

Multi-Atlas based Segmentation of Head and Neck CT Images using Active Contour Framework

Subrahmanyam Gorthi¹, Meritxell Bach Cuadra¹, Ulrike Schick², Pierre-Alain Tercier³, Abdelkarim S. Allal³, and Jean-Philippe Thiran¹

¹ Signal Processing Laboratory (LTS5), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, subrahmanyam.gorthi@epfl.ch,

² Department of Radiation Oncology, University Hospital of Geneva, Switzerland,

³ Service de Radio-oncologie, Hôpital Fribourgeois, Fribourg, Switzerland. *

Abstract. This paper presents the segmentation of bilateral parotid glands in the Head and Neck (H&N) CT images using an active contour-based atlas registration. We compare segmentation results from three atlas selection strategies: (i) selection of “single-most-similar” atlas for each image to be segmented, (ii) fusion of segmentation results from multiple atlases using STAPLE, and (iii) fusion of segmentation results using majority voting. Among these three approaches, fusion using majority voting provided the best results. Finally, we present a detailed evaluation on a dataset of eight images (provided as a part of H&N auto segmentation challenge conducted in conjunction with MICCAI-2010 conference) using majority voting strategy.

1 Introduction

Automated segmentation of structures in the Head and Neck (H&N) CT images is a challenging as well as important task for radiation treatment of H&N cancer [1]. Among various structures in the H&N region, parotid glands are one of the important organs at risk that need to be accurately segmented in treatment planning. Automated segmentation of parotid glands is challenging because of their low contrast and lack of distinctly visible boundaries with the surrounding structures. Thus, for an accurate segmentation of these structures, inclusion of prior knowledge is essential. Atlas-based segmentation methods are widely used for exploiting prior anatomical knowledge. There are two factors that mainly affect the accuracy of atlas-based segmentation methods: the type of registration algorithm used for mapping the atlas to the image to be segmented, and the closeness/similarity of the atlas to the image to be segmented.

In this paper, we perform segmentation of bilateral parotid glands in the H&N CT images, using an Active Contour-Based Atlas Registration (ACBAR)

* This work is supported in part by the Swiss National Science Foundation under Grant 205321-124797 and by CIBM of the Geneva–Lausanne Universities and EPFL, as well as the foundations Leenaards and Louis-Jeantet.

framework. This framework has been already proven to be successful in the segmentation of other structures in the H&N CT images, like lymph nodes [1], mandible and brainstem [2]. As we mentioned, atlas selection is another important factor that affect the accuracy of segmentation. In this paper, we mainly focus on the atlas selection strategies.

2 ACBAR

In the Section, we present a brief description of Active Contour-Based Atlas Registration (ACBAR) framework. Please refer to [1,2] for more details. ACBAR framework combines the forces coming from both optical flow framework (like pixel-based forces), and active contour framework (like region-based and boundary-based forces). The formulation of ACBAR has been intuitively derived from level set equation [3]. The generalized evolution equation of the registration model is given by:

$$\frac{\partial u(x, t)}{\partial t} = -S(x) \nu(\phi_L(x + u(x, t), 0)) \frac{\nabla \phi_L}{|\nabla \phi_L|} \quad (1)$$

where $u(x, t)$ is the deformation field vector at time t and spatial coordinates x , ν is a speed function that contains local segmentation and contour regularization constraints, $S(x)$ is a sign function that provides polarity information, and ϕ_L is a labeled level set representation of the contour proposed in [1].

In the current application, we perform a two-level hierarchical registration. In the first level, we perform a region-based registration driven by two selected structures of interest: external contour of the image and the mandible. The reason for using mandible is because it is adjacent to the actual structure to be segmented (parotid glands), and thus can influence in accurately segmenting the parotid glands on the target image. The region-based energy term used is inspired by region-based segmentation model proposed by Chan and Vese [4], and it is same the one described in [2]. In the second level, the registration is driven by pixel-based forces, and the corresponding speed term is same as the one used by [5]. The details of the registration parameters, preprocessing and postprocessing used here for the segmentation of parotid glands are presented in Section 4.

