Towards Automatic Plan Selection for Radiotherapy of Cervical Cancer by Fast Automatic Segmentation of Cone Beam CT Scans

Thomas Langerak[*], Sabrina Heijkoop, Sandra Quint, Jan-Willem Mens,
Ben Heijmen, and Mischa Hoogeman

Department of Radiation Oncology, Erasmus-MC Cancer Center,
Rotterdam, The Netherlands
`{t.langerak,s.heijkoop,s.quint,j.w.m.mens,`
`b.heijmen,m.hoogeman}@erasmusmc.nl`

Abstract. We propose a method to automatically select a treatment plan for radiotherapy of cervical cancer using a Plan-of-the-Day procedure, in which multiple treatment plans are constructed prior to treatment. The method comprises a multi-atlas based segmentation algorithm that uses the selected treatment plan to choose between two atlas sets. This segmentation only requires two registration procedures and can therefore be used in clinical practice without using excessive computation time. Our method is validated on a dataset of 224 treatment fractions for 10 patients. In 37 cases (16%), no recommendation was made by the algorithm due to poor image quality or registration results. In 93% of the remaining cases a correct recommendation for a treatment plan was given.

1 Introduction

One of the main treatment modalities for cervical cancer is External Beam Radiotherapy (EBRT) combined with brachytherapy and optionally chemotherapy. The aim of EBRT is to deliver an appropriate dose of radiation to the Clinical Target Volume (CTV), while minimizing the dose delivered to the Organs At Risk (OAR) that surround the CTV. For cervical cancer, the CTV typically consists of the upper part of the vagina, the cervix, the uterus and the nodal CTV. The most important OARs are the bladder in the anterior direction, the rectum in the posterior direction and the small bowels in the cranial direction (see Figure 1).

One of the main challenges when treating cervical cancer is that in the 23 treatment days the shape and position of the CTV exhibit a large day-to-day variation, depending on the amount of bladder filling (compare figure 1a to figure 1b). To ensure that the CTV is radiated, a large safety margin is used and as a result healthy tissue is unnecessarily irradiated. State-of-the-art treatment therefore uses a so-called Plan-of-the-Day procedure (Bondar et al., 2012) that selects at every treatment day the treatment plan that best fits the daily anatomy of the patient.

---

[*] Corresponding author.

P. Golland et al. (Eds.): MICCAI 2014, Part I, LNCS 8673, pp. 528–535, 2014.

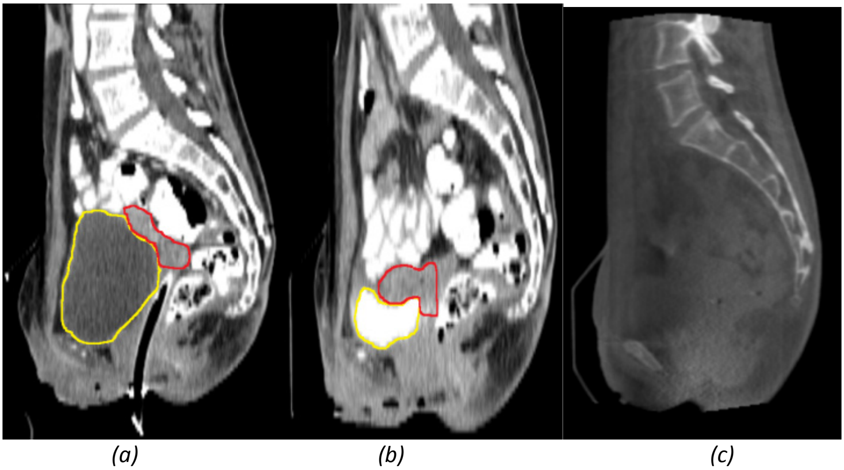

Fig. 1. Sagittal slices of CT scans of a patient (a) with a full bladder and (b) with an empty bladder. In both CT scans, both the bladder and the uterus are delineated. Figure c shows a CBCT scan of the same patient.

