An Automatic Multi-atlas Segmentation of the Prostate in Transrectal Ultrasound Images Using Pairwise Atlas Shape Similarity

Saman Nouranian[1], S. Sara Mahdavi[1], Ingrid Spadinger[2], William J. Morris[2], Septimiu E. Salcudean[1], and Purang Abolmaesumi[1]

[1] Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, Canada
[2] Vancouver Cancer Center, British Columbia Cancer Agency, Vancouver, Canada

Abstract. Delineation of the prostate from transrectal ultrasound images is a necessary step in several computer-assisted clinical interventions, such as low dose rate brachytherapy. Current approaches to user segmentation require user intervention and therefore it is subject to user errors. It is desirable to have a fully automatic segmentation for improved segmentation consistency and speed. In this paper, we propose a multi-atlas fusion framework to automatically segment prostate transrectal ultrasound images. The framework initially registers a dataset of a priori segmented ultrasound images to a target image. Subsequently, it uses the pairwise similarity of registered prostate shapes, which is independent of the image-similarity metric optimized during the registration process, to prune the dataset prior to the fusion and consensus segmentation step. A leave-one-out cross-validation of the proposed framework on a dataset of 50 transrectal ultrasound volumes obtained from patients undergoing brachytherapy treatment shows that the proposed is clinically robust, accurate and reproducible.

1 Introduction

The majority of the image-guided diagnosis and treatment methods proposed for prostate cancer rely on the segmentation of the prostate gland in transrectal ultrasound (TRUS) images. The current standard-of-care for prostate segmentation in TRUS images is either manual or semi-automatic, with primary reason being the appearance of large amount of speckles and ultrasound imaging artifacts in ultrasound data. Consequently, the clinical prostate segmentation techniques are highly affected by inter- and intra-observer variability. Moreover, the segmentation process is a very time consuming task, especially for regions close to base and apex where the boundary of the anatomy is not visible.

Over the last three decades, there have been significant efforts made to reduce the observer variability and streamline the prostate segmentation process (see e.g. [1–4]). A group of proposed solutions rely on using information from previously segmented images, suitable for the application of atlases. An atlas is usually

K. Mori et al. (Eds.): MICCAI 2013, Part II, LNCS 8150, pp. 173–180, 2013.
© Springer-Verlag Berlin Heidelberg 2013

a model produced from a set of segmented control datasets that represents variation of shape and size of the object in a population. In segmentation problems, the atlas is generated *a priori* and is transformed into the coordinates of a target test data to propagate information to the unknown target space. Usually, multiple atlases that capture the variation of shape and size of the anatomy within the population are used to improve the success rate and increase robustness. Subsequently, the transformed atlases are fused through a voting mechanism [5–8] to produce the final segmentation result. The implicit assumption of the fusion step is that the transformed atlases are perfectly registered to the target and to each other. In other words, if one atlas is geometrically transformed to the target image coordinates, the same transformation on the atlas label produces the desired target segmentation.

Multi-atlas segmentation methods are widely used for segmentation of MR brain images [6, 9, 10], and more recently for CT and MR images of prostate [11–13]. To increase accuracy, robustness and/or speed, some of these works use decision fusion on a subset of atlases, discarding the information from poorly matching atlases [6, 10, 14]. One of the common approaches for pruning atlases is to rank them based on the registration quality metric. However, the optimized image similarity metric does not necessarily indicate better label matching among registered atlases. This is observed specially in the case of ultrasound images, which suffer from substantial imaging artifacts and contain speckles.

In this paper, we propose a multi-atlas segmentation framework that employs joint pairwise atlas shape similarity and classification in pruning poorly matching atlases prior to the fusion step. Our proposed method aims to reduce the impact of the initial registration error on the overall segmentation accuracy and precision. Initially, the atlases of prostate TRUS images and their segmentations are registered to a target image. Subsequently, the Dice similarity metric between the registered prostate shapes is used to cluster the atlases into groups, containing similar shapes. Next, an iterative label fusion technique is used to provide the consensus segmentation. We evaluate this framework on a set of 50 clinical TRUS images. We demonstrate that the proposed automatic segmentation framework produces results comparable with other semi-automatic segmentation techniques that are currently used in a clinical practice [3].

