Automatic Segmentation of Parotids in Head and Neck CT Images using Multi-atlas Fusion

Xiao Han, Lyndon S. Hibbard, Nicolette P. O'Connell, and Virgil Willcut

Elekta CMS Software, 13723 Riverport Drive, St. Louis, MO 63043, USA
{Xiao.Han,Lyn.Hibbard,Nicole.Oconnell,Virgil.Willcut}@elekta.com

Abstract. Treatment planning for high precision radiotherapy of head and neck (H&N) cancer patients requires accurate delineation of many critical structures. Manual contouring is tedious and often suffers from large inter- and intra-rater variability. To reduce manual labor, we have previously developed a fully automated, atlas-based method for H&N CT image segmentation [1, 2]. In this work, we adapt the previous method and apply it to tackle the parotid segmentation problem in the MICCAI 2010 H&N Segmentation Challenge. The proposed method applies a hybrid deformable image registration to map parotid labels from an atlas image to the subject, the result of which is then refined using a deformable surface model approach. Segmentation fusion using multiple-atlases is also employed to further improve the segmentation accuracy. Validation results on eight clinical datasets distributed by the MICCAI workshop showed that the proposed method gave accurate segmentation results, with a volume overlap above 85% for most subjects.

1 Introduction

Treatment planning for high precision radiotherapy of head and neck (H&N) cancer patients requires accurate delineation of target volumes and OARs on planning computed tomography (CT) images. Manual contouring is tedious and time-consuming, and often suffers from large intra- and inter- rater variability. Tools for automated segmentation are thus needed.

We have previously developed a fully automated atlas-based segmentation method for H&N CT images [1, 2], which consists of a hierarchical atlas registration method and a multiple atlas fusion strategy. The method was shown to produce very accurate results for the segmentation of the mandible and the brainstem [2], as well as some other structures [1]. In this work, we adapt the previous method to the segmentation of the left and right parotids, as required by the MICCAI 2010 H&N Auto-Segmentation Challenge. In particular, a different atlas registration method is used here to account for the limitation of the training data (i.e., only parotids labels are available for each atlas). The rest follows a similar strategy as our earlier work: we apply a deformable surface model to improve the segmentation results of individual atlases and use multi-atlas fusion to achieve overall better accuracy. In the following, we first briefly summarize the multi-atlas segmentation framework and then describe the atlas registration

and segmentation refinement methods in more details. Finally, we present the validation results based on the test datasets provided by the workshop.

2 Method

2.1 Overview

The overall framework stays the same as in the previous work [2], where the segmentation of a new subject is computed by applying multiple atlases separately and then combining the individual segmentation results. The STAPLE (*Simultaneous Truth And Performance Level Evaluation*) algorithm, as originally introduced by Warfield et al. [3], is applied as the segmentation fusion method.

The multi-atlas fusion strategy still requires a reliable atlas registration and segmentation method in order to get good segmentation result for each individual atlas. In this work, we apply a hybrid non-linear image registration method to align the subject to a chosen atlas, following a global linear registration. The mapped structure labels are then refined by a deformable model method to further improve the accuracy. The linear registration method applied here is exactly the same as in [2]. Thus, in the next we only present the new non-linear image registration algorithm and the deformable model-based parotid refinement.

2.2 Dense Hybrid Deformable Registration

The goal of the deformable registration is to align details between the atlas and the subject images, given an initial linear registration. We use here a non-parametric transformation model, where the image transformation T is modeled directly as a vectorial displacement field $\mathbf{U}$, such that $T(\mathbf{x}) = \mathbf{x} + \mathbf{U}(\mathbf{x})$ for every image point $\mathbf{x}$. In order to be able to handle intensity contrast changes that often exist between CT images of different subjects in the parotid region, we have designed a hybrid image matching metric to be used at this step. This hybrid metric is a combination of the popular MI metric and a new normalized-sum-of-squared-differences (NSSD) metric:

$$\mathcal{J}_H(I, J, \mathbf{U}) = -\mathrm{MI}(I, J, \mathbf{U}) + w \cdot \mathrm{NSSD}(\tilde{I}, \tilde{J}, \mathbf{U}), \tag{1}$$

where I and J denote the atlas and the subject images respectively, and w denotes a relative weighting of the two terms. From our experience, this hybrid similarity measure provides better image alignment than using the MI metric alone since the latter cannot account for local image contrast changes. Thus, this new metric tends to improve the segmentation accuracy for all structures comparing to the MI-only method of [2], not just limited to the parotids.