In our implementation of this procedure two treatment plans (TP_f and TP_e) are constructed prior to treatment based on two CT scans of the patient with varying bladder fillings. These scans were already part of normal clinical practice in order to be able to determine the extent of motion of the uterus as a result of the amount of bladder filling. Treatment plan TP_f can be used to treat the patient when the bladder is full to half-full; if the patient has an empty to half-full bladder, then TP_e can be used. It is not difficult to generate more treatment plan in the range from full to empty bladder, but the benefit of such a strategy has not been proven. In fact, existing research (Bondar et al., 2012) concludes that there is no significant benefit to more than two plans.

On the day of treatment, based on a Cone Beam CT (CBCT) scan that is acquired just before treatment, one of these two plans is then selected to be delivered to the patient. CBCT scans typically exhibit low soft tissue contrast, which makes it difficult to determine the exact location and shape of all relevant organs (see Figure 1c). Therefore, in our clinical procedure if no treatment plan can be reliably selected, a back-up plan (TP_b) is chosen that guarantees adequate coverage of the CTV but does not spare the OAR. As patients are typically treated in 23 treatment fractions, this procedure has to be performed 23 times. Although this is not part of the current clinical practice yet, after the delivery of the treatment the CBCT scan can be segmented to support a detailed analysis of the treatment up to that treatment day. In such an analysis the segmentation of each CBCT scan is coupled to the selected treatment plan in order to be able to determine what dose has been delivered to what organ. By accumulating this information over all treatment fractions it can be determined, at the end of treatment, what dose has been delivered to the target and to healthy tissues.

Selection of the correct treatment plan is currently done by a team of radiotherapy technicians (RTTs), who need to quickly interpret the daily CBCT scan using only visual assessment and identify the best treatment plan. Because they are under time pressure, RTTs tend to select the back-up treatment plan if they are uncertain about the treatment plan to use, for example as a result of poor image quality. Currently, there are only two treatment plans to choose from, which limits the complexity of the plan choice. In the future, however, we hope to further improve the sparing of healthy tissues by extending the number of treatment plans, such that all possible shapes of the target are better represented. Consequently, the complexity of the plan choice and time pressure will increase and automatic treatment plan selection becomes a necessity.

In this paper we present, to our knowledge, the first attempt to automate plan selection for cervical cancer. Our method uses all the images acquired on previous treatment days as an atlas in a multi-atlas-based segmentation procedure of the image that is acquired on the current day. A measure that is based on the agreement of the set of atlases expresses the confidence of the correctness of the segmentation. Based on this segmentation, a recommendation is made for what treatment plan to be used.

In addition, we present a method that re-computes the segmentation using only a subset of the atlases. Selection of the atlases is done based on the recommendation for a treatment plan. It can be expected that using a more dedicated atlas set leads to a more accurate segmentation that can be used for an interim or retrospective evaluation of the treatment. If the segmentation is of sufficient quality it can even form the basis for constructing a new treatment plan which can then be added to the treatment plan library, although this falls outside the scope of this paper.

In section 2 the existing literature on atlas selection is briefly reviewed as well as the literature on the Plan-of-the-Day procedure. Section 3 proposes the method, section 4 describes the data and experiments, section 5 presents the results and section 6 discusses our findings, suggests further research, and draws a conclusion.

2 Review of Existing Literature

Multi-atlas based segmentation was popularized in the first decade of this century by, amongst others, Heckemann et al., 2006, Aljabar et al., 2009, Rohlfing et al., 2005 and Klein et al., 2008. Following the initial investigations into the concept of multi-atlas-based segmentation in general, the focus of new research came to lie on the label fusion aspect of multi-atlas based segmentation. Most of these methods are based on the pioneering work by Warfield et al., 2004, on expectation maximization and include variants that use atlas selection (Langerak et al, 2010), local similarity measures (Asman, 2012; Commowick, 2012) and probabilistic models 13 et al., in press). Atlas selection has been proposed based on image similarity (Aljabar et al., 2009) and on segmentation similarity (Langerak et al., 2013)

Plan-of-the-day procedures have been described for radiotherapy treatment of cervival cancer (Bondar et al., 2012), bladder cancer (Murthy et al., 2011) and prostate cancer (Gill et al., 2013). In all of these cases, using a Plan-of-the-Day treatment is motivated by variable bladder- and rectum filling that influences the shape and position of the target.

