Automatic Segmentation of Parotids from CT Scans
Using Multiple Atlases

Jinzhong Yang, Yongbin Zhang, Lifei Zhang, and Lei Dong

Department of Radiation Physics, University of Texas MD Anderson Cancer Center
1515 Holcombe Blvd., Houston, TX 77030, USA
{jyang4, yonzhang, lifzhang, ldong}@mdanderson.org

Abstract. Accurate delineation of normal structures is a crucial step in designing highly conformal treatment plan for radiotherapy. Manual contouring of these structures is generally tedious and time-consuming, and suffers inter-observer variations. Thus it is urgent to automate this process in the clinical practice. In this paper, we design an automatic atlas-based segmentation method for delineating parotid glands, which is an important organ to spare in head-and-neck cancer radiotherapy in order to reduce radiation-induced toxicity. The proposed approach uses deformable image registration to map the pre-defined atlas structures to a new patient image. A local image matching is implemented to find a set of optimal atlases. Statistical fusion of multiple atlas-generated structures is applied to reduce the inter-subject variability. The entire segmentation scheme is validated using eight clinical datasets. Volume overlap and Hausdorff distance (HD) is computed against the physician’s manual delineation. Our approach shows an averaged volume overlap of 85% and an averaged median HD of less than 7mm.

Keywords: Atlas-based segmentation, Deformable registration, STAPLE.

1 Introduction

The development of intensity-modulated radiation therapy (IMRT) treatment technique allows the radiation dose to be delivered to the target with higher precision, while minimizing radiation toxicity to the adjacent normal tissue cells. In order to achieve a successful IMRT treatment, it is a crucial process for the clinical specialists to accurately define targets and all the organs-at-risks (OAR). However, the above process has proven to be tedious and time-consuming. Particularly for HN patients, it has been shown in some studies that specialists may spend several hours on average to fully define the desired target for a single patient. In addition, manual contouring process introduces intra/inter-observer variability, which are caused by variable clinical experience and training of the specialists, as well as the limitation of medical image quality. As a result, there is an urgent need to define automatic segmentation systems to reduce contouring time drastically, and create more objective and standardized contours.
Recently, automatic atlas-based segmentation has drawn enormous attention in radiation therapy [1-3], the key idea of which is to exploit the prior knowledge encoded in those patients that have already been manually contoured in the database to segment the new patient images. In this project, the target structures that we are interested in are bilateral parotid glands, which are critical structures to spare in HN radiotherapy. First, we use deformable image registration approach to calculate dense correspondences between atlas image and new patient image; then this correspondence information in the form of displacement vector field will be used to warp the atlas structures onto the new image. Nonetheless, the high variations in intensity, contrast and geometry across different patients pose a significant challenge for intensity-based image registration. A bad atlas image usually introduces large registration discrepancies with the testing image, leading to inaccurate target segmentation. Therefore, a good atlas-based segmentation system requires not only a robust registration algorithm, but also an effective scheme to select optimal atlas templates that are close to the testing image from the patient database. In this paper, we have proposed an efficient atlas-based segmentation system that consists of three key components: atlas selection, deformable image registration and fusion of multiple atlases. First, a local image matching approach was used to select a set of optimal atlas candidates. Then, an improved “Demons” algorithm was used to perform registration between selected atlas images and patient image. The resultant displacement vector field mapped the atlas structures to the test patient image. Finally, STAPLE algorithm was employed to achieve a statistical estimation of true segmentation from multiple atlases.

2 Method and Materials

Eighteen clinical CT datasets were provided for the MICCAI 2010 Head and Neck Auto-Segmentation Challenge workshop by the Princess Margaret Hospital in Toronto, Canada. All images have the in-plane dimensions of 512×512 pixels with pixel size of approximate 0.98×0.98mm². The number of slices ranges from 108 to 191 with the slice thickness of 2mm. In all the 18 datasets, 10 of them are identified as the training datasets, for which the manual segmentations of parotids were provided. The manual delineations of the parotids were generated by an expert radiation oncologist and stored as binary masks. The other 8 datasets are used for the evaluation purpose.

2.1 Overview

We proposed a multi-atlas based segmentation method to delineate parotids from CT images automatically. Our method can be summarized in the following steps:

**Preprocess the training datasets:** we manually reviewed the expert delineated parotid contours and modified contours accordingly to minimize the intra-observer variability and to reduce the systematic error propagation during segmentation.

**Select optimal atlas candidates:** we selected 6 out of 10 training datasets as our atlases which are most representative and less registration error prone.

