A multi-atlas approach for prostate segmentation in MR images

Geert Litjens, Nico Karssemeijer, and Henkjan Huisman

Diagnostic Image Analysis Group, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

Abstract. Prostate segmentation is an important and often mandatory step for several tasks, for example volume estimation, radiotherapy planning and computer-aided detection of prostate cancer. In this paper we evaluate a multi-atlas segmentation technique to segment the prostate from transversal T2-weighted MR images on data from the Prostate MR Image Segmentation Challenge (PROMISE12). Atlases are registered using localized mutual information as a metric, after which the Selective and Iterative Method for Performance Level Estimation (SIMPLE)-algorithm is used to merge the atlas labels and obtain the final segmentation. Results obtained on the training data show good performance on average with a median Dice coefficient 0.83.

1 Background

Segmentation of the prostate on medical images is an important step in both clinical and image processing work flows. In the clinical setting prostate segmentations are used in for example radiotherapy, but also in prostate volumetry and calculation of diagnostically important metrics likes prostate specific antigen (PSA) density. In image processing segmentation of the organ of choice is usually a mandatory first step such that subsequent algorithms can focus on the region of interest. This usually reduces both algorithm complexity and computation time.

The complexity of segmenting the prostate on MR images is based on two components: the inherent variability of prostate size and shape between men, but also largely because of the MR scanning technique. MRI has a wide range of different imaging protocols, some which have a large impact on image appearance (for example the use of an endorectal coil). In addition, the intensity unit of MR is not standardized, as in CT.

In current literature a wide range of different segmentation techniques is available, both automatic and interactive. Examples are active appearance/shape models[1], multi-atlas registration [2], pattern recognition or combinations of these methods[3]. In this paper we investigate the use of an automatic multi-atlas segmentation algorithm to segment the prostate on transversal T2-weighted MR images. An automatic methods was chosen because in our hospital the number of prostate MRI's made each week is high (up to 40) and segmenting all these cases by hand, even with an interactive tool, is not feasible.

2 Methodology

The process of a multi-atlas algorithm consists of two distinct steps. First, the atlases are registered to the unknown case, after which the obtained transformation is applied to the segmentations of the atlases. Second, the transformed segmentations have to be merged to obtain a final binary segmentation of the organ of interest. The method developed in this paper is based on the methods presented by Klein et al.[2] and Langerak et al.[4], with adaptations to make it suitable to this specific problem. In the following subsections we will discuss these steps in more detail.

2.1 Atlas registration

Let the patient image to be segmented be denoted as $I(x)$, where x is a spatial location within the image. Then we are looking for a labeled image, $S(x)$ that contains an accurate segmentation of the prostate. The following steps are similar in most multi-atlas based systems, a set of N labeled images is non-rigidly registered to the unknown case $I(x)$. The i-th image in this set is denoted as $A_i(x)$. After registration the obtained transformation is applied to the label image of the atlas, $L_i(x)$.

The registration will try to maximize the similarity between $I(x)$ and $A_i(x)$ by deforming the latter. In this approach we will do the registration in two steps. First, we will rigidly align the images, which will give us a simple transformation matrix. Second, a nonrigid registration was applied, parameterized by cubic B-splines. These registrations were all performed in a multi-resolution manner using 3 resolution steps for both the rigid and the non-rigid registration. Downsampling the images was performed by applying Gaussian smoothing at scales of 3, 2 and 1 mm before resampling. The nonrigid B-spline grid spacing for these resolution was 32, 16 and 8 mm respectively. Experiments with both 300 and 2000 iterations were performed to asses the effect of the number of iterations. The optimization was performed using a stochatistic gradient optimization like described in [2].

Another important part of the registration procedure is the similarity metric that is used. In this approach we used the localized mutual information metric[5]. First, let us define the mutual information[6] as:

$$\mathrm{MI}(I, J, \Omega) = \sum_k \sum_m p_{\mathrm{IJ}}(k, m) \log \frac{p_{\mathrm{IJ}}(k, m)}{p_I(k) p_J(m)} \tag{1}$$

Here p_I and p_J denote the marginal intensity probabilities of two images I and J and p_{IJ} represents the discrete joint intensity probability. These can be estimated from a discrete set of intensity pairs, of which the coordinates are sampled from the continuous image volume Ω. However, an assumption of mutual information is that the intensity probabilities do not vary over the image domain. This assumption is typically violated in MR scans due to field inhomogeneities or coil profiles. A solution is to evaluate the mutual information on several subregions

in the image. Adding the resulting per-region mutual information values then gives us the localized mutual information:

$$\mathrm{LMI}(I, J, \Omega) = \frac{1}{N} \sum_{x_j \in \Omega} \mathrm{MI}(I, J; \mathcal{N}(x_j)) \tag{2}$$

Here $\mathcal{N}(x_j)$ is a spatial neighborhood centered around x_j. The total number of neighborhoods used is denoted as N. In this experiment we used a neighborhood size of 50 mm^3. Finally, all registrations in this paper were performed using the Elastix package[7].

