Multi Atlas Segmentation with Active Shape Model Refinement for Multi-Organ Segmentation in Head and Neck Cancer Radiotherapy Planning

Thomas Albrecht[1], Tobias Gass[1], Christoph Langguth[2], and Marcel Lüthi[2]

[1] Varian Medical Systems Imaging Laboratory, Baden, Switzerland
[2] Dept. of Mathematics and Computer Science, University of Basel, Switzerland

Abstract. We describe a segmentation method that was used in the Head and Neck Auto Segmentation Challenge held at the MICCAI 2015 conference. The algorithm consists of two building blocks. First, we employ a multi-atlas segmentation to obtain an initial segmentation for the considered organs at risk. Secondly, we use an Active Shape Model (ASM) segmentation to refine the initial segmentation of some of the organs. Leave-one-out experiments with the training data were used to determine suitable parameters for the individual steps of the segmentation. The ASM refinement resulted in improved segmentation for the optic nerves and submandibular glands, while for the brain stem, parotid glands, chiasm, and mandibular bone, the multi-atlas segmentation was preferable. Our submission achieved the second rank in the challenge.

1 Introduction

The aim of the Head and Neck Auto Segmentation Challenge held in conjunction with MICCAI 2015 was to evaluate fully automatic segmentation algorithms for the segmentation of organs at risk for radiotherapy planning of head and neck cancer. The challenge consisted of an "offline" part with 10 cases to be segmented before the challenge, and an "online" part with 5 cases to be segmented on the day of the challenge. In addition to these 15 test cases, 25 training cases with "ground truth" segmentations for the considered organs at risk were supplied by the organizers.

Our proposed algorithm uses a multi-atlas segmentation followed by an Active Shape Model (ASM) fitting [3] to refine the segmentation for individual organs. Figure 1 shows a schematic overview of the approach. We used the 25 test cases as atlases for the multi-atlas segmentation and as training data for the ASM models. No additional training data was used. We used leave-one-out experiments on the training cases to determine suitable values for the free parameters in the multi-atlas segmentation and the Active Shape Models. These experiments showed that for some of the structures, the multi-atlas segmentation result could be improved by the ASM refinement (optic nerves and submandibular glands), while for others there was no significant improvement. In these cases

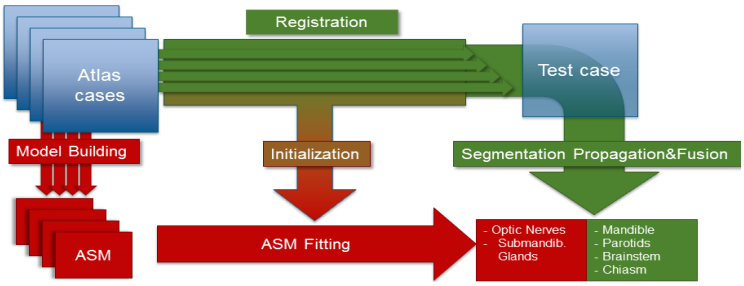

Fig. 1. Overview of the method: The atlas (training) cases are used for Active Shape Model (ASM) building and multi-atlas segmentation. For some organs, the multi-atlas result is used to initialize the ASM fitting (red), for others, it is used as the final segmentation result (green).

it was therefore preferable to use the multi-atlas results as the final segmentation (brain stem, parotid glands, chiasm, and mandibular bone).

2 Method

2.1 Multi-Atlas Segmentation (MAS)

Multi-atlas segmentation is based on non-rigidly registering a number of atlas images onto the input image, propagating the atlas segmentations to the target image, and then fusing them into a single segmentation with a fusion algorithm. The atlas cases are chosen from the training examples provided for the challenge. The MAS work flow is implemented using individual algorithms for registration and fusion. These were chosen due to their performance in our leave-one-out evaluations as well as their availability as an in-house solution or an open source algorithm. We present the details in the following.

Rigid Registration For the initial rigid registration of the atlases to the target image, we used Varian's in-house automatic rigid alignment algorithm. It is based on detecting a small number of obvious landmark points and then computing the unique rigid transform that minimizes the distance between these points.