3 Atlas Selection Strategies

Atlas selection strategy is a key issue for achieving accurate results in atlas-based segmentation [6,7]. In majority of works, the much deserved attention has not been given for atlas selection, and a single segmented image is almost randomly selected as an atlas for all the images to be segmented. Such random selection can lead to significant undesirable bias. One of the simple and effective approaches to overcome this problem is, for each image to be segmented, to adaptively select the most similar atlas from a given database of atlases [6]. Alternatively, instead

of using a single atlas, multiple atlases can be used. For instance, in [6], it is shown for the segmentation of brain in the confocal microscopy images, that, using multiple atlases can give more accurate results than with a single-best-atlas.

Segmentation results from multiple atlases can be fused in different ways [6,7,8]. In this paper we consider two popular fusion approaches: “Majority voting” [6] and “STAPLE” [8]. In majority voting, each voxel is assigned with the label that most segmentations agree [7]. Another popular algorithm is “Simultaneous Truth And Performance Level Estimation” (STAPLE) [8]. Intuitively, majority voting gives equal weight to segmentations from all atlases, while in STAPLE, the probabilistic estimate of the true segmentation is formed by estimating an optimal combination of the segmentations, weighting each segmentation depending upon the estimated performance level. In this paper, we assess three atlas selection strategies: (i) adaptively selecting the “most similar atlas” for each image to be segmented, (ii) fusion of segmentation results from multiple atlases using STAPLE, and (iii) fusion of segmentation results from multiple atlases using majority voting.

4 Results

The dataset is provided by Princess Margaret Hospital, Toronto, as a part of H&N auto-segmentation challenge [9]. It currently consists of 18 CT images. At the time of evaluation, expert segmentation for only 10 images (training data) are available to the participants, whereas for the remaining 8 images (testing data), expert segmentations are hidden from the participants and are available only with the organizers of the challenge; automated segmentations for the testing data are submitted by the participants to one of the organizers and then, the organizer has generated the evaluation results for the testing data. Out of 10 images in the training data, for 3 images (numbered: 01, 04 and 10), there are considerable artifacts in the mandible region; so we have constructed an atlas database with only the remaining 7 images. We note that we could actually remove those 3 images from the training data as well. However, we deliberately did not do that because, in practice, we can only select the images to be used in the atlas database, but not the images to be segmented. On the other hand, we could have introduced a preprocessing step before segmenting them, in order to remove the artifacts and thereby obtain more accurate evaluation.

In order to speed up the registration, all images are cropped in the Z-direction; the images are cropped such a way that they include all the structures of interest (parotid glands) as well as 3 additional axial slices both at the top and bottom of the images.

First of all, an affine registration is performed between the images to be segmented versus all the images in the atlas database. We use the AffineTransform available in ITK ⁴ along with Mean Square Error (MSE) as similarity metric.