2.2 SIMPLE atlas merging

After atlas registration we have a transformed label image $L_i(x)$ for each atlas. From these images we need to construct a single binary segmentation. The simplest method would be to use a majority voting approach which can be defined as:

$$l_c(x) = \frac{\sum_{i \in A} \delta[c, (L_i \circ T_i(x))]}{N} \tag{3}$$

$$S(x) = \arg\max_{c \in \mathcal{C}} l_c(x) \tag{4}$$

Here δ is the Kronecker delta function. This equation selects the most likely class for each voxel, 0 being background and 1 being prostate. For this paper we implemented the Selective and Iterative Method for Performance Level Estimation (SIMPLE) presented by Langerak et al.[4] This iterative method tries to optimize the final segmentation by removing badly registered atlases. The main assumption in this method is that most atlases are registered well.

The SIMPLE process works as follows: first a baseline segmentation is constructed using the majority voting approach. Then we calculate the Dice coefficient between this baseline segmentation and all transformed label images $L_i(x)$ of the atlases. We then remove the atlases with the worst Dice coefficient compared to the baselines segmentation based on a threshold θ. Then a new segmentation is constructed by weighting the remaining atlases with the Dice coefficient and again performing the majority voting equation. These steps iterate until no more atlases are removed. The resulting segmentation is considered the final segmentation.

The threshold θ depends on the number of remaining atlases. In the beginning we want to be careful, we do not want to throw away to many atlases at the start because we cannot get them back later. At the end we can be a little less strict because we have many well performing atlases left, so even the slightly less performing atlases can be removed. The threshold is constructed as:

$$\theta = \bar{D} - \alpha \sigma_D \tag{5}$$

where $\bar{D}$ is the average Dice coefficient over all cases compared to the segmentation at the current iteration, σ_D is the standard deviation of the Dice coefficient and α is a scalar number which is a function of the number of segmentation. We choose α to start at 2 and then linearly decrease to 1 when we only have 10 atlases remaining.

3 Experimental Design

Our approach was evaluated using the training data provided by the PROMISE12 challenge (http://promise12.grand-challenge.org/). The training data included a total of 50 cases from different centers, different vendors, field strengths and acquisition protocols. All experiments where done in a leave-one-out-manner. This means that for each segmentation we removed the image to be segmented from the set of atlases. Additionally, we only used the data provided in the challenge, we did not use any external data.

We decided to divide the atlases into two categories: atlases where an endorectal coil is present and atlases without. Before each registration we calculated the mutual information between the test image and the two sets. The set with the highest average mutual information was then used to perform the segmentation. In all cases the correct set was identified in this approach. This helps us reduce computation time and improve results. In a real clinical setting the task of selecting the correct set would be even easier, because then you could read this from the DICOM header.

In our experiments we will use the Dice coefficient to evaluate the results of the segmentation, as it is one of the metrics that is going to be used for the challenge. We will evaluate the atlas segmentation approach with four different settings:

1. 300 iterations, majority voting
2. 2000 iterations, majority voting
3. 300 iterations, SIMPLE
4. 2000 iterations, SIMPLE

As such we can evaluate whether our merging approach gives additional performance and we can evaluate how many iterations are needed for a accurate segmentation. For the segmentation of the test data in the challenge we will use approach 4.

4 Results and Discussion

4.1 Quantitative Results

The results of our experiments are visualized in figure 1 and table 1. In these figures we can see that the highest average and median Dice are obtained when using both 2000 iterations and SIMPLE label merging. The resultant mean and median Dice are then 0.78 and 0.83. We can see from the figures that in three cases the algorithm seems to give very poor results. In 40 of the 50 cases a Dice coefficient of above 0.7 is obtained and in 28 of the 50 cases a coefficient of above 0.8 is found. In all cases the prostate was correctly localized by the algorithm.

In three of the cases in experiment 4 the atlas segmentation performed poorly, even though it accurately located the prostate. This was caused by the fact those cases were not well represented by the atlases in the system.