Non-Rigid Registration After rigid alignment, the atlas images are registered non-rigidly onto the target image. Based on its performance in a number of recent public studies and internal tests, we selected the DEEDS algorithm [4] for computing the deformable registration. Its core strengths are a robust image similarity metric comparing self-similarity context (SSC) features and an efficient global optimization using a Markov random field defined on a minimum spanning tree. It is also comparatively efficient, with one registration requiring approximately 1-5 minutes of computation time.

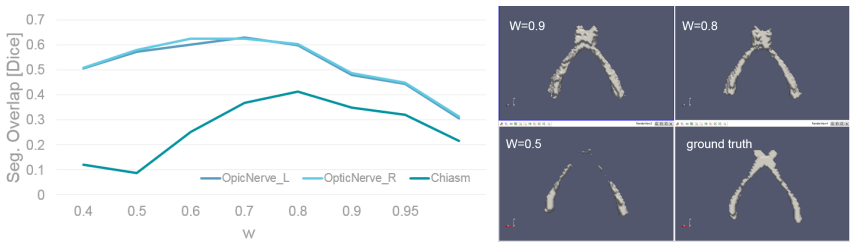

Fig. 2. Optimizing majority-vote segmentation fusion by adjusting the bias parameter w in Eq. (1).

Multi-Atlas Fusion Once the atlas images are registered, the atlas segmentations can be propagated to the target image with the deformation fields computed by the DEEDS algorithm. This results in multiple segmentation candidates per organ, which we fuse individually (one fusion per organ) using weighted majority voting. For N binary segmentation images $s_n(x)$, the fused result for weight $w \in [0, 1]$ is given as:

$$s(x) = \begin{cases} 0, & \text{if } \frac{1}{N} \sum_{n=1}^{N} s_n(x) > w \\ 1, & \text{otherwise.} \end{cases} \tag{1}$$

A weight $w = 0.5$ results in a fair majority vote. For $w > 0.5$ we introduce a bias resulting in more pixels classified as foreground. This can improve the segmentation quality, for example, for very thin structures such as the optic system that frequently have a local thickness of a single slice.

Group-wise Atlas Pre-Selection For the offline challenge, we used all N=25 training cases as atlas images. For the online challenge, this was not possible due to run-time constraints. We therefore selected a subset of the training images as atlases using a greedy group-wise scoring. We start by selecting the two atlases which perform best as single atlases, measured by the dice coefficient, in a leave-one-out study. Then, we provisionally add each of the remaining candidates as a third atlas, and compare the resulting $N-2$ different atlas sets in terms of multi-atlas segmentation quality on the remaining $N-3$ training images, finally choosing the one that results in the best MAS results. The same is repeated to choose the fourth atlas and so on. The resulting atlas set has the potential to cover a wide range of anatomical variability, while avoiding unsuitable atlases, e.g. due to image artifacts or abnormal anatomy.

2.2 Active Shape Model Fitting

The registrations and the MASs are used to initialize an Active Shape Model (ASM) fitting [3].

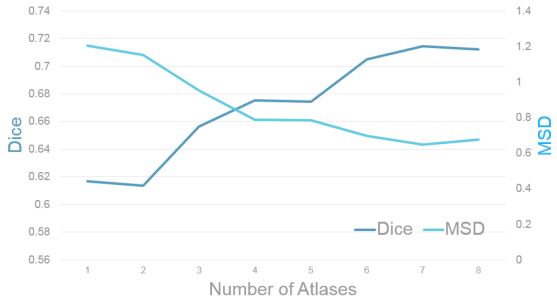

Fig. 3. Group-wise atlas selection on the training set as described in Sec. 2.1. Segmentation accuracy saturates at about seven atlases

ASM Computation We built a separate Active Shape Model for each organ, using only the 25 training cases provided for the challenge. For initial rigid alignment of the training data, we manually placed 4 landmarks on each of the organs. Because of the limited number of training datasets, the statistical shape model is likely to not fully capture the anatomical variability of the organs. Therefore, we slightly enlarge the shape variability by augmenting the model with smooth, low-frequency deformations, as described in [5]. In order to model the image intensities in the neighborhood of the shape model, we employ intensity profiles along the surface normal, as described in the original ASM publication [3].