3 Method

We assume that the first fraction CBCT scan, indicated as $CBCT_1$, was manually segmented by an expert and we use this scan to initialize both atlas sets A_f and A_e. It may seem more logical to use CT_f and CT_e to initialize the atlas sets, but these differ significantly from the CBCT scans, among others because a vaginal catheter is inserted to indicate the lower and upper part of the vaginal wall during the acquisition of the CT scans. This catheter is not present on the CBCT scans, which considerably hinders registration of the CT scans. Inserting a catheter during the acquisition of the CBCT scans to increase uniformity and make registration easier is not an option because of the decrease in patient comfort and a considerable increase in the time needed for treatment. Segmenting $CBCT_{i>1}$ was done using the following procedure:

1. $CBCT_1$ and $CBCT_i$ are registered in both directions using a rigid registration followed by a B-Spline driven non-rigid registration. Mutual information was used as the image similarity metric and we used a LBFGSB optimizer as provided by ITK.
2. The segmentations of all $CBCT_{j<i}$ are propagated to $CBCT_i$ indirectly, via $CBCT_1$. These propagated segmentations are denoted as S^0 to S^{i-1} and combined to a single result. To combine segmentations, the SIMPLE algorithm (Langerak et al., 2010) was used, but we do not expect any significant difference in outcome when using another label fusion method and therefore refer to the above mentiond paper for details of the label fusion method. The default parameters mentioned in the paper were used. The resulting segmentation is noted as S_i.
3. If the volume of the bladder $|B_i|$ in S_i is larger than $(|B_f|+|B_e|)/2$, then it is recommended to use TP_f for treatment. Otherwise TP_e is recommended. A confidence level is given by analyzing the variability of all individual atlas-based segmentations.
4. If the full bladder treatment plan was recommended then A_f is selected to be used in step 5, otherwise A_e is selected.
5. $CBCT_i$ is re-segmented using the selected atlas set. Note that this step hardly takes additional computation time because the necessary registrations were already computed in step 1. This is an automatic correction that differs from step 2 only in the fact that not all segmentations are used as an atlas, but only the subdivisions A_f and A_e.

6. $CBCT_i$ and its segmentation are manually corrected and approved by an RTT and added to the appropriate atlas set A_f or A_e.

The confidence level that is mentioned in step 3 is computed as the smallest agreement between any two propagated segmentations and is defined as $\min_{S1,S2\in[S_0,S_{i-1}]} DSC(S1,S2)$, where DSC is the Dice Similarity Coefficient. If the confidence level drops below a certain threshold, no recommendation for a treatment plan is given. In this case the segmentation is not added to either atlas set.

By using $CBCT_1$ as a reference image, this procedure only takes one forward and one backward registration per treatment fraction. The initial segmentation based on which a treatment plan is chosen is not necessarily very accurate because, due to the shape variation, some of the atlases will not register well to the target image. However, as it is not used as a final segmentation, but only to make a decision between the two atlas sets, it does not have to be very accurate. I.e. if bladder volumes vary between the full bladder volume $|B_f|$ and the empty bladder volume $|B_e|$ and n treatment plans are computed, then each treatment plan $T_{0\le m<n}$ covers the bladder volume range $[|B_e|+(|B_f|-|B_e|)\frac{m}{n},|B_e|+(|B_f|-|B_e|)\frac{m+1}{n,}]$, and therefore segmentation volumes are allowed to be off by $\frac{(|B_f|-|B_e|)}{n}$, assuming that B_f fully overlaps B_e. In other words: the fewer sub-ranges and the larger the volume of the bladder in the full-bladder scan, the less accurate the segmentation needs to be to be able to choose the correct treatment plan.

