Multi-object Segmentation of Head Bones

Dagmar Kainmueller[1], Hans Lamecker[2], Heiko Seim[1] and Stefan Zachow[1]

July 20, 2009

[1]Zuse Institute Berlin, Germany
[2]INRIA Sophia Antipolis, France

Abstract

We present a fully automatic method for 3D segmentation of the mandibular bone and cranium from CT data. The method includes an adaptation of statistical shape models of mandible and cranium, followed by graph based simultaneous optimization of adjacent deformable models. For the latter, a correspondence relation between vertices of adjacent surface meshes is established. The adaptation of the models to the image data is performed according to a heuristic model of the typical intensity distribution around the bone boundary, with special focus on an accurate discrimination of adjacent bones in joint regions. The initial positioning of the models in the image data is estimated automatically. An evaluation on 8 CT scans shows that a manual correction of the automatic segmentations is not necessary in at least 58% of the axial slices that contain the mandible.

1 Introduction

In CT data, bony tissue is well-contrasted to surrounding soft tissues. Furthermore, CT scanners are usually calibrated such that the absolute intensities of bony tissue vary only slightly between different scans and patients. Hence simple thresholding with a fixed threshold usually produces reasonable bone segmentations. However, in joint regions, thresholding is often not sufficient for separating adjacent individual bones from each other. Due to large slice distances or pathological changes of the bones, the joint space may be hard to detect even for a human observer. Considering the bones of the human head, another problem is the presence of teeth: First, metallic fillings or braces on teeth cause bright artifacts in the data. Second, the teeth themselves show higher intensities than the surrounding bone, which is problematic if the teeth shall be excluded from the bone segmentation. And last but not least the varying dentition state may pose a problem to an automatic segmentation approach for the mandibular bone or the skull bone.

Prior knowledge about the typical shape of the object of interest (in our case the mandibular bone) may be incorporated into an automatic segmentation approach to constrain the process where the image information is not reliable. The shape may be constrained by a mean shape or reference anatomies [12, 3], a statistical shape model (SSM) [2, 15, 10, 4, 6] or appearance model [1, 13], or more flexible single- or multi-object deformable models [8, 11, 5]. In this work we employ single-object statistical shape models of skull and mandible for a robust initial segmentation, followed by a multi-object deformable model optimization on coupled objects that produces accurate segmentations, especially in joint regions.

2 Statistical Shape Models of Mandible and Cranium

In this paper we apply an SSM of the mandible and another SSM of the cranium to the task of segmenting the mandibular bone from CT data. To build such an SSM, a set of training surfaces has to be reconstructed from medical image data. Point correspondences are established between all training surfaces by a consistent division into sub-regions (patches) on all surfaces, following the approach by Lamecker at al. [9]. Each surface may then be represented in a common vector space $\mathbb{R}^{3m}$, with m being the number of surface points. After alignment of all training surfaces to one reference surface, principal component analysis (PCA) performed on the resulting point cloud yields a linear model $S(\mathrm{b},T) = T(\bar{v}+\sum_k b_k \mathrm{p}_k)$ with $\bar{v} \in \mathbb{R}^{3m}$ representing the mean shape, $p_k \in \mathbb{R}^{3m}$ the modes of shape variation (shape modes), $b_k \in \mathbb{R}$ the shape weights and T an affine transformation. By adapting the model parameters, e.g. guided by image features, the SSM allows for the reconstruction of plausible shapes as a weighted linear combination of all training surfaces.

The SSM of the mandible was generated from 106 CBCT datasets from a previous study [4] (age 16 to 71, 56 female, 50 male). Each dataset has a resolution of 512^3 voxels with an isotropic voxel size of 0.3 mm and a FoV of approx. 15 cm^3. In each CBCT scan the bone was interactively labeled by an experienced dentist. The reconstructed surface meshes were then divided into eight patches (see Fig. 1(a)). For model generation the teeth were ignored due to a strongly varying dentition state. The SSM consists of 8561 surface points.