**Perform multi-atlas based image segmentation:** we used the selected atlases to perform segmentation on a given testing dataset. The overall framework of our me-
method is illustrated in Fig. 1. For a given testing dataset, the atlases are applied separately to register the testing dataset. A set of vector fields characterizing the individual registrations are then used to deformed the manually delineated contours separately for individual segmentations, which are then combined in a fusion framework to obtain the final segmented object.

![Fig. 1. Overall framework of multi-atlas based image segmentation.](image)

### 2.2 Preprocess training data

All training datasets were imported into Pinnacle Treatment Planning System (Philips Medical Systems) for manual review. The binary images of segmented parotids were converted to 2D slice contours and then imported into the Pinnacle Treatment Planning System. We first examined slice by slice to manually correct the contours where there exist obvious anatomically inaccurate segmentations. In addition to checking the contours in the transverse plane, we also checked the sagittal view and the coronal view of the contours and modify them to ensure its consistency between different slices. The inconsistency normally happens where the anatomical boundaries of parotids are ambiguous.

In order to minimize the intra-observer variability, we performed a manual contour refinement on each of the training datasets. In this step, we registered other 9 training datasets to the one in consideration, and deformed the parotids contours from other 9 training datasets. The 9 sets of deformed contours, as well as the manual delineated contours, describe the same parotid in the space of the training dataset in consideration. These 10 sets of contours were fed into the STAPLE (Simultaneous Truth and Performance Level Estimation) algorithm [4] to estimate a fused set of contours, which is considered to show the inherent contouring practice of the physician by
removing the variation across different subjects. Based on the STAPLE generated contours, we further refined the contours manually at the location that is anatomically ambiguous to be consistent with the physician practice. With this preprocessing we expect to minimize the systematic error propagation during the automatic segmentation in later steps.

2.3 Selection of Optimal Atlas Candidates

In this paper, we have proposed an approach that uses multiple representative atlases for parotid gland segmentation. First, a local appearance-based image matching approach was defined to find a set of closest templates from the training set. Secondly, multiple segmentation results were then fused to achieve final segmentation using STAPLE algorithm. As the target to be segmented is parotid gland, it is a reasonable assumption that good atlas candidates should possess similar intensity distribution within the testing image around parotid region. Image similarity metric was only calculated over a small region that covers the parotid, rather than the entire image. To start, all the images in training set were first registered to a common template. Subsequently, a small region of interest (ROI), which was expanded by 1 cm from the target, was cropped from each image sets, Fig. 2. Instead of directly matching images as used in [2], our approach performs image matching within a transformed low-dimensional space, which is more robust to the noises introduced by dental artifacts and registration errors. Principal Component Analysis (PCA) is used to learn a subspace representation of these image regions, whose basis vectors correspond to major variations of image intensities. By discarding those less important bases, significant dimension reduction can be achieved, and at the same time noises are suppressed. In this paper, we used a 5-d subspace to represent the training image regions, since we have found that the first five bases could capture over 80% the total variations, Fig. 2 illustrates our subspace learning procedure.

Once the subspace is given, we will perform image matching by comparing PCA coefficients using Normalized Correlation (NC) similarity metric. Leave-one-out strategy was used to find an optimal atlas set from the training set. In our experiment, a total of six images were selected as the atlas templates to perform atlas segmentation for the testing set. At the same time, we have also conducted a study to select an atlas set by visual assessment of images, where the image intensity, contrast and the head tilt, were considered as the major factors. The results were quite close to the automatic one. As a final note, we should em-
phasize that our atlas sets were selected off-line, and then used for all the testing images. Although it has achieved satisfactory results on this testing set, it is potentially a better solution to select the atlas online for any given patient’s image. This will require extra efforts to register the testing image to a common template.

2.4 Intensity-based Deformable Image Registration

Intensity-based deformable registration algorithms directly calculate correspondences from the intensity of pixels under the assumption that the intensity is consistent across images. Because Hounsfield units of the CT images are calibrated to the attenuation coefficient of water, intensity-based methods are thus preferred for CT-to-CT deformable registration. In this study, we chose to use an improved dual-force ‘Demons’ algorithm. The original ‘Demons’ algorithm [5] was further extended and validated by Wang et al. [6] with improved accuracy and speed. The diffusion process can be made more efficient within low-gradient regions, by introducing an additional force in the opposite direction, which can be characterized as:

\[
\vec{v} = \vec{v}_s + \vec{v}_m = (m - s) \times \frac{\vec{v}_s}{\sqrt{\vec{v}_s^2 + \alpha (m - s)^2}} + \frac{\vec{v}_m}{\sqrt{\vec{v}_m^2 + \alpha (s - m)^2}}
\]

where the parameter \(\alpha\) is introduced to control the weight between internal force and external force. Multi-resolution scheme was used to handle large deformation. Rigid alignment of image pair was first performed as a preliminary step to minimize the global positional difference. Instead of using the entire image for rigid alignment, we choose an object-specific alignment scheme, which is expected to be more flexible in achieving a good starting position for individual structures. In this paper, C2 was used as the alignment object due to its proximity to the parotid.