4 Data and Experiments

In a retrospective study, we investigated a total of 234 treatment fractions for 10 patients: 9 patients were treated in 23 fractions and 1 patient was treated in 27 fractions. These patients were selected from our database because full- and empty-bladder treatment plans were available for these patients because the variations in bladder filling resulted in large motion of the uterus. In all CBCT scans three structures were manually segmented by an expert: the uterus, the bladder and the rectum. These segmentations served as a ground truth segmentation.

Not taking into account the first treatment fraction (in which the CBCT scan was manually delineated), in total 224 treatment plan decisions were made. For each patient i and fraction j, the chosen treatment plans were recorded as $TP_j^i \in \{TP_e, TP_f, TP_b\}$, where TP_e means that an empty-bladder treatment plan was used, TP_f stands for a full-bladder treatment plan and TP_b represents the back-up plan.

In 42 cases a back-up plan was chosen by the RTTs and no choice was made between an empty- or a full-bladder treatment plan. In these cases, a treatment plan decision was made retrospectively by an expert clinician.

First, the treatment plan that was recommended by our method was tested against the manual treatment plan decisions to validate the automatic treatment plan suggestion. In addition, it was tested whether the confidence measure correctly warned for inaccurate treatment plan recommendations. Finally, the accuracy of the re-segmentation using a dedicated atlas set was tested against the manual ground truth segmentation of the CBCT scans.

5 Results

Of the 224 treatment plan decisions, our method did not make a recommendation in 37 cases. A minimal confidence level of 0.3 was used that was experimentally determined. In 24 of these cases, the method would indeed have recommended the wrong treatment plan. For 173 of the remaining 187 cases, a correct recommendation was made, so our method achieved a 93% accuracy. A clinical investigation into whether the treatment plan suggestion supported the RTTs in their decision as a result of which the back-up plan was chosen less often is left for further research.

The refined segmentation that was computed with either the empty-bladder or the full-bladder atlas set was compared to the manual segmentations. The results are shown in Figure 2 in the form of a boxplot that indicates the median and the quartiles of the distribution of the accuracy. The DSC score for the bladder seems higher than for the Uterus and Rectum, but this is mainly due to the size of the bladder. Cases in which confidence levels were too low to recommend a treatment plan, whether correct or not, were not included in these results.

From this figure, it can be concluded that the segmentations that are computed are highly accurate, especially considering the fact that the underlying images are CBCT images, but not good enough for fully automatic segmentation. In our clinical practice, these segmentations are therefore used as an initial estimate that is manually corrected.

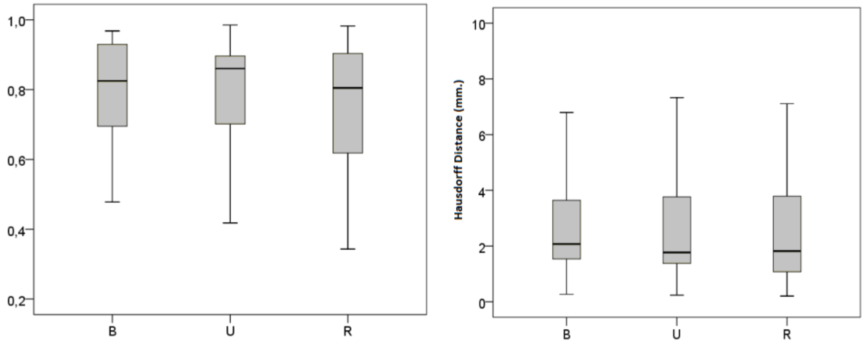

Fig. 2. Dice Similarity Coefficient (left) and Hausdorff distance in mm. (right) of the refined segmentations of the Bladder (B), Uterus (U), and Rectum (R).

