

Learning Distance Transform for Boundary Detection and Deformable Segmentation in CT Prostate Images

Yaozong Gao^{1,2}, Li Wang¹, Yeqin Shao^{1,3}, and Dinggang Shen¹

¹ Department of Radiology and BRIC, University of North Carolina at Chapel Hill, USA

² Department of Computer Science, University of North Carolina at Chapel Hill, USA

³ Institution of Image Processing & Pattern Recognition, Shanghai Jiao Tong University, China

Abstract. Segmenting the prostate from CT images is a critical step in the radiotherapy planning for prostate cancer. The segmentation accuracy could largely affect the efficacy of radiation treatment. However, due to the touching boundaries with the bladder and the rectum, the prostate boundary is often ambiguous and hard to recognize, which leads to inconsistent manual delineations across different clinicians. In this paper, we propose a learning-based approach for boundary detection and deformable segmentation of the prostate. Our proposed method aims to learn a boundary distance transform, which maps an intensity image into a boundary distance map. To enforce the spatial consistency on the learned distance transform, we combine our approach with the auto-context model for iteratively refining the estimated distance map. After the refinement, the prostate boundaries can be readily detected by finding the valley in the distance map. In addition, the estimated distance map can also be used as a new external force for guiding the deformable segmentation. Specifically, to automatically segment the prostate, we integrate the estimated boundary distance map into a level set formulation. Experimental results on 73 CT planning images show that the proposed distance transform is more effective than the traditional classification-based method for driving the deformable segmentation. Also, our method can achieve more consistent segmentations than human raters, and more accurate results than the existing methods under comparison.

1 Introduction

CT images are widely used in the image-guided radiotherapy planning (IGRT), as it provides Hounsfield units for all image voxels, which are necessary for dose calculation. In the IGRT for prostate cancer, the prostate and nearby organs (e.g., the bladder and the rectum) need to be segmented in order to optimize the dose plan for precisely targeting the radiation beams on the prostate and minimizing the radiation exposure to the surrounding tissues. The segmentation accuracy of the prostate could largely affect the efficacy of radiation treatment. However, due to the touching boundaries with the bladder and the rectum, the prostate boundary is often indistinct in CT images (Fig. 1), which imposes much difficulty upon the manual delineation. It usually takes an experienced clinician 10 – 12 minutes to manually delineate the prostate boundary in CT image of each patient. Despite of taking this long delineation time, manual segmentations still vary much among different raters [1,2]. Thus, an automatic and robust segmentation method is highly desired in this context.

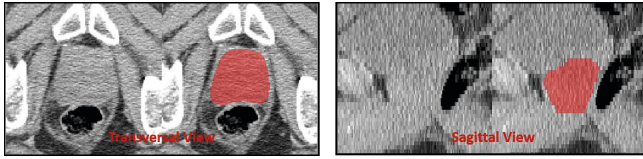


Fig. 1. A typical 3D planning CT image in the transversal view (left panel) and sagittal view (right panel). In each view, the left figure shows the original prostate slice, and the right figure shows the corresponding slice overlaid with the manually segmented prostate (red).

Previous works [3,4] often reply on learning a patient-specific model for addressing the aforementioned challenge, as the prostate appearance and shape variations are small for the image data acquired from the same patient. However, these methods are not applicable for prostate segmentation from the planning CT images, as no segmented CT images of the same patient are available for appearance and shape learning in the radiotherapy planning stage. Consequently, only population information (i.e., CT images of other patients) could be used for guiding the prostate segmentation in the planning CT images. Among the population-based methods, most of them utilize the shape constraint for deriving a robust segmentation. For example, Costa et al. [5] proposed a coupled deformable model for segmenting the prostate by imposing a non-overlapping constraint from the bladder. Chen et al. [6] proposed a Bayesian framework that incorporates anatomical constraints from the surrounding bones for prostate segmentation. However, due to the lack of an effective appearance model for guiding the deformable segmentation, the accuracy of these methods is very limited. Recently, classification-based methods [2,7] have been proposed to segment the prostate from CT images, and achieved significant improvement over the traditional intensity-based methods [5,6,8]. The main idea is to train a classifier for distinguishing prostate voxels from background voxels based on local patch appearance. The learned classifier could be used to label the intensity image into a prostate likelihood map for guiding the deformable segmentation.

In this paper, we propose to learn a distance transform for boundary detection and deformable segmentation of the prostate. The learned distance transform can map a new intensity image into the distance map of the target prostate boundary, which could be further utilized for anatomical boundary detection as well as deformable segmentation. In particular, regression forest is adopted to learn the non-linear relationship between a voxel's local image appearance and its 3D displacement to the nearest point on the target prostate boundary. Once the forest is learned, it can be used to predict the 3D displacement from any voxel in the new testing image to the target prostate boundary. By taking the magnitude of the displacement vector, the distance map of the prostate boundary can be obtained for a new testing image. To enforce the spatial consistency within the obtained distance map, we further combine the high-level context features extracted from the previously obtained distance map with the original image appearance features into the auto-context framework [9] for iterative refinement. Finally, the refined distance map will be integrated into a level set formulation for segmenting the prostate from CT images. Experimental results show that learning a boundary distance transform is more effective than prostate classification for guiding the deformable segmentation.