For additional guidance of our segmentation approach, especially at the temporomandibular joint, we make use of a second SSM of the cranium. Our cranial SSM is generated from 43 CT datasets acquired for maxillofacial therapy planning. After the manual segmentation of all dataset by a user, well trained in the anatomy of the skull, surface meshes of the cranium could be reconstructed. All training surfaces were then reduced to a common subset of anatomical regions, namely the upper jaw, the cheekbone, the orbit, the nasal bone and parts of frontal and temporal bone (cp. Fig. 1(b)). The cranial SSM consists of 46 unique surface patches.

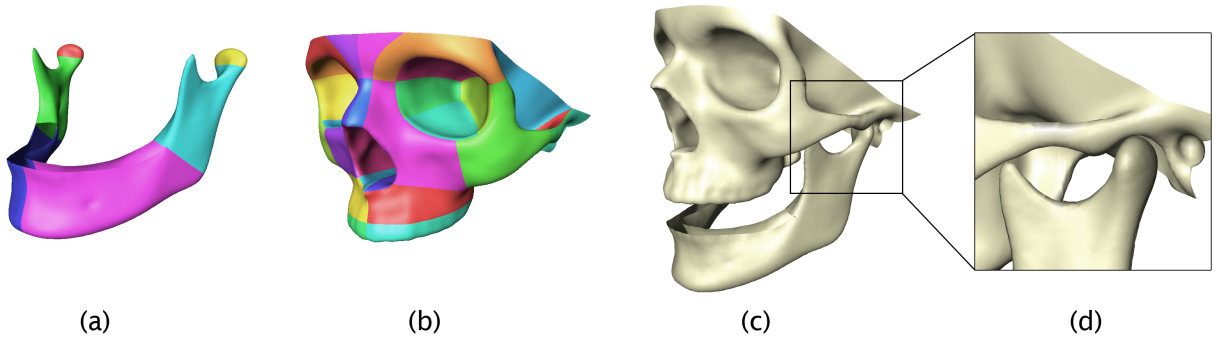

Figure 1: Statistical shape models of the mandible (a), the cranium (b) and a combined visualization to illustrate the anatomy of the temporomandibular joint (c-d).

3 Segmentation Framework

Our fully automatic framework for the segmentation of the mandible in CT data employs the following components: Initialization, image driven SSM adaptation, and graph based optimization of single or coupled objects. A simultaneous segmentation of the cranium is performed to produce accurate segmentations of the temporomandibular joint. This Section describes all steps as well as the overall segmentation algorithm that produced the segmentation results presented in Sec. 4.

3.1 Initialization

The pose initialization of a single object (mandible or skull) in CT data as contained in our framework closely follows a global approach for 3D object detection introduced by Khoshelham [7]. It is based on the Generalized Hough Transform (GHT). For a detailed description, see [14].

3.2 Image Driven SSM Adaptation

Segmentation using the SSM is the task of iteratively finding transformation and shape parameters (b,T) such that the shape $S(b,T)$ approximates the unknown target shape R^* as good as possible. Let $R^i = S(b^i,T^i)$ denote the segmentation in iteration i: A *displacement* vector field ΔR^i is computed that assigns a vector Δr_j to each vertex j of R^i. ΔR^i describes the desired deformation of the model towards R^* in the underlying image data I. Then, both transformation parameters T and shape parameters b are adapted to the target shape $(R^i + \Delta R^i)$, as originally proposed by Cootes et al. [2]. The following paragraph explains how ΔR^i is generated.