6 Conclusions and Discussion

Given previous attempts to register images of cervical cancer patients, our results are encouraging but there are some limitations to our findings, mainly the fact that it is unknown what the inter- and intra-observer variability of manual segmentations is. Existing research suggests that the inter-observer variability is large when segmenting CBCT images, but this has only been investigated for prostate images. Our results show that the accuracy for atlas-based segmentation of the bladder is higher for the bladder than for the rectum and uterus. In the future we will therefore investigate a strategy to select atlases based on uterus shape and rectum filling in addition to bladder filling.

Furthermore, in section 3 we derived that the required accuracy of the rough initial segmentation increases with the number of available treatment plans. In our clinical practice only two treatment plans are used and this represents the easiest test case. It remains to be shown that our method also works in situations where more treatment plans are used, as a result of which the demands on the accuracy of the method become more strict.

In a visual inspection of the cases in which confidence was low, we noticed that in most cases where the low confidence was unjustified, it was caused by a single failed segmentation. In the future we plan to investigate whether it makes more sense to measure confidence as the average overlap between propagated atlas segmentations rather than the minimal overlap.

References

1. Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D.: Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. Neuroimage 46, 726–738 (2009)
2. Asman, A.J., Landman, B.A.: Formulating Spatially Varying Performance in the Statistical Fusion Framework. IEEE Trans. Med. Imaging 31(6), 1326–1336 (2012)
3. Bondar, M.L., Hoogeman, M.S., Mens, J.W., Quint, S., Ahmad, R., Dhawtal, G., Heijmen, B.J.: Individualized nonadaptive and online-adaptive IMRT treatment strategies for cervical cancer patients based on pre-treatment acquired variable bladder filling CT-scans. Int. J. Radiat. Oncol. Biol. Phys. 83(5), 1617–1623 (2012)
4. Commowick, O., Akhondi-Asl, A., Warfield, S.K.: Estimating a reference standard segmentation with spatially varying performance parameter: local MAP STAPLE. IEEE Trans. Med. Imaging 31(8), 1593–1606 (2012)
5. Gill, S., Pham, D., Dang, K., Bressel, M., Kron, T., Siva, S., Tran, P.K., Tai, K.H., Foroudi, F.: Plan of the day selection for online image-guided adaptive post-prostatectomy radiotherapy. Rad. Onco. 107(2), 165–170 (2013)
6. Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. Neuroimage 33, 115–126 (2006)
7. Langerak, T.R., van der Heide, U.A., Kotte, A.N.T.J., van Vulpen, M., Viergever, M., Pluim, J.P.W.: Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). IEEE Transactions on Medical Imaging 29(12), 2000–2008 (2010)

8. Klein, S., van der Heide, U.A., Lips, I.M., van Vulpen, M., Staring, M., Pluim, J.P.W.: Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. Medical Physics 35(4), 1407–1417 (2008)
9. Langerak, T.R., Berendsen, F.F., van der Heide, U.A., Kotte, A.N.T.J., Pluim, J.P.W.: Multi-atlas-based segmentation with preregistration atlas selection. Med. Phys. 40(9), 091701 (2013)
10. Murthy, V., Master, Z., Adurkar, P., Mallick, I., Mahantshetty, U., Bakshi, G., Tongaonkar, H., Shrivastava, S.: 'Plan of the day' adaptive radiotherapy for bladder cancer using helical tomotherapy. Radiother. Oncol. 99(1), 55–60 (2011)
11. Rohlfing, T., Brandt, R., Menzel, R., Russakoff, D.B., Maurer Jr., C.R.: Quo vadis, atlas-based segmentation? In: Suri, J.S., Wilson, D.L., Laxminarayan, S. (eds.) Handbook of Biomedical Image Analysis. Topics in Biomedical Engineering International Book Series, pp. 435–486 (2005)
12. 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), 903–992 (2004)
13. Wu, G., Wang, Q., Zhang, D., Nie, F., Huang, H., Shen, D.: A generative probability model of joint label fusion for multi-atlas based brain segmentation. Med. Image Anal. (in press)