Definition of the MI metric is skipped here, which can be found in [4, 5]. The NSSD-term is an edge-based alignment metric, which is defined as follows:

$$\mathrm{NSSD}(\tilde{I}, \tilde{J}, \mathbf{U}) = \frac{1}{N} \sum_{\mathbf{x}} \|\tilde{I}(\mathbf{x}) - \tilde{J}(T(\mathbf{x}))\|^2, \tag{2}$$

where

$$\tilde{I}(\mathbf{x}) = \frac{I(\mathbf{x}) - \mu_I(\mathbf{x})}{\sigma_I(\mathbf{x})}, \tag{3}$$

and $\mu_I = G_\sigma * I$ and $\sigma_I^2 = G_\sigma * (I - \mu_I)^2$ denote the local intensity mean and local intensity variation for image I. G_σ denotes a Gaussian filter with kernel size σ (the kernel size σ is chosen to be two times the image voxel size in all later experiments). Similar notations hold for image J. We call $\tilde{I}$ and $\tilde{J}$ the *normalized local offset* images as implied by Eq. (3).

The solution for the optimal deformation field can be found using a compositive, explicit local search scheme similar to what we initially proposed in [5]. In particular, the deformation field $\mathbf{U}$ is updated iteratively according to the following equation:

$$\mathbf{U}^n = \mathbf{U}^{n-1} \circ (\mathbf{Id} + \mathbf{u}^n) + \mathbf{u}^n; \tag{4}$$

or equivalently,

$$\mathbf{U}^n(\mathbf{x}) = \mathbf{U}^{n-1}(\mathbf{x} + \mathbf{u}^n(\mathbf{x})) + \mathbf{u}^n(\mathbf{x}), \ \forall \mathbf{x}. \tag{5}$$

In Eq. (4), $\mathbf{Id}$ denotes the identity transformation, "$\circ$" denotes transformation composition, and $\mathbf{u}^n$ is a local update field that optimizes the hybrid image metric at each iteration step.

In this work, the update field in Eq. (4) is computed as the average of two separate update fields, one for each of the two similarity terms. Consider first the NSSD-term. Given the previous estimation of the deformation field $(\mathbf{U}^{n-1})$, an update field $(\mathbf{u}^n)$ can be found that minimizes the NSSD metric between $\tilde{I}$ and $\tilde{J}^n = \tilde{J} \circ (\mathbf{Id} + \mathbf{U}^{n-1})$:

$$\mathrm{NSSD}(\tilde{I}, \tilde{J}^n \circ (\mathbf{Id} + \mathbf{u}^n)) = \frac{1}{N} \sum_{\mathbf{x}} \|\tilde{I}(\mathbf{x}) - \tilde{J}^n(\mathbf{x} + \mathbf{u}^n(\mathbf{x}))\|^2. \tag{6}$$

Since Eq. (6) is a summation of N-independent terms, the optimal update $\mathbf{u}^n(\mathbf{x})$ at each atlas image location $\mathbf{x}$ can be found separately. In particular, we search the local neighbors of $\mathbf{x}$ in the deformed subject image $\tilde{J}^n$ and find the neighbor $\mathbf{x}'$ that minimizes $\|\tilde{I}(\mathbf{x}) - \tilde{J}^n(\mathbf{x}' = \mathbf{x} + \mathbf{u}^n(\mathbf{x}))\|^2$. This gives the update field corresponding to the NSSD-term. A second update field can be computed similarly that optimizes the MI-term, as initially described in [5]. We then average the two update fields as mentioned above to get the final update field for the whole hybrid image metric, which is further smoothed with a spatial Gaussian filter. The regularized update field is applied to update the total deformation field according to Eq. (4), and the result is regularized using another Gaussian filter. The iteration is repeated until a user-specified number of steps or until the hybrid metric in Eq. (1) stops decreasing.

This dense deformable registration is the most time-consuming part of the whole atlas-based segmentation method. To improve the computation efficiency and make multi-atlas segmentation feasible in practice, we have implemented this deformable registration algorithm on NVIDIA GPUs using the NVIDIA CUDA programming model. As we have observed, the GPU-based implementation easily offers a speed-up of more than 25× comparing against a modern CPU.