3.3 Image Analysis for Displacement Vector Computation

The displacement vector field ΔR for surface R is computed by analyzing 1D *intensity profiles* in the image data $I : \mathbb{R}^3 \to \mathbb{R}$: For each vertex j of R, I is sampled over a length L along the surface normal u_j at vertex position v_j. A cost function $c_j : P_j \to \mathbb{R}_0^+$ is computed on the set of sampling points $P_j = \{v_j^n := v_j + (\frac{n-1}{N_j-1} - 0.5) \cdot L \cdot u_j | n = 1,\ldots,N_j\}$. In SSM adaptation, the displacement vector at vertex j is then defined as $\Delta r_j = v_j^* - v_j$, with $v_j^* = \text{argmin}_{v_j^n} c_j(v_j^n)$, i.e. the minimum cost sample point on a profile serves as a desired (locally optimal) new position for the respective vertex. In the following we drop the index j for clarity. Our framework employs two cost functions: A "standard" cost function c_S, and a "conservative" cost function c_C. We define $c_S(v^n) =$

$$(2i+1)\left(\frac{-g}{dI(v^n)} + 2\frac{|m-n|}{N}\right) \text{ if (2)}, \quad 7\left(\frac{-g}{dI(v^n)} + 2\frac{|m-n|}{N}\right) \text{ if (3)}, \quad 30 + 2\frac{|m-n|}{N} \text{ else}, \tag{1}$$

with

$$I(v^n) \in [t+iw, t+(i+1)w] \wedge dI(v^n) < -g, i = 0,1,2 \tag{2}$$

$$I(v^n) \in [t, t+3w] \text{ and } dI(v^n) < -0.5 \cdot g \tag{3}$$

Here, t defines an intensity threshold, w an intensity window width, and g a threshold for gradient magnitude. $dI(v^n)$ denotes the directional derivative of I along u, and $m \in \left\{1, \frac{N+1}{2}, N\right\}$ is a "preferred sample index", i.e. sample points closer to v_m get slightly lower costs. c_C is defined as $c_C(v^n) = c_S(v^n)$ if there are two distinct intervals of sample points with $c_S(v^n) < 30$, and at least one sample point in between with $dI(v^n) > g_{min}$, and $c_C(v^n) = c_I(v^n)$ otherwise. where c_I is defined as $c_I(v^n) = 0$ if $n = \frac{N+1}{2}$, and $c_I(v^n) = 100$ otherwise. Thus, features are ignored if only one downward gradient (from bone to background) is found along the profile. This makes sense for the mandibular condyles: A downward gradient is sometimes barely present at the mandible surface, but only inside the skull. Note that we determined all cost functions in a heuristic manner. None of the parameters were automatically learned from the training images. The cost functions are specifically designed to segment bones in CT data, especially in joint regions.

3.4 Graph Based Multi-object Optimization

To obtain very accurate segmentations, a fine grain segmentation step is necessary after SSM adaptation to be able to leave the shape space spanned by the SSM. For this purpose, graph cut algorithms allow for a *global* optimization of the sum of costs for each vertex displacement while respecting hard constraints on surface smoothness and optionally on the distance between multiple surfaces. For the latter, multiple surfaces must be coupled with *shared intensity profiles* at individual vertices. For more details on graph construction see [11]. In [5], we proposed a construction algorithm for shared intensity profiles on pairs of adjacent triangular surface meshes P and Q that yields a bijective mapping between *coupled patches* on both meshes. In the process, the connectivity of parts of the original meshes is modified. For more details, see [5].

3.5 Overall Segmentation Algorithm

The segmentation algorithm consists of a series of steps combining the methods presented above and employing the SSMs of mandible and cranium as described in Sec. 2. First, each SSM is initialized via the GHT. For the specific parametrization, see [14]. If the GHT for the mandible yields a transformation such that the mandibular condyles lie outside the range of the image data, we use this transformation also for initializing the skull, as the skull initialization might fail in this case due to the missing image data. Otherwise a separate GHT is performed for the skull.