⁴ <http://www.itk.org/>

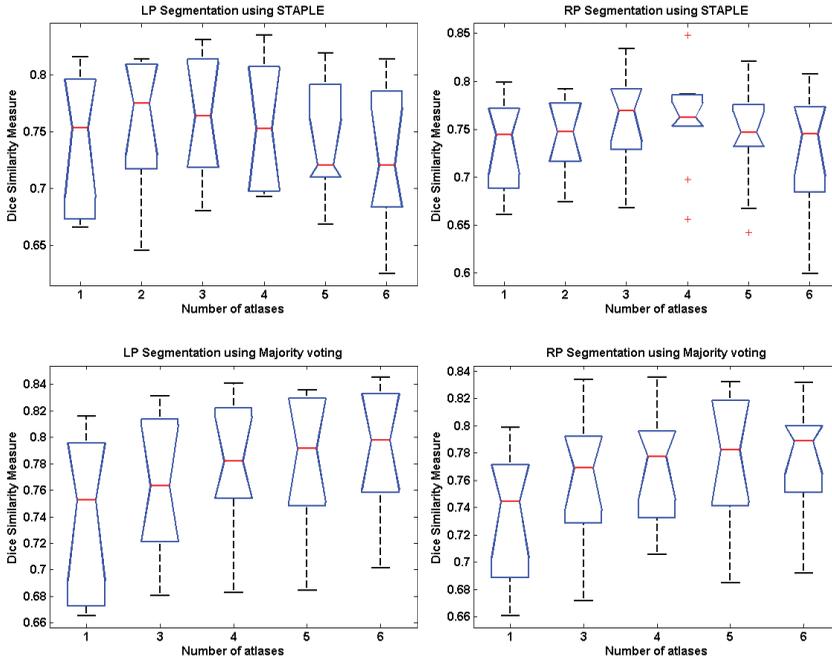


Fig. 1: Box plots of Dice Similarity Measure (DSM) for the segmentation of Left Parotid (LP) and Right Parotid (RP) glands, for varying number of atlases. Box plot corresponding to a single atlas represents the results for single-best-atlas selection.

Then, for each image to be segmented, atlases are ordered based on the resulting MSE at the end of affine registration. This is based on our hypothesis that smaller MSE indicates better similarity of the atlas to the image, and thus, is ordered first. This ordering is used later while merging segmentations from multiple atlases. Then, as described in Section 2, a 2-level active contour-based registration is performed; in the first level, region-based forces and curvature forces are used, and in the second level, pixel-based forces are used. The parameters for ACBAR are same as those used in [2]. For the segmentation of parotid glands on testing data, merged segmentation results from multiple atlases are post-processed by connected-component-thresholding to contain only the principal component, followed by Gaussian smoothing.

4.1 Comparison of Atlas Selection Strategies

Three types of atlas-selection strategies are compared on the training data of 10 images. While selecting atlases for each image, obviously, the same image in the atlas database is left out and is not considered as an atlas-candidate. Figure 1 shows box plots of Dice Similarity Measure (DSM) for the segmentation parotid glands, with varying number of atlases. Top row and bottom row respectively

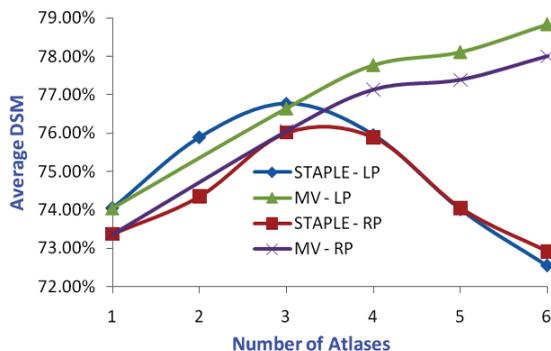


Fig. 2: Graph showing the average values of DSM for the segmentation of Left Parotid (LP) and Right Parotid (RP) glands, with varying number of atlases. Note that since atlases for each image are ordered in the decreasing order of similarity, the values shown for single atlas (i.e., values corresponding to x-label: “1”) represent the average DSM values for “single-best-atlas” selection strategy.

Atlas Selection Strategy	Optimal no. of Atlases	Corresponding DSM (mean \pm SD)	
		Left Parotid	Right Parotid
Adaptive Single Best Atlas	-	74.05% \pm 6.02%	73.37% \pm 5.06%
STAPLE	3	76.77% \pm 5.13%	76.02% \pm 5.05%
Majority Voting	6	78.85% \pm 5.01%	78.01% \pm 4.18%

Table 1: Comparison of the three approaches for parotid glands segmentation.

show the statistics for atlas fusion using STAPLE and Majority Voting (MV) algorithms. First and second columns respectively show the statistics for left and right parotid glands. Please note that for each image to be segmented, atlases are ordered in the decreasing order of similarity; in other words, the first atlas for each image represents the “most similar atlas” for that image. Thus, in the box plots of both STAPLE and majority voting approaches, the boxes corresponding to single atlas (i.e., the first box in each figure with x-label: “1”) actually represent the statistics for “single-most-similar” atlas selection. Because of this reason, we have not separately shown the box plots for single best atlas selection strategy. Figure 2 shows average values of DSM for parotid glands, over the complete training data, with varying number of atlases. Table 1 summarizes the comparison of the three atlas selection approaches. Note that number of atlases for which the average value of DSM is maximum, is considered as the “optimal number of atlases” for that strategy.