2.3 Refinement of Parotid Segmentation using Deformable Surface Model

In the last step of parotid segmentation using a single atlas, a deformable surface model method is applied to improve the initial segmentation result as produced by atlas registration and label mapping. Refinements for the left and right parotids are computed separately at this step.

Mathematically, we denote the initial parotid surface (either left or right) by a vector-valued parametric function $\mathbf{x}_0(r,s) = [x_0(r,s), y_0(r,s), z_0(r,s)]$, which gives the position of each surface point (indexed by (r,s)[1]) in the subject image's coordinate system. We model the surface deformation using a surface deformation field $\mathbf{d}(r,s)$, such that $\mathbf{x}(r,s) = \mathbf{x}_0(r,s) + \mathbf{d}(r,s)$ gives the final location of each surface point. The surface refinement is then formulated as finding the optimal solution, $\mathrm{d}_{\mathrm{opt}}$, that minimizes the following energy functional:

$$E(\mathbf{d}) = \int\int \exp(-\|\nabla P(\mathbf{x}(r,s))\|^2) dr ds + \int\int \|\nabla \mathbf{d}(r,s)\|^2 dr ds, \qquad (7)$$

where $P(\mathbf{x}) \propto \exp(-(J(\mathbf{x}) - \mu_p)^2/\sigma_p^2)$ measures the probablity that a point $\mathbf{x}$ of the subject image J belongs to the parotid region. The parotid intensity mean μ_p and variance σ_p^2 are estimated based on the initial segmentation result. In the above equation, the first term drives the surface to the boundary of the estimated parotid region and the second term enforces the regularity or smoothness of the surface displacement field. Regularizing the surface deformation instead of the surface itself allows capturing fine details of the object shape whereas at the same time enforcing the deformed surface to stay close to the initial shape. The solution is computed iteratively with a gradient-descent scheme. We limit the refinement to a local neighborhood of the initial segmentation and this step usually takes less than ten seconds. An example result is illustrated in Fig. 1.

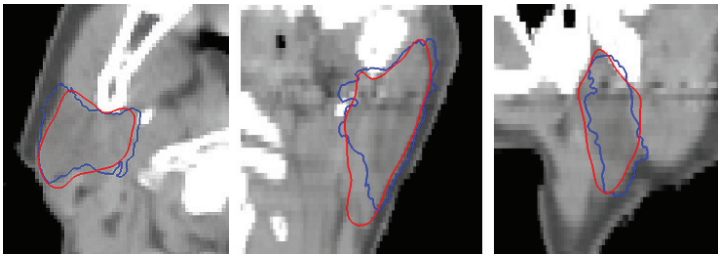

Fig. 1. Illustration of deformable model based refinement of parotid segmentation. The images show three cross-sections for one subject. The blue curves are the initial results before refinement and the red curves are the final results.

[1] The parameters (r,s) simply indicate that the surface is intrinsically a two-dimensional object. In practice, we approximate the surface using a triangulated surface mesh and each surface node is indexed by a unique number.

3 Results and Discussions

A total of 18 datasets were provided by the MICCAI 2010 H&N Auto-Segmentation Challenge workshop at the offline competition stage. Among the 18 datasets, 10 were provided as the training data, for which expert manual segmentations of the parotids were made available. We used the 10 training datasets as the atlases and applied the proposed method to process the remaining 8 test datasets. As described in Section 2, for each test dataset, 10 individual segmentation results were first computed using each of the 10 atlases, and the results were then fused using the STAPLE algorithm to get the final auto-segmentation result. The fusion was performed independently for the left and the right parotids. The final auto-segmentation results for all test subjects were submitted to the organizers of the workshop and evaluated against manual expert segmentation.

All experiments were carried out on a HP xw8400 desktop computer equipped with an Intel Xeon Quad-core 2.66 GHz CPU and a NVIDIA GTX 280 graphics card. The computation time was about 1 minute for each dataset if a single atlas is used and the total computation time was about 10 minutes for 10 atlases.

Fig. 2 illustrates the segmentation results for the left and right parotids for one test dataset. The figures were generated by an independent reviewer (an organizer of the workshop). As we can see from the figures, the auto-segmentation results match the manual segmentation pretty well except for slices that are corrupted by streaking artifacts.

For quantitative evaluation, the auto- and manual- segmentations were compared in a slice-by-slice fashion for each dataset and the following quantitative measures were computed for each slice: the symmetric Hausdorff distance (HD) and the Dice similarity coefficient. A volumetric Dice similarity coefficient was also computed for each dataset to assess the overall volume overlap. Details of these evaluation criteria and their computation can be found in [6].

