A Multi-atlas Approach for the Automatic Segmentation of Multiple Structures in Head and Neck CT Images

Antong Chen[1], Benoit M. Dawant[2]

[1]Applied Mathematics and Modeling, Merck & Co., Inc., West Point, PA 19422, U.S.A.
[2] Department of Electrical Engineering and Computer Science, Vanderbilt University, 2201 West End Ave, Nashville, TN 37235, U.S.A.

Abstract. A multi-atlas approach is proposed for the automatic segmentation of nine different structures in a set of head and neck CT images for radiotherapy. The approach takes advantage of a training dataset of 25 images to build average head and neck atlases of high-quality. By registering patient images with the atlases at the global level, structures of interest are aligned approximately in space, which allowed multi-atlas-based segmentations and correlation-based label fusion to be performed at the local level in the following steps. Qualitative and quantitative evaluations are performed on a set of 15 testing images. As shown by the results, mandible, brainstem and parotid glands are segmented accurately (mean volume DSC>0.8). The segmentation accuracy for the optic nerves is also improved over previously reported results (mean DSC above 0.61 compared with 0.52 for previous results).

1 Introduction

Modern radiotherapies for patients with head and neck cancer have critical requirements on the delivery of radiation to the regions to be treated while maintaining a low level of radiation dose to the surrounding organs at risk (OAR). To fulfill the requirements, a precise delineation of both treated regions and OARs must be performed in the planning phase of the therapy, which is an extremely time-consuming task performed conventionally by the physicians on CT images manually.

In recent years, automatic approaches have been proposed to delineate various structures to be treated and/or spared in CT images, which include the parotid gland [1-7], brainstem [6-8], thyroid gland [9], mandible [7], optical track [8, 10], and neck lymph node regions [11]. A plethora of techniques and quantitative assessments can be found in [12]. Nevertheless, most of the automatic methods focus on one or a few structures, and a comprehensive assessment on the various automatic segmentation techniques for multiple structures has not yet been reported, in part, due to a lack of publicly available datasets with associated manual delineations.

Based on such a dataset that is made available by the organizers of the 2015 MICCAI Head and Neck Auto Segmentation Challenge, we propose an automatic multi-atlas approach for the segmentation of 9 structures in head and neck CT images.

By comparing the automatic segmentation results with the provided manual delineations, the robustness and flexibility of the proposed approach are demonstrated.

2 Data Description

A total of 33 clinical head and neck CT images (transverse scan) were provided by the organizers of the challenge for training purpose. Among the 33, 25 were assigned as the primary training set, in which the slice thickness was 3 mm for 14, 2.5 mm for 10, and 1.25 mm for one. All of the primary training images were associated with manual delineations for 7 structures which were the brainstem, mandible, left and right optic nerves, optic chiasm, and left and right parotid glands. Manual delineations for the bilateral submandibular glands (two individual structures) were provided for 21 out of the 25 cases on the left side and 19 out of the 25 cases on the right side. Meanwhile, the remaining 8 training images were assigned as an optional training set, on which manual delineations were provided for all the aforementioned structures but not for the mandibles and the left and right submandibular glands.

A total of 10 CT images were provided for offline testing and 5 images were provided for online testing. Manual delineations were not provided at the testing stage but were made available after the submission of testing results. Among the 15 images, slice thickness was 3 mm for 8, 2.5 mm for 6, and 2 mm for the remaining one.

All of the training and testing images had in-plane dimension of 512×512 voxels. The in-plane voxel sizes were isotropic and vary from 0.76 mm to 1.27 mm, while for the majority of the cases the in-plane voxel sizes were around 1 mm.

3 Method

3.1 Overview

The framework introduced here is based on the multi-atlas approach proposed previously for the segmentation of thyroid glands [9]. The framework was built based on the primary training set, while tuning and adjustments were achieved with the help of the optional training set.

3.2 Global Registrations and Creation of Average Head and Neck Atlases

One image in the training set is selected as a template in the initial step, such that all the other training images can be aligned with it at the global level. Due to the flexible neck structure in head and neck CT images, a global-level affine registration with 12 degrees of freedom (rotation, translation, scaling, and sheering) is performed at the initial step to align images with the template. Following the affine registration, a non-rigid registration at a more detailed level is needed to further align the images. In order to reduce the bias that could be introduced by the anatomy of any individual template image, our solution was to build an average atlas of the population. Using the 25 images in the primary training set, we were able to build average atlases fol-

lowing the method proposed in [13]. Since nearly equal portions of the training images are of either 3 mm or 2.5 mm slice thicknesses, we elected to build two atlases, one with 3 mm slice thickness and the other with 2.5 mm slice thickness. After a visual inspection, we identified two images (0522c0003 for the 3 mm atlas and 0522c0253 for the 2.5 mm atlas) covering sufficient region from the top of the skull to the upper chest region and showing mostly normal anatomy and posture as the initial templates.