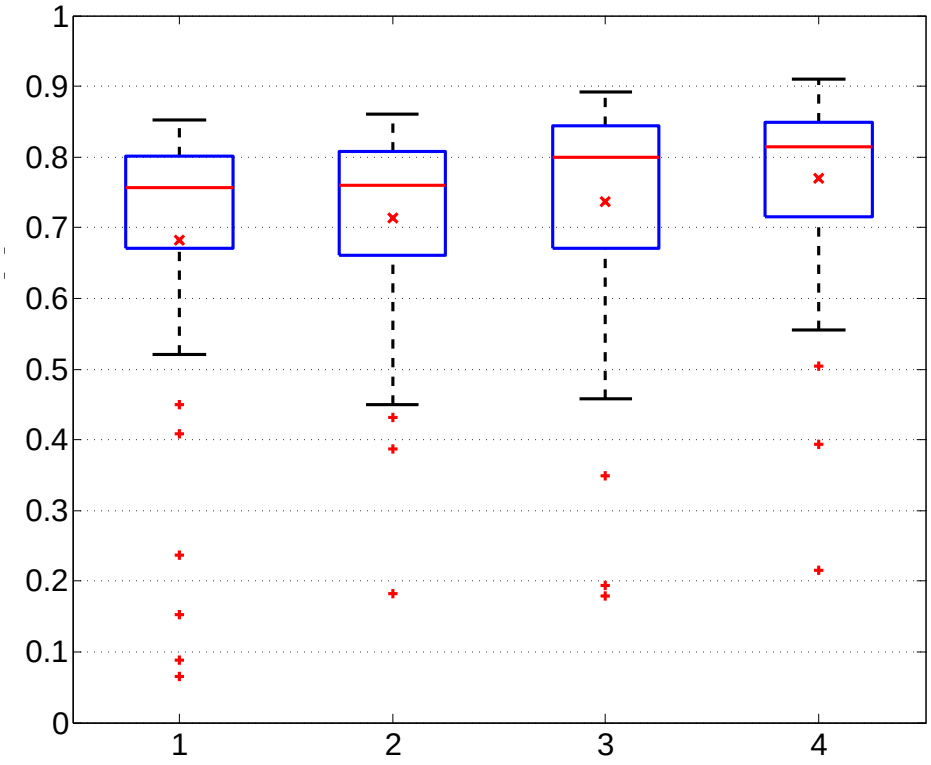


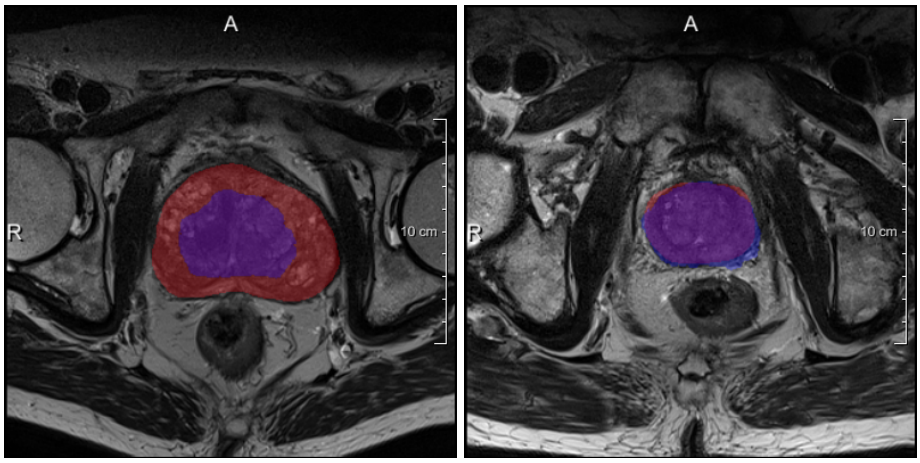
Fig. 1: Dice coefficients for experiments 1 through 4 visualized in box-plots. Experiments 1 and 2 are using 300 registrations iterations, experiments 1 and 3 are using majority voting for label merging instead of SIMPLE. Means are indicated with a red cross inside the box.

Table 1: Numerical representation of the results obtained in the experiments. Experiments 1 and 2 are using 300 registrations iterations, experiments 1 and 3 are using majority voting for label merging instead of SIMPLE. Specified from left to right are the mean, the standard deviation, the median, the inter-quartile range, the minimum and maximum of the Dice coefficients over all cases in the training set

	Mean	STD	Median	IQR	Min	Max
Experiment 1	0.68	0.19	0.75	0.13	0.06	0.85
Experiment 2	0.71	0.14	0.76	0.15	0.18	0.86
Experiment 3	0.74	0.16	0.80	0.17	0.18	0.89
Experiment 4	0.78	0.12	0.83	0.13	0.22	0.91