Multi-atlas-based segmentation results clearly outperformed the single-best-atlas selection, with an optimal selection of “number of atlases”. Between the multi-atlas selection strategies, STAPLE gave its best results with 3 atlases whereas results from MV kept improving till 6 atlases. There is approximately

Dataset No.	Mean HD	Median HD	No. of slices (HD > 3 mm)
11	9.65	9.61	34 (34)
12	7.40	7.44	28 (28)
13	10.34	9.79	26 (26)
14	11.65	9.69	24 (24)
15	6.42	5.12	23 (23)
16	6.69	4.98	29 (28)
17	14.26	14.71	33 (33)
18	8.76	8.22	24 (24)
Mean±SD	(9.40±2.68)	(8.70±3.11)	-

Table 2: Hausdorff distance(HD) statistics for left parotid segmentation.

2.7% improvement in DSM from single-best-atlas to the best results of STAPLE, and there is further improvement of 2.1% from the best results of STAPLE to that of MV. Based on these results, for the final evaluation on the testing data, we have chosen majority voting strategy with 6 atlases.

4.2 Evaluation on Testing Data

As mentioned in the preceding Subsection, the final evaluation on the testing data of 8 images is performed using majority voting with 6 atlases. The quantitative evaluation is performed using various Hausdorff Distance (HD)-based and overlap (OV)-based metrics. Some of these metrics are evaluated both slice-wise and volumetric-wise. Please refer to [9,10] for a detailed description of the evaluation metrics.

Figure 3 shows ground truth and automated segmentation of parotid glands for one of the images. Table 2 and Table 3 respectively present the Hausdorff distance and overlap statistics for left parotid gland segmentation. Similarly, Table 4 and Table 5 respectively present the Hausdorff distance and overlap statistics for right parotid gland segmentation. As expected, the quantitative evaluation results for left and right parotid glands have similar behavior, because of the similarities of these two structures. Average total volume overlap is around 75.5% whereas the average mean-HD is around 9 *mm*.

5 Conclusions

In this paper, we have presented the segmentation of parotid glands in the H&N CT images using active contour-based atlas registration framework. We have evaluated three atlas selection strategies: single-best-atlas, STAPLE and majority voting; among the three approaches, majority voting gave the best results. We then performed the final evaluation on the testing data using majority voting.

Although the current automated segmentations still need manual corrections before using them in treatment planning, these results are quite promising con-

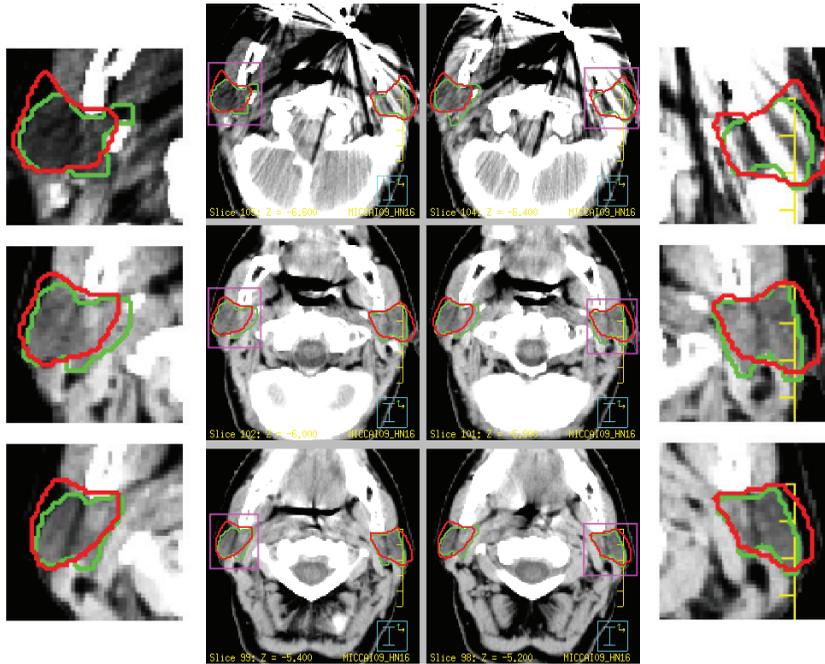


Fig. 3: Middle two columns show the segmentation of parotid glands for one of the images in the testing data. First and last columns show a zoom of the selected regions in the middle image. Ground truth and automated segmentations are respectively shown in green and red colors.