2.5 Automatic segmentation

The vector field we obtained from deformable registration indicates how each point of the atlas moves to match the testing image. It can be used to transform the manual segmented parotids on the atlas to match the parotids on the testing image. In order to achieve an accurate contour transformation, the binary image of the segmented parotids is first converted to closed triangular surface meshes. The vector field is applied to each mesh vertex to transform the parotid surface to match the testing image. Although the vertex positions have been changed, the topology of the surface is preserved. After the vertex mapping, the new surfaces represent the segmented parotids on the testing image. The surface meshes are then converted back to a binary image. In the clinical practice, the organ is contoured slicewise, normally in transversal planes. In order to imitate the clinical practice of contouring, we first cut through the surface meshes using an axial plane slice by slice, and then converted each slice to a 2D binary image. The final 3D binary segmentation is stacked from all those 2D images. This segmentation process was applied to each atlas to generate a set of individual segmentations for the same parotids on the testing image.
We observed that there exist big variations in the superior and inferior parts of parotids when we deformed all segmented parotids from multiple atlases. This observation is illustrated in Fig. 2, where we transformed the parotids on 9 training datasets to match the rest one. At the same time we also noticed that the group consensus of the individual segmentations very closelys to the manual segmentation (Fig. 3). This is basically because the inter-subject variations can be reduced when multiple segmentations of the same object are available. Therefore, we are motivated to take advantage of the STAPLE algorithm [4]. By applying the STAPLE algorithm to all the individual segmentations, the inter-subject variations are minimized and a fused binary image is produced as the final segmentation of parotids.

3 Results

We selected 6 datasets from the 10 training datasets (Set# 3, 5, 6, 8, 9, and 10) as our atlases based on our atlas selection rule. These 6 datasets are considered the most representative when they are used for the segmentation of the given 8 testing datasets. For each test dataset, 6 individual segmentation results were first computed using the 6 atlases, and then the STAPLE algorithm was applied to the 6 individual segmentations to obtain the final auto-segmentation result. The STAPLE fusion was performed separately for left and right parotids. The final segmentation results for all testing datasets were submitted to the organizers of the workshop for independent evaluation.

Fig. 4 illustrates the segmentation results for both the left and right parotids for one testing dataset #16. The figures were generated by the workshop organizer. As we can see from the figures, the auto-segmentation results match the manual segmentation well in most locations except for some ambiguous parts where it is not easy to determine the boundaries even for experts.

Quantitative evaluation was performed slice-by-slice for each dataset to compare the auto-segmentation with the manual segmentation results. The quantitative measures include the symmetric HD and the Dice similarity coefficient. A volumetric Dice similarity coefficient was also computed to evaluate the overall volume overlap for each dataset. The statistics of the quantitative evaluation are summarized in Tables 1-2. The tables show that the median slicewise HD is less than 7mm for most datasets,
and the median slicewise Dice coefficient is mostly above 0.85 for left parotids and 0.82 for right parotids. The total volume overlap is close to or above 0.85 for most datasets. We can observe that, except for dataset #17, our method provides accurate and robust segmentation of parotids for most of the given testing datasets.

Fig. 4: Snapshots of the parotids segmentation for one testing dataset. Red curves indicate auto-segmentation results and green curves are the corresponding manual segmentation.

All our experiments were carried out on a PC with an eight-core Intel Xeon CPU and 8GB memory. Multithread computing is enabled in our algorithm. For segmentation using a single atlas, the computation time is less than 10 seconds. When all 6 atlases are used, the overall segmentation time is around 1 minute.

4 Conclusion

We have developed an automatic segmentation method for head-and-neck CT images using multiple sets of atlas to increase the robustness and accuracy of parotids segmentation. The experiment results demonstrated that our method was able to produce accurate segmentation of parotids using clinical data. We also showed a practical atlas selection strategy and demonstrated its computational efficiency of our method when multiple atlases were used for segmentation.
Table 1. Hausdorff distance (HD) and overlap (OV) statistics for left parotid segmentation in the testing datasets.