The overall statistics of the quantitative measures are summarized in Tables 1-2 for the left parotid and Tables 3-4 for the right parotid. From these tables, it can be seen that the median slice-wise Hausdorff distance mostly ranges between 4–6 mm. The median slice-wise Dice coefficient is mostly above 0.85 for both the left and the right parotids. The total volume overlap for most subjects is also close to or above 0.85. Thus, our method provided quite accurate segmentation for this soft-tissue structure. We note that the dataset No. 13 has a relatively low accuracy, which is due to the fact that No. 13 has a truncated field of view and the parotid region is not fully contained in the image. The accuracy for the dataset No. 17 is also low, which is because the parotids of this subject are much smaller than all the atlas subjects.

We note that the segmentation accuracy is lower than our previous results on the mandible and the brainstem [2]. This is largely due to the fact that the parotid is a structure with low contrast in typical H&N CT images. In addition, the parotid region is often corrupted by the streaking artifacts, as observed in many of the provided datasets. The artifacts (in either the atlas or the subject image) make the atlas registration more difficult and also limit the accuracy of the model-based refinement.

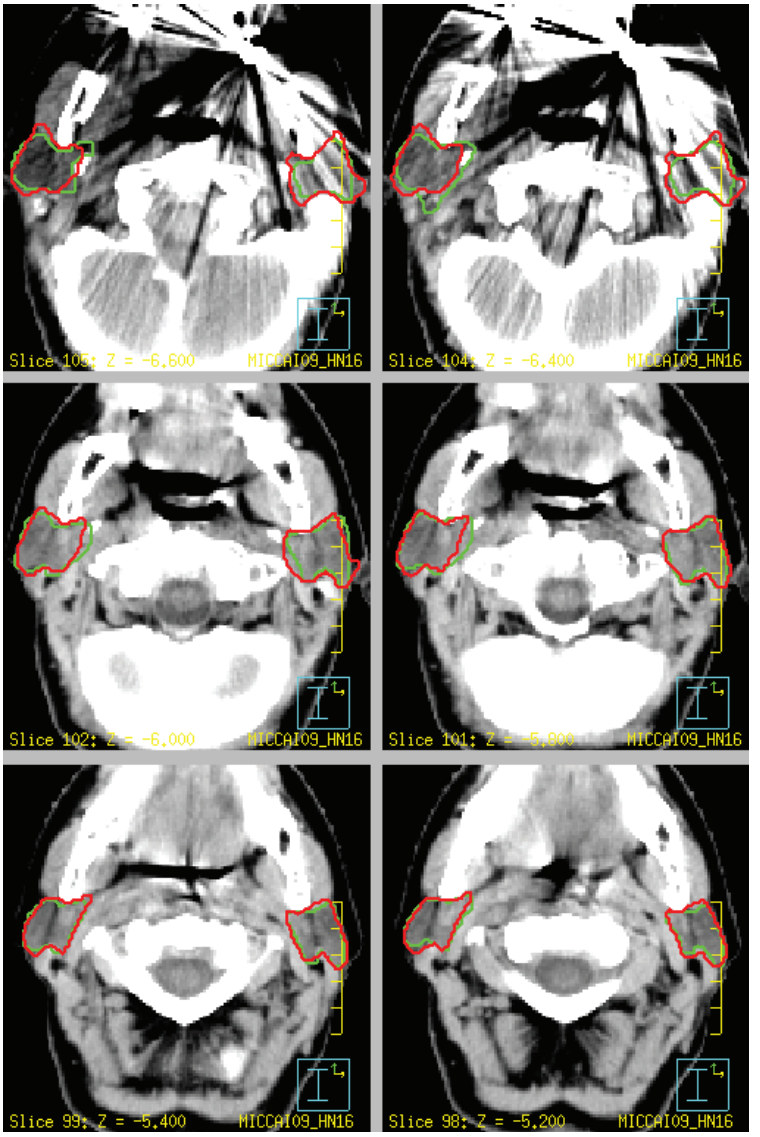

Fig. 2. Snapshots of the parotid segmentation result for one test subject. Red curves indicate the auto-segmentation result and green curves are the corresponding manual segmentation.