Dataset No.	Average slice OV	Median slice OV	Total volume OV
11	69.4 %	72.3 %	77.3 %
12	72.1 %	78.8 %	77.0 %
13	69.0 %	75.7 %	74.2 %
14	68.6 %	78.5 %	74.6 %
15	81.9 %	87.6 %	85.9 %
16	73.0 %	79.0 %	75.9 %
17	60.7 %	63.0 %	63.5 %
18	69.4 %	73.8 %	75.4 %
Mean±SD	(70.51±5.89)%	(76.09±7.02)%	(75.48±6.10)%

Table 3: Overlap(OV) statistics for left parotid segmentation on testing data.

sidering the low contrast of the parotid glands and the presence of artifacts. These two issues can be clearly noticed from Figure 3.

The main contribution of this paper is the evaluation of three atlas selection strategies. We note that while ordering the atlases for fusion, we have arbitrarily selected MSE as a measure of similarity. In future work, we would like to eval-

Dataset No.	Mean HD	Median HD	No. of slices (HD > 3 mm)
11	6.81	5.69	32 (32)
12	12.30	9.62	30 (30)
13	9.69	7.01	26 (26)
14	12.23	9.81	24 (24)
15	8.13	5.27	26 (24)
16	7.03	6.02	31 (30)
17	10.17	10.28	27 (27)
18	6.62	5.90	25 (25)
Mean±SD	(9.12±2.33)	(7.45±2.10)	-

Table 4: Hausdorff distance(HD) statistics for right parotid segmentation.

Dataset No.	Average slice OV	Median slice OV	Total volume OV
11	77.8 %	83.1 %	81.7 %
12	66.6 %	76.2 %	71.7 %
13	72.4 %	75.8 %	77.9 %
14	62.1 %	70.6 %	68.1 %
15	77.7 %	85.1 %	84.4 %
16	69.5 %	75.4 %	73.8 %
17	64.7 %	72.2 %	67.5 %
18	73.8 %	82.3 %	80.2 %
Mean±SD	(70.58±5.86)%	(77.59±5.30)%	(75.66±6.34)%

Table 5: Overlap(OV) statistics for right parotid segmentation on testing data.

uate and compare the effect of using other similarity measures, like normalized correlation coefficient and mutual information.

References

1. Gorthi, S., Duay, V., Houhou, N., Bach Cuadra, M., Schick, U., Becker, M., Allal, A.S., Thiran, J.P.: Segmentation of head and neck lymph node regions for radiotherapy planning using active contour-based atlas registration. *IEEE Journal on Selected Topics in Signal Processing* **3**(1) (2009) 135–147
2. Gorthi, S., Duay, V., Bach Cuadra, M., Tercier, P.A., Allal, A.S., Thiran, J.P.: Active contour-based segmentation of Head and Neck with adaptive atlas selection. In: *MICCAI workshop on 3D segmentation challenge for clinical applications*. (2009)
3. Osher, S., Sethian, J.: Fronts propagating with curvature-dependent speed - algorithms based on hamilton-jacobi formulations. *Journal of Computational Physics* **79**(1) (1988) 12–49
4. Chan, T.F., Vese, L.A.: Active contours without edges. *IEEE Transactions on Image Processing* **10**(2) (2001) 266–277
5. Vemuri, B.C., Ye, J., Chen, Y., Leonard, C.M.: Image registration via level-set motion: Applications to atlas-based segmentation. *Medical Image Analysis* **7**(1) (2003) 1–20

6. Rohlfing, T., Brandt, R., Menzel, R., Maurer Jr., C.R.: Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *NeuroImage* **21**(4) (2004) 1428–1442
7. Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J., Rueckert, D.: Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *NeuroImage* **46**(3) (2009) 726 – 738
8. Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging* **23**(7) (2004) 903–921
9. Head and Neck auto-segmentation challenge 2010: <http://www.grand-challenge2010.ca/>
10. Pekar, V.: <http://www.grand-challenge2010.ca/evaluation.pdf>