2 Method

The outline of the proposed method is shown on Figure 1. Hereafter, we refer to the target image and its hidden segmentation with V and L, respectively. Each atlas is denoted by $\{I_i, S_i\}$, $i \in \{1, 2, ..., N\}$ representing pair of the intensity and corresponding binary segmentation volume $S_i : \mathbb{R}^d \rightarrow \{0, 1\}$ from total number of N atlases in a d dimensional space.

Transformations obtained from intensity-based registration of each individual atlas volume to the target image is propagated onto the atlas labels, producing dataset of registered atlases denoted by $\{\widetilde{I}_i, \widetilde{S}_i\}$.

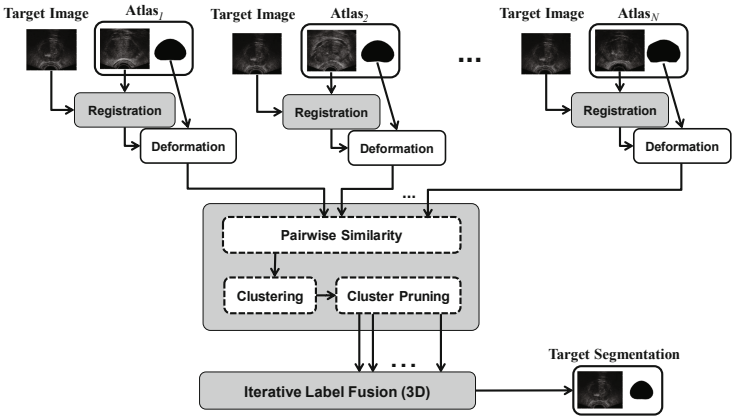

Fig. 1. Block diagram of the proposed segmentation algorithm

We define the pairwise atlas similarity vector $\mathbf{F}_i$ for the atlas $\{\widetilde{I}_i, \widetilde{S}_i\}$ by calculating its Dice similarity coefficient (DSC) $d_{\widetilde{S}_i,\widetilde{S}_j}, j \in \{1, ..., N\}$ with all registered atlas labels:

$$\mathbf{F}_i = (d_{i,1}, d_{i,2}, ..., d_{i,N}), d_{\widetilde{S}_i,\widetilde{S}_j} = \frac{2\left|\widetilde{S}_i \cap \widetilde{S}_j\right|}{\left|\widetilde{S}_i\right| + \left|\widetilde{S}_j\right|}, \tag{1}$$

Generated feature vectors are divided into K clusters. All members of each cluster are fused together and compared against the gold-standard. Since, the best cluster is unknown, we rank clusters based on the mean image similarity of its members to the target image. Top ranked clusters are merged together to produce the pruned dataset of atlases to fuse. An iterative voting method is used to produce the final estimation of the target segmentation $\hat{L}$. The detailed configuration of the proposed framework is explained below.

2.1 Registration

We use non-parametric diffeomorphic Demon's registration algorithm for bringing all atlases along with the coordinates of the target image [15]. This algorithm adapts the original Demon's underlying optimization procedure to a space of diffeomorphic transformations. It is shown that transformations obtained from this method are much smoother and closer to the gold-standard. The image similarity metric that we optimize is Sum of Squared Differences (SSD).

2.2 Atlas Clustering

By propagating the transformations obtained from the registration step, pairwise atlas shape similarity vectors $\mathbf{F}_i$ are calculated for each atlas. The Dice Similarity Coefficient (DSC) used for reconstruction of the $\mathbf{F}_i$ represents

atlas shape similarity by providing higher coefficients for pairs of the most similar atlas labels. Atlases are divided into K clusters $\mathbf{G} \in \{G_1, ..., G_K\}$ using the k-means clustering algorithm on the feature vectors $\mathbf{F}_i, i \in \{1, ..., N\}$. Using the cosine between pairwise similarity vectors $\mathbf{F}_i$, the k-means algorithm minimizes the within cluster distance between members $\mathbf{F}_i$ and the centroid of each cluster $\mathbf{C}_1, ..., \mathbf{C}_K$:

$$\arg\min_{\mathbf{G}} \sum_{i=1}^{K} \sum_{\mathbf{F}_j \in G_i} (1 - \frac{\mathbf{F}_j . \mathbf{C}_k}{\|\mathbf{F}_j\| \|\dot{\mathbf{C}}_k\|}). \tag{2}$$

Centroids of the clusters are randomly initiated prior to the optimization step. The initiation step is repeated for 50 times and the best clustering result with the minimum summation of the distance to the centroids is selected. Intuitively, the clustering approach on $\mathbf{F}_i$ groups similar shape atlases together, while distinguishing between variation of dissimilarity between them. k-means algorithm with the configuration above is performed 10 times in each experiment and combined to reduce the sensitivity of the atlas clustering to the random initialization.