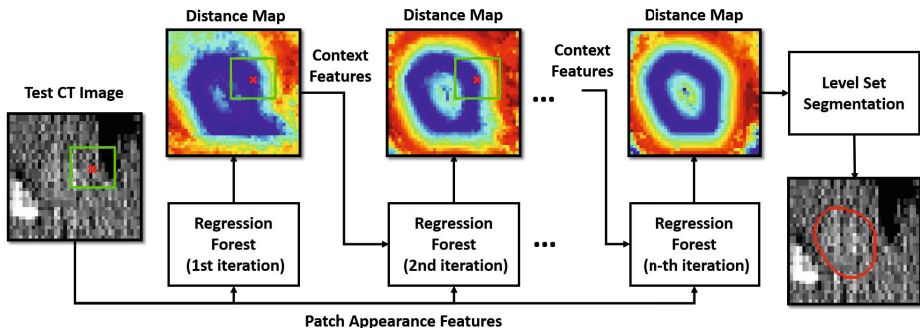


Fig. 2. The flowchart of our method. Green boxes show the local patches where appearance and context features are extracted for the voxel marked as red crosses. Cold and warm colors in the figure indicate voxels with small and large predicted distances to the prostate boundary.

In addition, our method can achieve more consistent segmentations than human raters, and also more accurate results than existing methods under comparison.

2 Method

Our method consists of three components: 1) regression forest for learning boundary distance transform, 2) iterative refinement of the predicted distance map by context features, and 3) distance-map-guided boundary detection and deformable segmentation with level sets. Fig. 2 shows the flowchart of our method.

2.1 Learning Boundary Distance Transform by Regression Forest

Regression forest, as a non-linear regression model, has recently been used for efficient anatomy detection [10], i.e., detecting the bounding box of one specific organ. In this paper, we extend it to learn the distance transform for a specific organ boundary (e.g., prostate boundary). The learned distance transform is used for mapping a new 3D intensity image into the distance map of the target boundary. More specifically, given any voxel in the new testing image, we want to predict its nearest distance to the target boundary. Hence, distance transform learning is essentially a regression problem. In our work, regression forest is particularly used for learning the non-linear relationship between a voxel’s local image appearance and its 3D displacement vector to the nearest point on the target boundary. By taking the magnitude of the 3D displacement vector, the distance of this voxel to the target boundary can be obtained. Thus, the learned regression forest can be regarded as a boundary distance transform. In the next paragraphs, we will show how the regression forest is trained for learning boundary distance transform, and how the learned forest could be applied to a new testing image for predicting the distance map.

To learn the distance transform for a specific organ boundary, we first randomly sample voxels near the boundary in every training image according to a Gaussian distribution: $p(\mathbf{x}) = 1/(\sqrt{2\pi}\sigma) \times \exp(-d(\mathbf{x})^2/2\sigma^2)$, where $p(\mathbf{x})$ indicates the probability of

voxel $\mathbf{x} \in \mathbb{R}^3$ in a training image to be sampled, $d(\mathbf{x})$ is the nearest distance of voxel \mathbf{x} to the target boundary in this training image, and σ controls the size of narrowband for sampling. In this way, the majority of sampled voxels will be close to the target boundary, thus making the learned model more specific on detecting the target boundary. This sampling strategy is important for accurate organ segmentation, as boundary voxels are usually the most difficult to characterize. Afterwards, the sampled voxels from all training images are used as our training dataset. For each sampled voxel in one training image, we extract randomized 3D Haar-like features from an intensity patch centered at this voxel for capturing the local image appearance around it. The Haar-like features are defined as follows.

$$f(I) = \sum_{i=1}^M t_i \sum_{\|\mathbf{x}-\mathbf{c}_i\|_\infty \leq s_i} I(\mathbf{x}) \quad (1)$$

where $f(I)$ denotes one 3D Haar-like feature extracted from intensity patch I , M is the number of 3D cubic functions used in this Haar-like feature, and $t_i \in \{+1, -1\}$, \mathbf{c}_i and s_i are the polarity, the center and the size of the i -th cubic function, respectively. By randomizing the parameters M , t_i , \mathbf{c}_i and s_i in Eq. 1, we can generate an unlimited number of 3D Haar-like features for regression forest learning. In this work, M is limited to $\{1, 2\}$, s_i is limited to $\{3, 5\}$, and \mathbf{c}_i is not limited as long as the 3D cubic function stays within intensity patch I of size $30 \times 30 \times 30$.