After initialization, first the skull SSM is adapted to the image data. The resulting approximate segmentation of the skull is then used for mandible SSM adaptation as a region where costs are uniformly high. This prevents the condyles of the mandible SSM from adapting to skull features in the joint region. SSM adaptation for both mandible and skull model starts with profile length $L := 20$mm while adapting 10 modes of shape variation, and gradually reduces L to 6mm and increases the number of considered shape modes to 80 for the mandible and all 42 for the skull. Sample points on the profiles are placed equidistantly every 0.5mm. For both SSMs, cost function c_S is employed for all surface patches.

The surface models resulting from SSM adaption are coupled with shared displacement directions of length 18mm. Graph based optimization of the coupled surfaces is performed with minimum distance constraint 0mm, maximum distance constraint 15mm, and shape preserving constraint 1mm For the coupled patches of the mandible we use cost function c_C. For all other surface patches of both mandible and skull, cost function c_S is employed.

4 Results

A comparison to expert segmentations was performed on 8 clinical CT datasets acquired for radio therapy planning at the Princess Margaret Hospital in Toronto. Each dataset contains the full neck region, the lower part of the head and the upper thorax. The slice resolution slightly varies between 0.8^2 to $0.9^2 mm$ with a slice distance of $2mm$. Our segmentation method was applied to all datasets on an Intel DualCore 2.66GHz taking aprox. 15 minutes per dataset. The automatic segmentation results of our approach were then compared to the expert segmentations in terms of their Hausdorff distance (HD) and their volumetric overlap (VO) (cp. Table 1). Both measures were evaluated slice-wise on all slices where a manual segmentation was available. Screenshots of an example segmentation (Dataset 12) are can bee seen in Figure 2.

The mean and medium HD lie in a range of about 4.5 to $15.6mm$ and 2.2 to $5.3mm$, respectively. The maximum error can be found for the segmentation of dataset 14. This is also the dataset where a majority

	Hausdorff distance (mm)			Volume overlap		
Dataset	**Average** per slice	**Median** per slice	**No. of slices** (> 3 mm)	**Average** per slice	**Median** per slice	**Total volume**
11	6.94	2.18	39 (8)	78.4 %	82.9 %	84.3 %
12	7.44	3.09	40 (26)	81.2 %	84.4 %	85.9 %
13	4.51	2.76	35 (11)	88.1 %	88.7 %	90.8 %
14	15.61	5.34	34 (26)	70.2 %	73.8 %	81.0 %
15	11.44	2.49	37 (14)	84.7 %	89.7 %	88.1 %
16	10.41	2.76	35 (13)	80.3 %	84.3 %	88.2 %
17	6.16	2.93	43 (20)	84.6 %	85.2 %	88.0 %
18	6.60	2.50	37 (9)	84.4 %	88.7 %	89.3 %

Table 1: Hausdorff distance and volumetric overlap for mandible segmentation in the testing datasets.

of the slices (26 of 34) show a HD larger than 3mm, which is often a critical value for a segmentation to be accepted by clinicians. All other datasets reveal significantly smaller errors with a minimum medium HD of 2.2*mm*. The VO exposed a similar tendency: for all datasets an average VO larger than 82.9% was calculated, except for dataset 14, where an average VO of 72.8% can be oberved. The total VO value is, in all cases, larger than the average or the median VO.

In addition to this evaluation, we compared our automatic segmentation results to a second set of manual segmentations generated a student experienced in bone segmentation. Here, we computed three 3D surface distance measures: The 3D Hausdorff, roots mean square, and mean distance. The average results are 0.3 mm, 0.7 mm and 5.8 mm.

5 Discussion

At the time writing this paper the ground truth segmentations of the results presented in Table 1 were not known to the authors. Therefore, the following discussion is based on the quantitative result tables provided by the organizers of the workshop *MICCAI 2009 3D Segmentation Challenge for Clinical Applications* as well as qualitative evaluations on previously provided pre-segmented training datasets.