<table>
<thead>
<tr>
<th>Dataset No.</th>
<th>Mean per slice</th>
<th>Median per slice</th>
<th>No. of slices (&gt; 3 mm)</th>
<th>Mean per slice</th>
<th>Median per slice</th>
<th>Total volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>6.60</td>
<td>5.69</td>
<td>34 (33)</td>
<td>81.5</td>
<td>85.1</td>
<td>85.0</td>
</tr>
<tr>
<td>12</td>
<td>6.82</td>
<td>6.55</td>
<td>28 (26)</td>
<td>84.1</td>
<td>86.2</td>
<td>85.1</td>
</tr>
<tr>
<td>13</td>
<td>6.57</td>
<td>6.21</td>
<td>26 (26)</td>
<td>86.1</td>
<td>85.8</td>
<td>84.9</td>
</tr>
<tr>
<td>14</td>
<td>8.95</td>
<td>10.16</td>
<td>24 (20)</td>
<td>81.4</td>
<td>87.2</td>
<td>85.6</td>
</tr>
<tr>
<td>15</td>
<td>7.65</td>
<td>6.15</td>
<td>23 (21)</td>
<td>83.4</td>
<td>87.1</td>
<td>86.6</td>
</tr>
<tr>
<td>16</td>
<td>5.23</td>
<td>4.88</td>
<td>29 (26)</td>
<td>82.2</td>
<td>84.8</td>
<td>84.1</td>
</tr>
<tr>
<td>17</td>
<td>12.90</td>
<td>9.77</td>
<td>33 (33)</td>
<td>73.6</td>
<td>77.6</td>
<td>76.3</td>
</tr>
<tr>
<td>18</td>
<td>6.26</td>
<td>5.28</td>
<td>24 (24)</td>
<td>81.0</td>
<td>85.5</td>
<td>83.6</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>7.62</strong></td>
<td><strong>6.84</strong></td>
<td><strong>94.3 % &gt;3 mm</strong></td>
<td><strong>81.7</strong></td>
<td><strong>84.9</strong></td>
<td><strong>83.9</strong></td>
</tr>
</tbody>
</table>

Table 2. Hausdorff distance (HD) and overlap (OV) statistics for right parotid segmentation in the testing datasets.

<table>
<thead>
<tr>
<th>Dataset No.</th>
<th>Mean per slice</th>
<th>Median per slice</th>
<th>No. of slices (&gt; 3 mm)</th>
<th>Mean per slice</th>
<th>Median per slice</th>
<th>Total volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>6.96</td>
<td>6.69</td>
<td>32 (31)</td>
<td>82.8</td>
<td>86.4</td>
<td>85.2</td>
</tr>
<tr>
<td>12</td>
<td>7.34</td>
<td>6.87</td>
<td>30 (30)</td>
<td>79.6</td>
<td>81.8</td>
<td>82.3</td>
</tr>
<tr>
<td>13</td>
<td>6.81</td>
<td>6.40</td>
<td>26 (26)</td>
<td>83.1</td>
<td>83.3</td>
<td>83.7</td>
</tr>
<tr>
<td>14</td>
<td>6.04</td>
<td>5.69</td>
<td>24 (23)</td>
<td>84.6</td>
<td>88.0</td>
<td>87.0</td>
</tr>
<tr>
<td>15</td>
<td>6.97</td>
<td>6.18</td>
<td>26 (26)</td>
<td>82.3</td>
<td>88.7</td>
<td>87.1</td>
</tr>
<tr>
<td>16</td>
<td>6.95</td>
<td>6.91</td>
<td>31 (29)</td>
<td>80.4</td>
<td>81.8</td>
<td>81.3</td>
</tr>
<tr>
<td>17</td>
<td>10.27</td>
<td>8.30</td>
<td>27 (27)</td>
<td>78.7</td>
<td>82.0</td>
<td>81.8</td>
</tr>
<tr>
<td>18</td>
<td>5.20</td>
<td>4.17</td>
<td>25 (21)</td>
<td>81.7</td>
<td>88.2</td>
<td>86.0</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>7.07</strong></td>
<td><strong>6.40</strong></td>
<td><strong>96.3 % &gt;3 mm</strong></td>
<td><strong>81.7</strong></td>
<td><strong>85.0</strong></td>
<td><strong>84.3</strong></td>
</tr>
</tbody>
</table>

References