The algorithm for nonrigid registration on the global level was Adaptive Bases Algorithm (ABA) [14] modeling the nonrigid deformation field as a linear combination of 3D radial basis functions (RBF). Normalized mutual information was used as the measure of similarity between the target and moving images in a gradient-descent optimization. Since the alignment was intended to achieve a normalization of global posture and an approximate spatial alignment of corresponding anatomical structures, 3D RBF's were placed at a density of one per 20 mm, which led to a smooth deformation. The registration was implemented in a multi-resolution manner, in which the first iteration started 2 levels lower than the full-resolution. At the end of each iteration, the forwardly registered training images were averaged and then deformed by the average of all the backward deformation fields from the current atlas. The deformed image was then used as the atlas in the next iteration. The procedure converged after 5 iterations.

3.3 Multi-atlas Segmentation of Structures on a Local Level

With the average atlas established, we performed global affine and nonrigid registrations between the atlas and the patient images to be analyzed using the same aforementioned registration parameters. In the following step, since corresponding anatomical structures were already aligned approximately, we were able to define local bounding boxes with the structures of interest at the center while maintaining margins comparable to the scale of the structure in all 6 directions in 3D, and then automatically place them back onto the globally aligned patient images for cropping.

Since the training images were also globally aligned with the average atlas, cropping local bounding boxes on the aligned training images allowed us to have a set of local-level atlases. Local nonrigid registrations were then performed between the cropped patient image and all the cropped atlases, with parameters favoring detailed deformations, e.g. the RBF density was increased to one per 8 mm.

Due to the existence of structures having similar density as adjacent tissues, e.g. brainstem against the surrounding soft tissues and mandible against the teeth, which results in less crisp boundaries in CT scans, we believed it was preferable to have more flexibility in the way local deformation fields were modeled. Therefore in addition to ABA we also performed local registrations using the Symmetric Normalization (SyN) algorithm from the Advance Normalization Tools (ANTS) [15] for modeling highly regularized small-scale local displacements.

After all atlases were aligned with the new image on the local level, local nonrigid transformations were applied to the manual delineations of the structures of interest (deformed from the patient image space using global transformations and then

cropped locally) when they were available. This would give us a set of N segmentations for the structure. We fused them into segmentation M as a weighted sum:

$$M = \sum_{i=1}^{N} w_i M_i \text{ , where } w_i = \frac{cc_i}{\sum_{i=1}^{N} cc_i} \quad (1)$$

The correlation coefficient cc_i was calculated between the local patient image and the ith locally deformed atlas image. As a result, segmentations from atlases that were more similarly aligned with the patient image would have higher weights in the combined segmentation. To focus on the regions of interest, the correlations were only calculated inside a mask defined as the union of all segmentations from the N local atlases. Furthermore, to eliminate the impact of dissimilar atlases, the weights were ranked and those in the lower half were set to zero in the combination.

Since it was convenient to define bounding boxes to cover multiple structures when they were close to each other, the left and right optic nerves and the chiasm were segmented within one box covering the region of the eyes, and the left and right submandibular glands were segmented in one box covering the lower chin region. All the other 4 structures were segmented with individual boxes.

3.4 Post-processing for Automatic Segmentation Results

After the locally combined segmentations were generated, they were reconstructed in the global space and transformed back onto each patient image using the inverse transformations obtained from global registrations, and finally thresholded into binary masks. For most masks, since the original binary manual delineations were normalized to [0, 255], and the default threshold was set to 127 (half of the maximum). The threshold was set to 105 in the masks for the optical nerves and chiasm due to a lower rate of overlapping in the combination (combined masks having maximum value around 210-220 in general) as a result of fine-tuning (thresholds tried at 115, 105, 95, and 85) on the optional training set. For the brainstem, following the guideline for manual delineations, landmarks like the posterior clinoid process and C1 vertebrae were localized on the average atlases and transformed back into the patient images with global transformations to determine the levels for trimming.