According to the HD and the VO it is obvious that dataset no. 14 is the worst case regarding the segmentation quality. We mainly attribute this to an under-segmentation of the bone at the front teeth row in multiple slices. In such a case, the slice-wise evaluation of the HD quickly leads to large deviations, since the maximum of the minimal distances is computed between the front and the back of the mandible. This behavior, i.e. the disregard of the z-dimension, might also be the reason for the relatively large number of slices with a HD larger than 3*mm*. By looking at the volume overlap two characteristic properties can be observed: (1) The total VO is larger than the average and the mean VO and (2) the mean VO is larger than the average VO. The first fact might be caused by very large VO errors for slices with only few bone voxels but large differences of manual to automatic segmentations, e.g. at the mandibular condyles. Such slices have less influence on the total VO, which is the only measure evaluating the 3D volume as a whole. The reason for the mean VO being larger than the average VO indicates that there are a few slices with a very low VO reducing the average VO. Such outliers are ignored by a mean metric.

Another behavior, which we observed in the previously provided training datasets, is a slight under-segmentation of the automatic method compared to ground truth (see Fig. 2). The calculation of the cost for

SSM and graph-based adaptation is mainly guided by the gradient magnitude, whereas a human observer might justify a segmentation primarily on a gray value threshold. This argument is supported by the fact that the ground truth segmentations of the test training datasets tend to lie outside the automatically obtained results of our method. Getting closer to the manual segmentation requires a higher influence of the threshold within the evaluation of the cost-function. However, for a full insight on the *under-segmentation issue* it would be valuable to have additional expert segmentations available. Those could also be used to determine the inter-observer variability, a useful measure when it comes to determine the practical value of a method.

5.1 Conclusion

We presented an algorithm for automatic segmentation of the mandibular bone in CT data. Its main components are global initialization, statistical shape adaptation and constrained multi-object free-form deformation, which can directly be applied to other segmentation problems as well. The computation of the cost function that guides model deformations was designed to solve the general problem of segmenting bones in CT data, with special attention to accurate discrimination of adjacent bones in joint regions. Results on eight test CTs with ground truth segmentations as provided by the *MICCAI 2009 3D Segmentation Challenge for Clinical Applications* imply that on average no manual corrections are necessary on at least 58% of the axial slices that contain the mandibular bone. However, to put the error measures into perspective, it is indispensable to evaluate a second manual segmentation by comparison to the above mentioned ground truth.

As future work, an articulated, composite statistical shape model of mandible and skull shall be established. Such a composite model with joint flexibility presumably makes the SSM adaptation process more robust, as both adjacent bones are adapted simultaneously while allowing only plausible joint postures. The segmentation of the mandibular condyle would not depend on a previous segmentation of the mandibular fossa.

The initiative of providing a common pool of test data as well as well-defined measures for evaluation is indispensable for further progress in the field of automatic image segmentation. Yet it remains difficult to establish the correct ground truth, due to inter- and intra observer variations in manual segmentations. At least as difficult is the definition of decisive error measures for evaluation. E.g. the slice based Hausdorff metric, while giving a hint which slices have to be corrected manually, may be misleading if multiple connected components are to be segmented in a slice and only one component - possibly just one pixel - is missing in the automatic segmentation.

Acknowledgments

D. Kainmueller is supported by DFG Collaborative Research Centre SFB 760. H. Seim s supported by the European Commission under the FP6 IST Project DeSSOS (027252). Thanks to Jana Malinowski for providing additional manual segmentations of the mandibular bone.