2.3 Cluster Ranking

Clusters of atlases are ranked based on their average similarity to the target image. The mean SSD between members of each cluster and the target image is calculated and used as the ranking measure for clusters. Therefore, members of the top ranked clusters maintain maximum shape similarity with other members of the cluster while sharing a high image similarity to the target image. We assign a ranking parameter r_{G_i} to cluster G_i as:

$$r_{G_i} = \frac{1}{|G_i|} \sum_{\widetilde{I} \in G_i} \sum_{\mathbf{x} \in \widetilde{I}} \left(\widetilde{I}(\mathbf{x}) - V(\mathbf{x}) \right)^2, \tag{3}$$

where $\mathbf{x}$ is a voxel in the registered atlas image $\widetilde{I}$. By sorting clusters based on r_{G_i}, we select top $M, M \approx \frac{K}{2} + 1$ clusters and merge their members to form a pruned set of atlases P for the fusion step, described next.

2.4 Iterative Fusion

Simultaneous Truth and Performance Level Estimation (STAPLE) is a common method for the decision fusion [5]. The STAPLE approach uses Expectation Maximization to iterate between the estimation of the hidden ground-truth and the estimation of performance parameters for each of the raters. Performance or reliability of the raters are calculated based on the sensitivity and specifcity parameters of their segmentation.

Members of the pruned atlases P, are either fused using the STAPLE or further filtered in an iterative fusion algorithm [16]. We integrate the STAPLE in an iterative fusion algorithm outlined below:

1. An initial estimation of $\hat{L}$ (iteration template) is made using fusion of all labels. We use STAPLE to initiate the algorithm.
2. For each atlas $\{\widetilde{I}_i, \widetilde{S}_i\} \in P$, the Dice coefficient is calculated $d_{\widetilde{S}_i, \hat{L}}$ with the iteration template $\hat{L}_0$.
3. For all $\widetilde{S}_i \in P$, $\alpha_1\%$ of the atlases with lower Dice are discarded.
4. For all $\widetilde{I}_i \in P$, $\alpha_2\%$ of the atlases with higher SSD are discarded.
5. The remaining atlases are fused together to produce a new estimation of $\hat{L}$. We use STAPLE to reproduce the iteration template $\hat{L}$.
6. The algorithm is repeated from the step 2 until convergence, i.e. when the new iteration template is the same as the previous iteration.

3 Experiments

Transrectal ultrasound (TRUS) volumes of 50 patients undergoing brachytherapy treatment in our local cancer center are used for evaluation of the proposed segmentation algorithm. The dataset covers a range of prostate volume sizes between 20 cc and 70 cc. For each patient a series of 8 to 14 B-mode images have been captured from the prostate in the transverse plane. Images are 5 mm apart starting from the prostate base and ending to the apex. All TRUS volumes have been delineated plane by plane by an expert, producing the binary segmentation volumes we call the *gold-standard*.

Prior to the registration step, all TRUS volumes are aligned based on their anatomical axes, normalized using the histogram matching technique and downsampled by a factor of three in order to speed up the registration process.

Consensus segmentation is evaluated against the gold-standard in three ways: (1) A naive method of using top N atlases based on their global SSD with the target image. In these experiments, we select $N = 10$, since the performance of the approach does not change by increasing N beyond 10; (2) Using STAPLE on all members of each cluster to find the best cluster; (3) The proposed method which uses the pruned dataset by cluster ranking to perform iterative fusion. We call this the sub-optimal atlas selection approach. In this paper, α_1 and α_2 are set to 10 and 40; however, we found that the result did not change significantly with slight changes in these values.

We use leave-one-out cross validation for all TRUS volumes within the dataset. In each experiment, one patient volume is excluded from the dataset and is used as the target volume considering all remaining 49 items as the atlas dataset. Similarly to [3], we use DSC for shape similarity and the volume difference (V_{diff}) for size similarity as well as the mean and maximum absolute radial distance (MAD and MAXD) between mid-gland contours.