4.2 Qualitative Results

In figure 2 we show both the worst and the best results from our segmentation algorithm. As you can see, even in the worst result, localization of the prostate was accurate, but the size of the prostate was outside of the scope of the atlases. In the best case we can see that we can get excellent results with this segmentation technique



(a) Worst case result (b) Best case result

Fig. 2: Results for the worst and the best case. In red the reference segmentation is shown and in blue the segmentation resulting from the algorithm is shown

4.3 Implementation details and efficiency

The registrations in this paper were performed with the open-source software package Elastix[7], which is based on the Insight Toolkit (ITK, http://www.itk.org/). The SIMPLE algorithm was implemented in MeVisLab http://www.mevislab.de/[8]. All underlying computationally intensive code was implemented in C++. SIMPLE was partly implemented in Python. The registration was performed on a cluster of standard personal computers. For the registration containing 300 iterations each registration took, on average, 80 seconds. For the registrations using 2000 iterations it took, on average, around 500 seconds. The SIMPLE algorithm takes around 45 seconds to generate the final segmentation. When the algorithm would have been run on a single computer it would have taken 34 minutes for the 300 iteration segmentation and 210 minutes for the 2000 iteration segmentation. However, when using the cluster each computer does one registration so then the time it would take per segmentation would be 2 minutes for the 300 iteration case and 10 minutes for the 2000 iteration case. More details can be found in table 2.

Table 2: Overview of the efficiency and implementation details of our algorithm

	Parameter	Value
Algorithm	*Language:*	C++/Python
	Libraries/Packages:	Elastix, Insight Toolkit, MeVisLab
	GPU Optimizations:	-
	Multi-Threaded:	No
Machine	*CPU Clock Speed:*	2.26 GHz
	Cluster CPU Core Count:	25
	Machine Memory:	24 GB
	Memory Used During Segmentation:	200 MB (peak) during registration
Time	*Segmentation Time (Single):*	33 - 210 minutes
	Segmentation Time (Cluster):	2 - 10 minutes

5 Conclusion

In this paper we evaluated an atlas segmentation system in the context of the PROMISE12 challenge. Our results show that an atlas system is capable of accurately segmenting the prostate in MR images. Additionally, the extra iterations and the SIMPLE label merging technique result in a higher performance compared to a lower number of iterations and majority voting-based label merging.

If we look more closely at the results where the atlas method did not perform well we notice that this usually happens in cases that are not well represented in the atlas set. For example, the worst performing case, Case23, which had a Dice coefficient around 0.2 in all experiments. The prostate in this case had a volume of 325 mL. This is very far from the average prostate volume in the atlases. Specifically including atlases in the system which can cope with the extremes in the data would solve this problem. As it is impossible to add an unlimited number of atlases to the system, adding a second step after atlas segmentation might also improve results, for example, by using an active shape model.

Concluding, we have shown that, when care is taken in defining the set of atlases, a multi-atlas based method for prostate segmentation on T2-weighted MR images gives good results.

References

1. Toth, R., Bloch, B.N., Genega, E.M., Rofsky, N.M., Lenkinski, R.E., Rosen, M.A., Kalyanpur, A., Pungavkar, S., Madabhushi, A.: Accurate prostate volume estimation using multifeature active shape models on T2-weighted MRI. Acad Radiol **18** (2011) 745–754
2. 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. Med Phys **35** (2008) 1407–1417

3. Makni, N., Puech, P., Lopes, R., Dewalle, A.S., Colot, O., Betrouni, N.: Combining a deformable model and a probabilistic framework for an automatic 3d segmentation of prostate on mri. Int J Comput Assist Radiol Surg **4** (2009) 181–188
4. Langerak, T.R., van der Heide, U.A., Kotte, A.N.T.J., Viergever, M.A., van Vulpen, M., Pluim, J.P.W.: Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). IEEE Trans Med Imaging **29** (2010) 2000–2008
5. Studholme, C., Drapaca, C., Iordanova, B., Cardenas, V.: Deformation-based mapping of volume change from serial brain mri in the presence of local tissue contrast change. IEEE Trans Med Imaging **25**(5) (May 2006) 626–639
6. Viola, P., Wells III, W.M.: Alignment by maximization of mutual information. Int J Comput Vis **24** (1997) 137–154
7. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.W.: elastix: a toolbox for intensity-based medical image registration. IEEE Trans Med Imaging **29** (2010) 196–205
8. Ritter, F., Boskamp, T., Homeyer, A., Laue, H., Schwier, M., Link, F., Peitgen, H.O.: Medical image analysis: a visual approach. IEEE Pulse **2** (2011) 60–70