References

[1] T. F. Cootes and C. J. Taylor. Statistical models of appearance for medical image analysis and computer vision. volume 4322, pages 236–248. SPIE, 2001. 1

[2] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active Shape Models - Their Training and Application. *Comput. Vis. Image Underst.*, 61(1):38–59, 1995. 1, 3.2

[3] S. Faisan, N. Passat, V. Noblet, R. Chabrier, J.-P. Armspach, and C. Meyer. Segmentation of head bones in 3-D CT images from an example. In *ISBI*, pages 81–84. IEEE, 2008. 1

[4] D. Kainmueller, H. Lamecker, H. Seim, M. Zinser, and S. Zachow. Automatic Extraction of Mandibular Nerve and Bone from Cone-Beam CT Data. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2009. 1, 2

[5] D. Kainmueller, H. Lamecker, S. Zachow, and H.-C. Hege. Coupling Deformable Models for Multi-object Segmentation. In *International Symposium on Computational Models for Biomedical Simulation (ISBMS)*, pages 69–78, 2008. 1, 3.4

[6] D. Kainmueller, H. Lamecker, S. Zachow, and H.-C. Hege. An Articulated Statistical Shape Model for Accurate Hip Joint Segmentation. In *Accepted at the IEEE Conference on Engineering in Medicine and Biology (EMBC)*, Minnesota, Minneapolis, USA, Sept. 2009. 1

[7] K. Khoshelham. Extending Generalized Hough Transform to Detect 3D Objects in Laser Range Data. In P. Rönnholm, H. Hyyppä, and J. Hyyppä, editors, *Proc. ISPRS Workshop Laser Scanning and SilviLaser*, volume 36 of *IAPRS*, pages 206–211, 2007. 3.1

[8] T. Klinder, J. Ostermann, M. Ehm, A. Franz, R. Kneser, and C. Lorenz. Automated model-based vertebra detection, identification, and segmentation in ct images. *Medical Image Analysis*, 13(3):471 – 482, 2009. 1

[9] H. Lamecker, M. Seebaß, H.-C. Hege, and P. Deuflhard. A 3D statistical shape model of the pelvic bone for segmentation. In J.M. Fitzpatrick and M. Sonka, editors, *Proceedings of SPIE - Volume 5370 Medical Imaging 2004: Image Processing*, pages 1341–1351, May 2004. 2

[10] H. Lamecker, S. Zachow, A. Wittmers, B. Weber, H.-C. Hege, B. Elsholtz, and M.Stiller. Automatic segmentation of mandibles in low-dose CT-data. *Int. J. Computer Assisted Radiology and Surgery*, 1(1):393–395, 2006. 1

[11] K. Li, S. Millington, X. Wu, D. Z. Chen, and M. Sonka. Simultaneous Segmentation of Multiple Closed Surfaces Using Optimal Graph Searching. In *Information Processing in Medical Imaging (IPMI)*, pages 406–417, 2005. 1, 3.4

[12] M. Lilja, V. Vuorio, K. Antila, H. Setälä, J. Järnstedt, and M. Pollari. Automatic Segmentation of the Mandible from Limited-Angle Dental X-Ray Tomography Reconstructions. In *International Symposium on Biomedical Imaging (ISBI)*, pages 964–967, 2007. 1

[13] S. Rueda, J. A.Gil, R. Pichery, and M. Alcañiz. Automatic segmentation of jaw tissues in ct using active appearance models and semi-automatic landmarking. volume 9, pages 167–74, 2006. 1

[14] H. Seim, D. Kainmueller, M. Heller, H. Lamecker, S. Zachow, and H.-C. Hege. Automatic Segmentation of the Pelvic Bones from CT Data Based on a Statistical Shape Model. In *Eurographics Workshop on Visual Computing for Biomedicine (VCBM)*, pages 93–100, Delft, Netherlands, 2008. 3.1, 3.5

[15] S. Zachow, H. Lamecker, B. Elsholtz, and M. Stiller. Reconstruction of mandibular dysplasia using a statistical 3D shape model. In *Proc. Computer Assisted Radiology and Surgery (CARS)*, pages 1238–1243, Berlin, Germany, June 22-25 2005. 1