4 Results and Discussion

In order to measure dependency of the segmentation accuracy to the number of clusters, we report the leave-one-out cross validation results for different number

Table 1. Volume and contour similarity metrics calculated from leave-one-out cross validation of the proposed method using $K = 6$

	DSC	$\mathbf{V_{diff}}$	**MAD**	**MAXD**
Best cluster	92.08±1.96%	-2.83±6.67%	1.22±0.57 mm	3.47±1.48 mm
Sub-optimal selection	91.51±2.64%	-2.80±9.33%	1.39±0.72 mm	3.86±1.65 mm
Top 10	91.07±2.83%	-5.32±9.43%	1.43±0.70 mm	3.91±1.52 mm

of clusters in Figure 2(a). For different values of K, best cluster consistently results in a mean DSC value about 92%. However, the best results for the sub-optimal atlas selection approach is obtained with $K = 6$.

In Table 1, we list the volume and contour metrics for the case of $K = 6$. A paired t-test over the DSC values of the best cluster and the naive method of top 10, shows a significant improvement in accuracy ($p < 0.05$). According to the results shown in Figure 2(b), the best atlas selection and the sub-optimal atlas selection methods lead to DSC of 87% at 75% of cases, compared to DSC of 84% at 75% of cases for the top 10 approach.

In the context of brachytherapy [3, 17], the inter-observer variability of manually segmented contours has been reported to be in the order of $7.25 \pm 0.39\%$ and $6.64 \pm 2.36\%$ for volume error $(1 - DSC)$ and V_{diff}, respectively. Moreover, the intra-observer variability of manually segmented contours has been reported to be in the order of $5.95 \pm 1.59\%$. Our proposed method produces results comparable to the manual segmentation approach considering the variability above.

One should consider that the volume study and the brachytherapy procedure are often days apart. Besides, the segmentation of the prostate from the volume study, which is currently performed manually in many centres, takes three to four minutes [3]; while a non-optimized implementation of the proposed method takes less than three minutes on a standard PC (Intel Core i7, 2.93GHz, 8GB RAM). Moreover, since the individual registrations of atlases to the target volume are

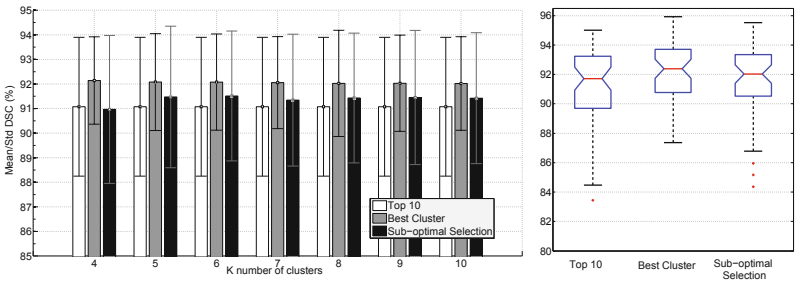

Fig. 2. (a) Mean and standard deviation of DSC from the leave-one-out cross validation vs number of K clusters. (b) Comparison of DSC distribution among ideal, sub-optimal and naive top 10 based leave-one-out result.

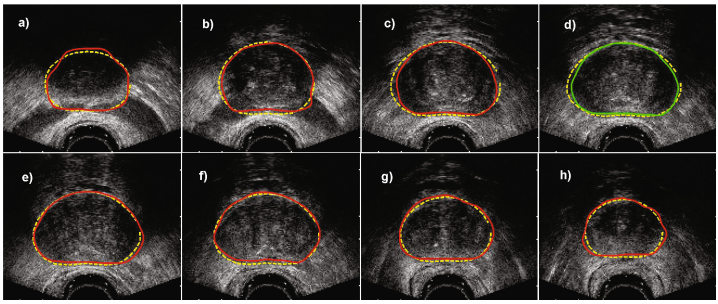

Fig. 3. Segmentation of a typical TRUS volume from post-base (a) to pre-apex (h). Dashed lines show the gold standard and solid lines are the algorithm output. Image (d) corresponds to the mid-gland image.

independent of each other, a parallel implementation on multiple CPU cores and possibly GPU should reduce the time further to the order of seconds. Therefore our method has the potential to handle the requirement of real-time dosimetry for brachytherapy with intra-operative planning.