3.5 Final Selection of Global Atlases and Local Registration Algorithms

The two global atlases and the two local registration algorithms led to a total of 4 combinations which were denote atlas_2.5mm-ABA, atlas_2.5mm-SyN, atlas_3mm-ABA, and atlas_3mm-SyN. When applied on patient images, the final structure-dependent selections of global atlases and local registration algorithms were made based on experiments carried out on the optional training set. Measured by volume Dice similarity coefficient (DSC) compared to manual delineations, atlas_2.5mm-ABA performed the best on parotid (L and R), optic nerve (L and R), and chiasm,

while atlas_2.5mm-SyN yielded the best performance on brainstem. Due to the lack of manual segmentations associated with the optional training set, atlas_2.5mm-ABA was selected for submandibular gland (L and R) for showing the least amount of under-segmentations and atlas_3.0mm-SyN was selected for the mandible for showing the least amount of over-segmentations into the teeth based upon visual inspection.

Table 1. Volume DSC between manual and automatic segmentations for all structures on both testing datasets.

Struct / Case	Brain Stem	Chi-asm	Man-dible	Nerve _L	Nerve _R	Parot-id_L	Parot-id_R	Sub_ L	Sub_ R
0555	0.729	0.341	0.883	0.596	0.533	0.791	0.830	0.571	0.574
0576	0.889	0.505	0.857	0.711	0.702	0.814	0.870	0.594	0.501
0598	0.814	0.550	0.919	0.675	0.669	0.826	0.868	0.813	0.756
0659	0.808	0.539	0.943	0.531	0.530	0.781	0.760	0.745	0.671
0661	0.714	0.482	0.927	0.747	0.649	0.793	0.823	0.707	0.684
0667	0.870	0.353	0.911	0.699	0.638	0.850	0.880	0.776	0.757
0669	0.893	0.552	0.932	0.734	0.550	0.811	0.831	0.565	0.725
0708	0.717	0.121	0.915	0.563	0.679	0.851	0.865	0.724	0.782
0727	0.821	0.399	0.940	0.650	0.703	0.860	0.800	0.795	0.802
0746	0.756	0.077	0.888	0.516	0.498	0.770	0.649	0.597	0.740
0788	0.780	0.432	0.935	0.624	0.726	0.843	0.843	0.746	0.562
0806	0.884	0.391	0.941	0.609	0.540	0.721	0.744	0.790	0.767
0845	0.846	0.386	0.920	0.649	0.564	0.862	0.827	0.560	0.568
0857	0.802	0.451	0.929	0.709	0.578	0.777	0.764	0.791	0.714
0878	0.717	0.125	0.915	0.647	0.588	0.819	0.861	0.757	0.543
Mean	0.803	0.380	0.917	0.644	0.610	0.811	0.814	0.702	0.676

4 Results

For the 15 images in both testing datasets, as shown in Table 1, 3D volume DSC was calculated between the automatic and manual segmentations. It can be seen that the automatic segmentation reached 0.917 mean DSC for mandible, and the mean DSC's for brainstem and parotid (L and R) were all above 0.8. The mean DSC's for optic nerve (L and R) were both above 0.61 while the lowest mean DSC was observed on the Chiasm (0.380). Measured by the 95th percentile of Hausdorff distance (HD), as shown in Table 2, mandible and optic nerve (L and R) showed the best results (mean HD<3.2 mm), while parotid (L and R) showed the worst (mean HD>6.4 mm). Qualitative segmentation results are shown in Fig. 1 for one case (0522c0788) from the onsite testing set as an example, where the manual contours are in green color and the contours of the corresponding automatic segmentations are in red color.

Table 2. The 95th percentile Hausdorff distance (in mm) between manual and automatic segmentations for all structures on both testing datasets.