Dataset No.	Mean HD	Median HD	No. of slices (HD > 3 mm)
11	5.20	4.02	34 (31)
12	5.45	4.63	28 (23)
13	6.89	6.02	26 (26)
14	7.33	6.73	24 (22)
15	5.78	4.97	23 (21)
16	4.48	4.03	29 (23)
17	6.93	5.48	33 (31)
18	5.41	4.34	24 (20)

Table 1. Hausdorff distance (HD) statistics for left parotid segmentation in the testing datasets.

Dataset No.	Average slice OV	Median slice OV	Total volume OV
11	83.4 %	86.9 %	87.0 %
12	87.0 %	89.1 %	88.2 %
13	79.0 %	81.7 %	76.3 %
14	82.7 %	84.1 %	86.6 %
15	85.0 %	89.5 %	89.8 %
16	82.1 %	85.2 %	84.4 %
17	77.6 %	81.6 %	83.4 %
18	80.8 %	86.8 %	85.4 %

Table 2. Overlap (OV) statistics for left parotid segmentation in the testing datasets.

The last column in Table 1 and Table 3 shows the number of slices with a Hausdorff distance greater than 3 mm for the left and the right parotid segmentation results respectively. Typically, a Hausdorff distance greater than 3 mm means that the auto-segmentation results need be edited before they can be used clinically [6]. From the tables, it is seen that most slices still require some editing. But from our experience, contour editing usually takes much less time than drawing contours from scratch. Thus, the auto-segmentation method can still help reduce manual contouring time in general. In addition, it has been observed that editing a common auto-segmentation result usually helps improve the consistency of contouring by different users. Of course, the current data is insufficient to validate any of these claims.

4 Conclusion

In this work, we have developed an atlas-based method for fully automatic segmentation of parotids in H&N CT images. From the experimental results generated for the MICCAI 2010 H&N Auto-segmentation Challenge workshop, it was shown that the method offered good accuracy on real clinical data, as indicated by a mean volume Dice coefficient above 0.85 for most cases. On the other hand, automatic parotid segmentation is still a challenging problem due to the

Dataset No.	Mean HD	Median HD	No. of slices (HD > 3 mm)
11	6.16	6.25	32 (29)
12	6.87	5.69	30 (30)
13	6.42	5.78	26 (25)
14	6.31	5.25	24 (24)
15	7.06	5.27	26 (24)
16	3.57	3.09	31 (19)
17	4.81	3.91	27 (23)
18	4.31	3.73	25 (18)

Table 3. Hausdorff distance (HD) statistics for right parotid segmentation in the testing datasets.

Dataset No.	Average slice OV	Median slice OV	Total volume OV
11	85.7 %	89.4 %	88.1 %
12	82.4 %	85.6 %	84.9 %
13	83.4 %	87.5 %	82.8 %
14	81.4 %	85.3 %	85.3 %
15	83.4 %	88.2 %	88.1 %
16	86.2 %	87.4 %	86.8 %
17	77.3 %	84.8 %	83.8 %
18	82.8 %	88.7 %	88.2 %

Table 4. Overlap (OV) statistics for right parotid segmentation in the testing datasets.

usual presence of streaking artifacts in this region and the relatively low image contrast. Future work will further improve the accuracy by studying adaptive atlas selection, and investigating the use of statistical shape model for parotid refinement.

References

1. Han, X., Hoogeman, M., Levendag, P., Hibbard, L., Teguh, D., Voet, P., Cowen, A., Wolf, T.: Atlas-based auto-segmentation of head and neck CT images. In: MICCAI 2008, Part II. LNCS 5242, Springer-Verlag, Heidelberg (2008) 434–441
2. Han, X., Hibbard, L., O'Connell, N., Willcut, V.: Automatic segmentation of head and neck CT images by GPU-accelerated multi-atlas fusion. In: The MIDAS Journal – H&N Auto-Segmentation Challenge, http://hdl.handle.net/10380/3136 (2009)
3. Warfield, S., Zou, K., Wells, W.: Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. IEEE Trans. Med. Imag. **23**(7) (2004) 903–921
4. Viola, P., Wells, W.: Alignment by maximization of mutual information. Int. J. Comput. Vision **24**(2) (1997) 137–154
5. Han, X., Hibbard, L., Willcut, V.: GPU-accelerated, gradient-free MI deformable registration for atlas-based MR brain image segmentation. In: MMBIA 2009: IEEE Workshop on Mathematical Methods in Biomedical Image Analysis. (2009)
6. Pekar, V.: http://www.grand-challenge2010.ca/Evaluation.pdf.