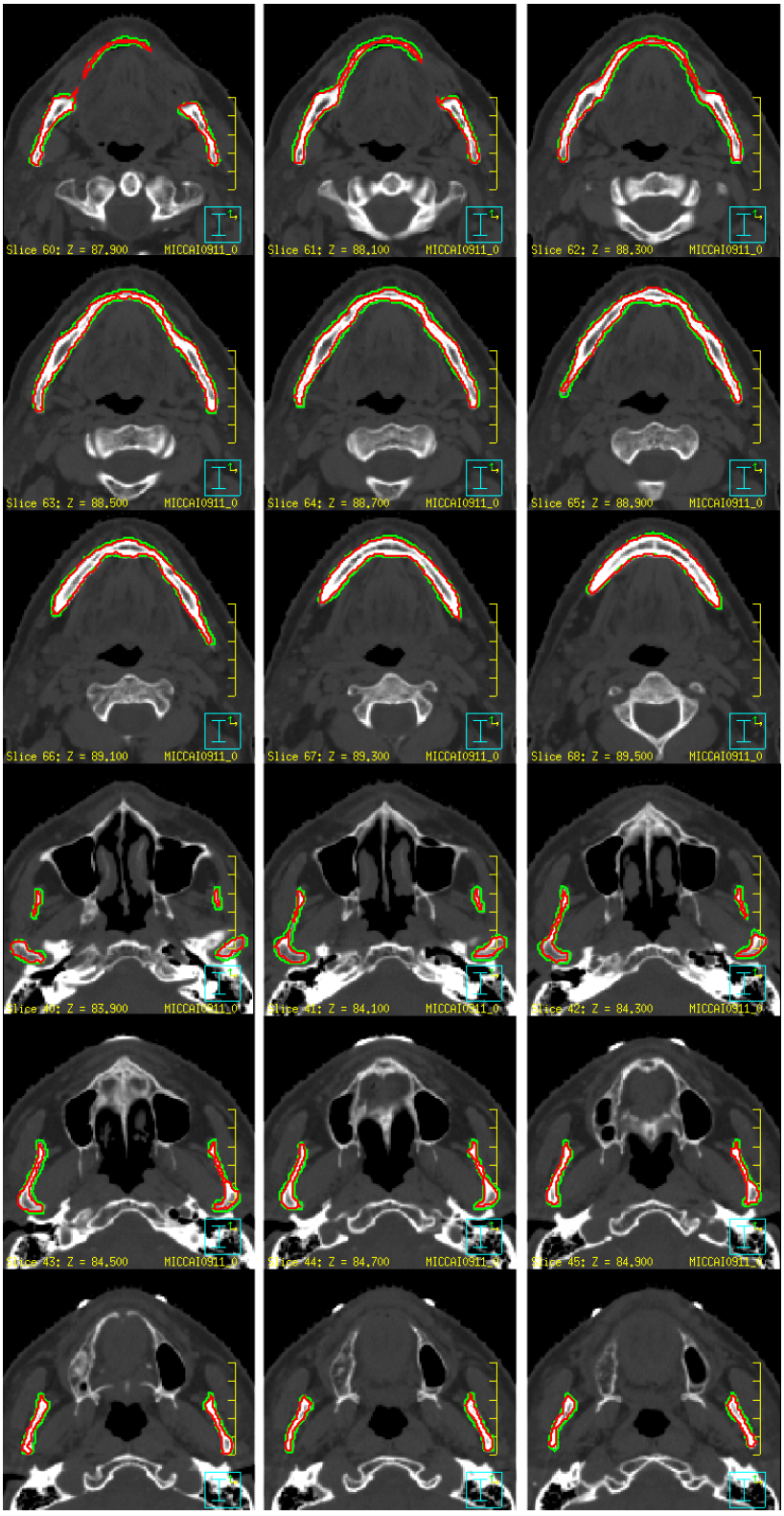

Figure 2: Comparison of manual (green contour) vs. automatic segmentation (red contour) for one dataset.