ASM Initialization In order to fit the shape model to the target image it first needs to be aligned to this image with a rigid transformation. In principle, this can be done for all organs at once, using a single global transformation. However, our experiments have shown that this global transformation can result in poor alignment for some of the organs. Instead, we propose to align each organ individually. This can be achieved by transforming the known landmarks of each atlas organ into the target image using the non-rigid registrations computed during the MAS. This results in several hypotheses for each landmark (one per atlas case and registration), that are fused using their component-wise median and then used to compute an optimal rigid transformation. Furthermore, to constrain the model to remain roughly in the area of the initialization, we compute a posterior shape model using the propagated landmarks [1]. Finally, we compute an initial guess for the model parameters by fitting the model's shape to the surface that resulted from the MAS.

ASM Fitting Starting from this initial shape, the original ASM fitting algorithm described by Cootes and Taylor in [3] is used to fit the ASM to the CT image. This algorithm aims at placing the boundary points of the shape model at image points that have similar intensity profiles. The result is a shape model instance whose boundary is expected to coincide with the organ boundary.

Initializations	Global rigid	Local rigid	Local rigid and MAS
brainstem	0.61	0.86	0.84
optic chiasm	0.00	0.51	0.51
optic nerves	0.00	0.56	0.59
parotid glands	0.30	0.74	0.78
submand. glands	0.16	0.65	0.67

Table 1. Median dice coefficients for different initialization methods for the ASM.

	Multi Atlas			ASM		
Organ	Dice	Hausdorff	Average	Dice	Hausdorff	Average
brainstem	0.84	**6.3**	**1.7**	**0.85**	6.7	1.8
optic chiasm	**0.51**	4.9	**1.2**	0.43	**4.6**	1.3
parotids	0.79	**13.2**	**2.2**	**0.80**	14.3	2.4
optic nerves	0.59	**3.7**	0.9	**0.64**	3.8	**0.8**
submand.	0.67	**10.5**	2.4	**0.71**	10.7	**2.2**

Table 2. Median metrics for initial segmentation and after ASM fitting (best values for each organ in bold face). For our challenge submissions, the organs in the lower half were segmented with the ASM, while for the organs in the upper half we directly submitted the multi-atlas segmentation.

3 Results

We evaluated all parameters using a leave-one-out strategy on the 25 training cases. I.e., the multi-atlas segmentations were computed using up to 24 segmentation hypotheses per organ, and ASMs were built from the same cases.

Registration and Multi-Atlas Fusion First, parameters for DEEDS non-rigid registration [4] where tuned based on segmentation overlap of propagated and target segmentations. For the offline challenge (10 test cases delivered before the challenge), we used an image pyramid with 7 levels with downsampling factors 7,6,5,4,3,2,2; 5,4,3,2,2,2,2 registration label samples per cardinal direction, and 3,2,1,1,1,1,1 voxels spacing between samples. The regularization parameter α was empirically set to 1.2. The resulting run time was about 5 minutes. For the online challenge, this had to be improved in order to be able to segment all test cases in under 2 hours. We therefore changed the image pyramid to 3 levels with downsampling factors 8,4,2; the samples to 6,4,2 and the spacing to 3,2,1. Furthermore, we selected a subset of 7 atlases based on their joint performance on segmenting the remaining atlases as described in Section 2.1. The resulting segmentation quality using different atlases can be seen in Figure 3.

For the segmentation fusion, we found best results by setting w to 0.66 in Equation (1), classifying a pixel as foreground if more than 34% of the target

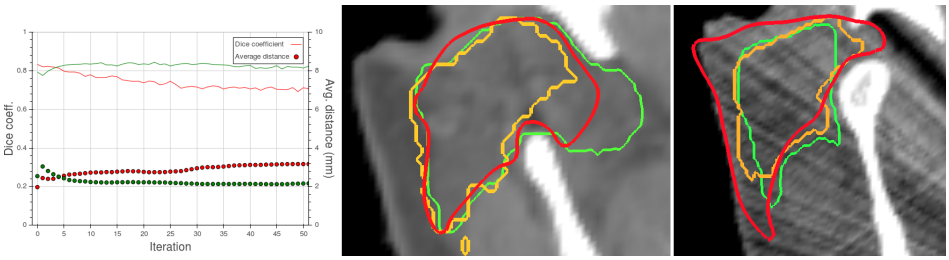

Fig. 4. Examples of ASM fitting, ground truth shown in green, MAS in orange, ASM in red. For case 0057 (middle), the ASM improves the MAS result, for case 0147 (right) the ASM fit is worse than the MAS. The dice coefficients and average surface distances are depicted on the left, showing how successive iterations improve or degrade the result for these two cases.

