray films using mathematical morphology. *IEEE Trans. Inf. Forensics Secur.* **1**(2), 178–189 (2006)
10. Silva, G., Oliveira, L., Pithon, M.: Automatic segmenting teeth in x-ray images: trends, a novel data set, benchmarking and future perspectives. *Expert. Syst. Appl.* **107**, 15–31 (2018)
11. Steger, S., Jung, F., Wesarg, S.: Personalized articulated atlas with a dynamic adaptation strategy for bone segmentation in CT or CT/MR head and neck images. In: *Medical Imaging 2014: Image Processing*, vol. 9034, p. 90341I. International Society for Optics and Photonics (2014)
12. Steger, S., Kirschner, M., Wesarg, S.: Articulated atlas for segmentation of the skeleton from head & neck CT datasets. In: *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1256–1259. IEEE (2012)
13. Wang, C.W., et al.: A benchmark for comparison of dental radiography analysis algorithms. *Med. Image Anal.* **31**, 63–76 (2016)