Struct / Case	Brain Stem	Chi-asm	Man-dible	Nerve _L	Nerve _R	Parot-id_L	Parot-id_R	Sub_ L	Sub_ R
0555	5.65	4.08	4.48	2.75	4.26	4.96	5.13	6.03	8.23
0576	2.90	4.11	3.52	1.79	2.50	5.96	3.76	6.62	8.27
0598	4.54	4.48	1.41	2.67	2.33	4.99	3.99	4.10	3.60
0659	5.05	2.57	1.42	3.63	3.15	5.52	6.01	6.90	6.47
0661	8.03	3.08	1.96	1.99	2.48	6.29	6.27	3.56	4.41
0667	3.17	4.40	2.50	2.37	3.14	8.98	3.57	7.44	4.39
0669	2.92	2.92	2.50	2.03	3.61	7.52	7.13	5.53	3.76
0708	8.86	3.98	2.08	3.22	2.50	3.92	3.56	4.62	3.76
0727	3.96	4.02	2.33	3.40	3.00	4.21	8.86	2.96	2.93
0746	6.00	4.47	2.89	4.12	7.28	6.87	10.18	5.93	4.65
0788	7.35	5.17	2.81	2.67	2.33	8.25	4.49	3.96	6.23
0806	3.43	4.29	1.79	3.36	2.27	15.62	10.18	3.41	3.92
0845	4.82	4.08	2.50	2.15	3.20	4.66	6.28	9.07	8.02
0857	4.10	4.39	2.81	2.09	2.92	9.24	11.17	3.81	5.05
0878	6.44	5.86	2.33	3.11	2.33	7.54	5.89	6.31	12.45
Mean	5.15	4.13	2.49	2.76	3.15	6.97	6.43	5.35	5.74

5 Conclusion and Discussions

In this work we proposed a multi-atlas approach for the segmentation of multiple structures in the head and neck CT images for radiotherapies. Based on global affine and nonrigid registrations, a patient image was first aligned with an average head and neck CT atlas on the global level. The aligned image was further processed by performing local nonrigid registrations with multiple atlases. The subsequent labels deformed from the atlases were then combined with weights determined by the local correlation coefficients between the patient image and the registered atlas images.

Compared with previous results in [8], where mean DSC was around 0.52, the segmentation of optic nerves showed improvements. Since delineation of thin structures can be affected drastically by the image's slice thickness, in addition to the contribution of multiple atlases, the improvement could also be attributed to atlases with 2.5 mm slice thickness (3 mm in [8]). Moreover, the method in [11] which incorporated 30 atlases in which 27 had slice thickness of 2.5 mm or less yielded better results on optic nerves when evaluated by DSC (mean DSC=0.74), although the results were not shown to be better when assessed by HD (mean HD=3.75 mm).

Tested on a workstation with two quad-core 2.50 GHz Intel Xeon CPU and 16 GB of memory, the total running time for segmenting a structure consists of around 5 minutes for global affine registration, 25 minutes for global nonrigid registration, and around 60 minutes for local nonrigid registration in the largest bounding box. Other

processing time (image cropping, mask padding, thresholding, and deformation) can take around another 10 minutes. The processing time could potentially be reduced by further tuning the local registration parameters without significantly sacrificing the quality of the registration result, and the impact could be minimized after atlas fusion.

The proposed framework has the flexibility to introduce more atlas images into the local multi-atlas step. Moreover, with more atlases available, more sophisticated selection and fusion techniques could be applied to improve the segmentation of structures of more inter-subject variability, e.g. the submandibular gland and the top side of the parotid gland.