segmentations classify it as foreground. For very 'thin' structures such as the optic nerves, with an extent of only one voxel in some directions, an even stronger bias ($w = 0.8$) further improved the segmentation, see Figure 2.

ASM Computation, Initialization and Fitting Our experiments led us to employ a relatively large number of profiles (up to 2600) even for small structures. Each profile consists of 7 points, spaced evenly with a distance of 1 mm. We apply Gaussian smoothing ($\sigma = 1mm$) to all input images. Experiments with multi-resolution ASMs showed no improvement in fitting accuracy.

We evaluated the different initialization methods described in Section 2.2, see Table 1. It can be observed that the global rigid transform computed from the provided set of landmarks fails to align the models properly. Using the propagated landmarks to estimate a "local" rigid transform for each organ individually substantially improves the initial results. Further fitting the model parameters to the multi-atlas segmentation (MAS) results in a slightly better initialization. Due to the relatively good initialization, the number of iterations needed in the final ASM fitting is small, and was empirically determined to lie between 5 and 10, depending on the organ. For some of the cases, we observed a degradation of the fitting result with increasing numbers of iterations (see Figure 4). Likely causes for such degradations are poorly generalizing models due to the small number of training cases or misleading intensity information. However, the effect of potential degradation is limited by the low number of fitting iterations.

Table 3. Challenge results

		Case	Brain-stem	Chiasm	Mandib.	Optic Nerves		Parotids		Submand.	
						Left	Right	Left	Right	Left	Right
offsite	Dice	0555	0.73	0.48	0.92	0.65	0.74	0.83	0.83	0.59	0.51
		0576	0.87	0.49	0.72	0.70	0.74	0.76	0.84	0.57	0.68
		0598	0.84	0.79	0.92	0.65	0.76	0.87	0.88	0.69	0.58
		0659	0.84	0.68	0.93	0.53	0.71	0.80	0.77	0.72	0.72
		0661	0.89	0.65	0.90	0.53	0.63	0.88	0.80	0.79	0.82
		0667	0.76	0.59	0.89	0.69	0.59	0.85	0.89	0.75	0.78
		0669	0.90	0.64	0.92	0.71	0.65	0.83	0.83	0.66	0.77
		0708	0.88	0.64	0.90	0.45	0.68	0.88	0.86	0.73	0.80
		0727	0.85	0.50	0.91	0.61	0.74	0.84	0.82	0.80	0.82
		0746	0.88	0.28	0.78	0.41	0.51	0.82	0.61	0.77	0.80
		Avg.	0.84	0.57	0.88	0.59	0.67	0.84	0.81	0.71	0.73
	Distance(95th)	0555	7.83	3.00	2.00	2.39	1.69	3.92	5.16	7.80	9.36
		0576	3.97	3.21	5.61	2.50	2.35	6.58	4.45	6.96	5.74
		0598	4.75	2.00	1.41	3.25	3.00	4.24	3.50	6.07	8.59
		0659	4.27	1.84	2.01	3.85	1.84	4.97	5.62	4.80	5.04
		0661	3.05	2.48	2.48	3.97	3.27	3.46	5.12	3.87	2.97
		0667	10.01	2.37	2.37	2.62	2.62	6.92	2.84	6.61	4.01
		0669	2.50	1.81	2.35	2.36	2.62	5.09	5.60	6.20	3.76
		0708	3.24	2.00	2.47	4.24	2.50	2.91	3.76	5.03	3.65
		0727	4.60	3.16	2.67	3.40	2.08	5.63	8.25	3.76	3.54
		0746	3.10	4.42	5.76	5.97	3.60	5.44	12.28	4.43	4.98
		Avg.	4.73	2.63	2.91	3.45	2.56	4.92	5.66	5.55	5.17
		Rank	3	**1**	4	**1**		2		2	
onsite	Dice	0788	0.89	0.54	0.90	0.70	0.67	0.81	0.80	0.76	0.80
		0806	0.85	0.44	0.91	0.56	0.56	0.76	0.76	0.76	0.78
		0845	0.85	0.47	0.87	0.66	0.55	0.83	0.81	0.68	0.79
		0857	0.85	0.58	0.88	0.75	0.55	0.83	0.80	0.76	0.64
		0878	0.86	0.58	0.87	0.47	0.50	0.82	0.77	0.82	0.55
		Avg.	0.86	0.52	0.89	0.63	0.57	0.81	0.79	0.76	0.71
	Distance(95th)	0788	3.15	3.26	2.67	3.00	4.79	5.89	6.81	4.86	5.10
		0806	4.44	3.96	2.37	3.16	4.60	5.87	7.76	5.12	3.79
		0845	5.00	2.42	2.50	2.52	4.02	4.90	5.64	5.71	4.18
		0857	4.69	2.34	3.00	2.08	4.79	6.97	6.63	6.55	6.73
		0878	4.22	3.48	2.81	3.21	4.28	3.89	8.52	3.81	9.92
		Avg.	4.30	3.09	2.67	2.79	4.49	5.50	7.07	5.21	5.94
		Rank	2	**1**	4	3		2		2	