In summary, results show improvement in DSC comparing to the naive top 10 based atlas selection, which means less modifications are required in a clinical application. Low standard deviation of the calculated metrics with respect to the mean, correlates with the robustness and repeatability of the proposed algorithm. It is also observed that this method is not very sensitive to the number of clusters. By propagating segmentation from atlases produced by different clinicians, this method can project a clinician-specific delineation style in the consensus segmentation. Future work will involve devising solutions to find the single cluster that generates the most accurate consensus segmentation. This framework can be integrated with more sophisticated fusion techniques or clustering approaches to produce more accurate segmentation results.

Acknowledgments. Funding from the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canadian Institutes of Health Research (CIHR), and the Institute for Computing, Information and Cognitive Systems (ICICS) at the University of British Columbia is gratefully acknowledged. The authors would also like to thank the physicians, therapists and staff at the Vancouver Cancer Center who have contributed in this project.

References

1. Ding, M., Chiu, B., Gyacskov, I., Yuan, X., Drangova, M., Downey, D.B., Fenster, A.: Fast prostate segmentation in 3D TRUS images based on continuity constraint using an autoregressive model. Medical Physics 34, 4109–4125 (2007)
2. Abolmaesumi, P., Sirouspour, M.R.: An interacting multiple model probabilistic data association filter for cavity boundary extraction from ultrasound images. IEEE Trans. Medical Imaging 23(6), 772–784 (2004)

3. Mahdavi, S.S., Chng, N., Spadinger, I., Morris, W.J., Salcudean, S.E.: Semi-automatic segmentation for prostate interventions. Medical Image Analysis 15(2), 226–237 (2011)
4. Qiu, W., Yuan, J., Ukwatta, E., Tessier, D., Fenster, A.: Rotational-slice-based prostate segmentation using level set with shape constraint for 3D end-firing TRUS guided biopsy. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 537–544. Springer, Heidelberg (2012)
5. 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 Trans. Medical Imaging 23(7), 903–921 (2004)
6. Artaechevarria, X., Munoz-Barrutia, A., Ortiz-de Solorzano, C.: Combination strategies in multi-atlas image segmentation: application to brain MR data. IEEE Trans. Medical Imaging 28(8), 1266–1277 (2009)
7. Isgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M.A., Van Ginneken, B.: Multi-atlas-based segmentation with local decision fusion–application to cardiac and aortic segmentation in CT scans. IEEE Trans. Medical Imaging 28(7), 1000–1010 (2009)
8. Sabuncu, M.R., Thomas Yeo, B.T., Van Leemput, K., Fischl, B., Golland, P.: A generative model for image segmentation based on label fusion. IEEE Trans. Medical Imaging 29(10), 1714–1729 (2010)
9. 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(1), 115–126 (2006)
10. 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(3), 726–738 (2009)
11. Acosta, O., Simon, A., Monge, F., Commandeur, F., Bassirou, C., Cazoulat, G., de Crevoisier, R., Haigron, P.: Evaluation of multi-atlas-based segmentation of CT scans in prostate cancer radiotherapy. In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 1966–1969. IEEE (2011)
12. 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, 1407 (2008)
13. Dowling, J.A., et al.: Fast automatic multi-atlas segmentation of the prostate from 3D MR images. In: Madabhushi, A., Dowling, J., Huisman, H., Barratt, D. (eds.) Prostate Cancer Imaging 2011. LNCS, vol. 6963, pp. 10–21. Springer, Heidelberg (2011)
14. Wang, H., Suh, J.W., Das, S.R., Pluta, J., Craige, C., Yushkevich, P.A.: Multi-atlas segmentation with joint label fusion. IEEE Trans. Pattern Analysis and Machine Intelligence 35(3), 611–623 (2012)
15. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: efficient non-parametric image registration. NeuroImage 45(1, suppl.), S61–S72 (2009)
16. Ou, Y., Doshi, J., Erus, G., Davatzikos, C.: Multi-atlas segmentation of the prostate: A zooming process with robust registration and atlas selection. In: Medical Image Computing and Computer-Assisted Intervention: MICCAI Workshop (2012)
17. Mahdavi, S.S., Spadinger, I., Chng, N., Salcudean, S.E., Morris, W.J.: Semiautomatic segmentation for prostate brachytherapy: Dosimetric evaluation. Brachytherapy (2011)