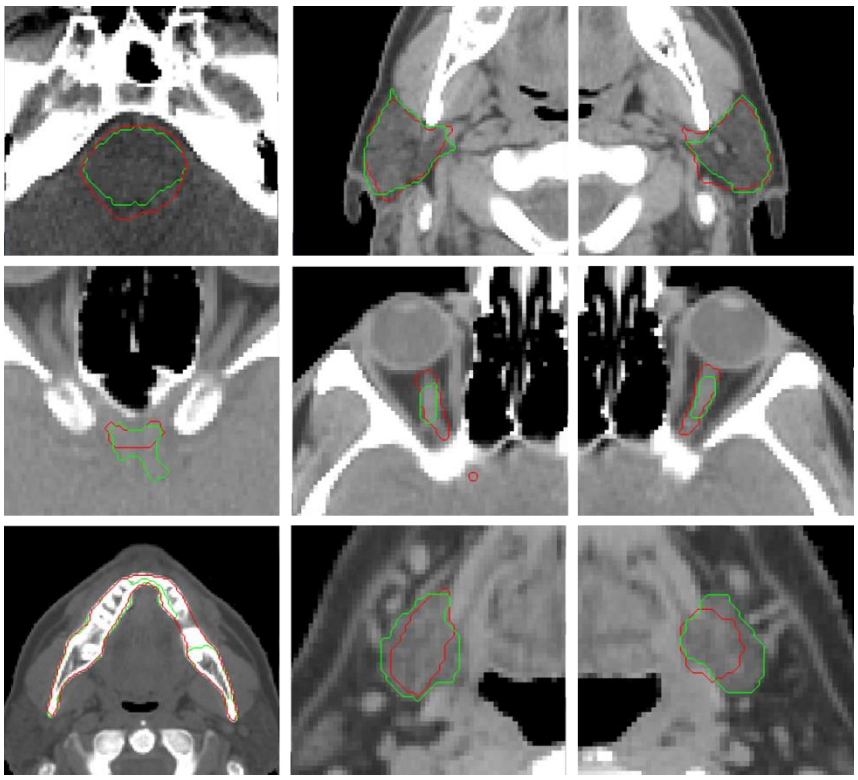

Fig. 1. Segmentation contours overlapped with CT image for 0522c0788. Top row left to right: BrainStem, Parotid_L, and Parotid_R. Middle row left to right: Chiasm, OpticNerve_L, and OpticNerve_R. Bottom row left to right: Mandible, Submandibular_L, and Submandibular_R. Auto: red. Manual: green.

References

1. Ramus, L. and Malandain, G.: Multi-atlas based segmentation: application to the head and neck region for radiotherapy planning. In proc. MICCAI 2010 Workshop Head & Neck Autosegmentation Challenge, 281–288. (2010)

2. Han, X., Hibbard, S., O'Connell, P., and Willcut, V.: Automatic segmentation of parotids in head and neck CT images using multi-atlas fusion. In proc. MICCAI 2010 Workshop Head & Neck Autosegmentation Challenge, 297-304. (2010)
3. Yang, J., Zhang, Y., Zhang, L., and Dong, L.: Automatic segmentation of parotids from CT scans using multiple atlases. In proc. MICCAI 2010 Workshop Head & Neck Autosegmentation Challenge, 323–330. (2010)
4. Chen, A., Noble, J.H., Niermann, K.J., Deeley, M.A., & Dawant, B.M.: Segmentation of parotid glands in head and neck CT images using a constrained active shape model with landmark uncertainty. In SPIE Medical Imaging (pp. 83140P-83140P). (2012)
5. Wachinger, C., Sharp, G.C., & Golland, P.: Contour-driven regression for label inference in atlas-based segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013 (pp. 211-218). (2013)
6. Fritscher, K.D., Peroni, M., Zaffino, P., Spadea, M.F., Schubert, R., & Sharp, G.: Automatic segmentation of head and neck CT images for radiotherapy treatment planning using multiple atlases, statistical appearance models, and geodesic active contours. *Med. Phys*, 41(5), 051910. (2014)
7. Qazi, A.A., Pekar, V., Kim, J., Xie, J., Breen, S.L., Jaffray, D.A.: Auto-segmentation of normal and target structures in head and neck CT images: a feature-driven model-based approach. *Med. Phys*, *38*(11), 6160-6170. (2011)
8. Deeley, M.A., Chen, A., Datteri, R., Noble, J.H., Cmelak, A.J., Donnelly, E.F., ... & Dawant, B.M.: Comparison of manual and automatic segmentation methods for brain structures in the presence of space-occupying lesions: a multi-expert study. *Physics in Medicine and Biology*, 56(14), 4557. (2011)
9. Chen, A., Niermann, K.J., Deeley, M.A., & Dawant, B.M.: Evaluation of multiple-atlas-based strategies for segmentation of the thyroid gland in head and neck CT images for IMRT. *Physics in medicine and biology*, *57*(1), 93. (2012)
10. Chen, A., Deeley, M.A., Niermann, K.J., Dawant, B.M.: Combining registration and active shape models for the automatic segmentation of the lymph node regions in head and neck CT images. *Med. Phys*, 37(12), 6338-6346. (2010)
11. Harrigan, R.L., Panda, S., Asman, ... & Landman, B.A.: Robust optic nerve segmentation on clinically acquired computed tomography. *Journal of Medical Imaging*,*1*(3), 034006-034006. (2014)
12. Sharp, G., Fritscher, K. D., Pekar, V., Peroni, M., Shusharina, N., Veeraraghavan, H., & Yang, J.: Vision 20/20: perspectives on automated image segmentation for radiotherapy. *Med. Phys*, 41(5), 050902. (2014)
13. Guimond, A., Meunier, J., & Thirion, J. P.: Average brain models: A convergence study. Computer vision and image understanding, 77(2), 192-210. (2000)
14. Rohde, G. K., Aldroubi, A., & Dawant, B. M.: The adaptive bases algorithm for intensity-based nonrigid image registration. *Medical Imaging, IEEE Transactions on*, *22*(11), 1470-1479. (2003)
15. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C:. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1), 26-41. (2008)