Challenge Results Table 2 shows the final median segmentation quality we could achieve on the training data with the MAS and the ASM fitting. The best values are displayed in bold. Based on these, we chose to employ the ASM for the optic nerves and submandibular glands and use the multi atlas result for the remaining structures. For the mandible, experiments showed that the MAS substantially outperformed the ASM.

Our submission fared well in the competition, see Tab. 3, achieving second rank for both the offline and on-site challenge. Interestingly, our results did not vary substantially between the offline and online competition, indicating that using the seven preselected atlases and the sped-up registration parameterisation did not substantially deteriorate the results.

4 Conclusion

We submitted a combination of a state-of-the-art multi atlas segmentation (MAS) and ASM fitting to the head-neck segmentation challenge, and achieved good results with an average second rank. For the ASM segmentations, a big impact of accurate initialization using the propagated landmarks from the registered atlases could be observed. For the MAS, it is notable that competitive performance could be achieved by using a group-wise preselection of atlases and a fast registration method that resulted in a MAS runtime of about 7 minutes per test case. Further improvements could potentially be achieved (at the cost of execution speed) by employing more advanced segmentation fusion algorithms like such as STEPS [2].

References

1. Albrecht, T., Lüthi, M., Gerig, T., Vetter, T.: Posterior shape models. Medical Image Analysis 17(8), 959 – 973 (2013), http://www.sciencedirect.com/science/article/pii/S1361841513000844
2. Cardoso, M.J., Leung, K., Modat, M., Keihaninejad, S., Cash, D., Barnes, J., Fox, N.C., Ourselin, S.: STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcelation. Medical Image Analysis (mar 2013), http://linkinghub.elsevier.com/retrieve/pii/S1361841513000200
3. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., et al.: Active shape models-their training and application. Computer vision and image understanding 61(1), 38–59 (1995)
4. Heinrich, M.P., Jenkinson, M., Brady, M., Schnabel, J.A.: MRF-based deformable registration and ventilation estimation of lung CT. IEEE Transactions on Medical Imaging 32(7), 1239–48 (jul 2013), http://www.ncbi.nlm.nih.gov/pubmed/23475350
5. Lüthi, M., Jud, C., Vetter, T.: A Unified Approach to Shape Model Fitting and Non-rigid Registration. In: Wu, G., Zhang, D., Shen, D., Yan, P., Suzuki, K., Wang, F. (eds.) Machine Learning in Medical Imaging, Lecture Notes in Computer Science, vol. 8184, pp. 66–73. Springer International Publishing (2013), http://dx.doi.org/10.1007/978-3-319-02267-3_9