Once the feature representation of each voxel is determined, a regression forest can be trained for predicting the 3D displacement from any image voxel to the nearest point on the target boundary. Given a new testing image, the learned forest can be applied to voxel-wisely estimate the 3D displacement for every image voxel. By taking the magnitude, a boundary distance map can then be obtained.

2.2 Iterative Refinement of Distance Map by Context Features

As the displacement from each image voxel to the target boundary is predicted independently, the estimated distance map for a new testing image is often spatially inconsistent, as shown in the leftmost distance map of Fig. 2. To overcome this limitation, we integrate the proposed distance transform learning with the auto-context model [9] for iteratively refining the estimated distance map. The main idea is to train a sequence of distance transforms, each utilizing both the local image features extracted from the original intensity image, and the high-level context features extracted from the output of the previous distance transform for gradually improving the quality of the estimated distance map.

During the training stage, after the distance transform of the first iteration is learned as described in Section 2.1, it can be used to predict a boundary distance map for every training image. Then, the additional high-level context features can be extracted from the estimated distance map, and further combined with the original image features to form a new feature representation for each voxel. Afterwards, a new distance transform can be learned by using the updated feature representation. This iterative training procedure continues until a specified number N of distance transforms is obtained. In our work, the high-level context features are also the randomized Haar-like features as

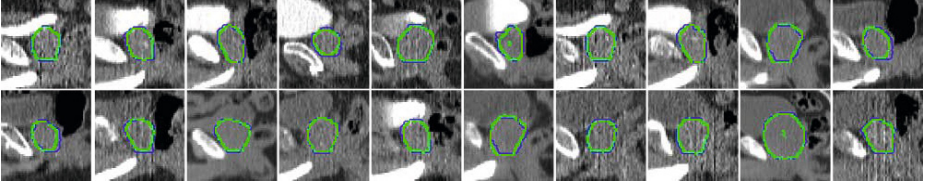


Fig. 3. Typical planning CTs (sagittal view) in our dataset. Green and blue contours indicate the prostate boundaries automatically detected by our boundary detection method and manually delineated by the expert, respectively.

defined in Eq. 1. Different from image appearance features, these context features are extracted from the distance map estimated by the previous distance transform. Since the rough distances of nearby voxels to the target boundary have been encoded in the previously estimated distance map, the new distance transform learning can utilize this valuable information to impose the spatial consistency on the to-be-estimated distance map, thus improving the overall prediction accuracy. In the testing stage, the learned distance transforms can be applied sequentially as shown in Fig. 2 to iteratively refine the estimated distance map for a new testing image.

2.3 Distance-Map-Guided Boundary Detection and Deformable Segmentation (with Level Sets)

Once the boundary distance map is estimated for a new testing image, it can be used for either boundary detection or level set segmentation.

Boundary Detection: In most cases, the estimated distance map of a new testing image will be directly utilized for the final segmentation (e.g., to guide the deformable segmentation). However, sometimes if the organ-specific boundary segments are desired, we can also adopt non-minima suppression and hysteresis thresholding, similar as in the canny edge detector [11], to detect these organ-specific boundaries from the estimated distance map.

Level Set Segmentation: Since the target boundaries are located in the valley of the estimated distance map, the local means in the estimated distance map should be similar for both sides of the zero level set. Based on this assumption, we can design the following evolution flow to segment the prostate from the boundary distance map:

$$\frac{\partial \phi}{\partial t} = \delta(\phi)(u_1(\mathbf{x}) - u_2(\mathbf{x})) + v\delta(\phi)\text{div}\left(\frac{\nabla \phi}{\|\nabla \phi\|}\right) \quad (2)$$

$$u_1(\mathbf{x}) = \frac{\int K(\mathbf{y} - \mathbf{x})H(\phi(\mathbf{y}))\Omega(\mathbf{y})d\mathbf{y}}{\int K(\mathbf{y} - \mathbf{x})H(\phi(\mathbf{y}))d\mathbf{y}}, u_2(\mathbf{x}) = \frac{\int K(\mathbf{y} - \mathbf{x})(1 - H(\phi(\mathbf{y})))\Omega(\mathbf{y})d\mathbf{y}}{\int K(\mathbf{y} - \mathbf{x})(1 - H(\phi(\mathbf{y})))d\mathbf{y}} \quad (3)$$

where ϕ is the level set function with $\phi > 0$ as the inner part and $\phi < 0$ as the outer part, δ is the Delta function, $u_1(\mathbf{x})$ and $u_2(\mathbf{x})$ are the local means of the inner and outer parts, respectively, K is a Gaussian kernel function with the standard deviation of 3, H is the Heaviside step function [12], and Ω denotes the estimated boundary distance map. The first data-fitting term attracts the zero level set to the valley of the estimated distance map Ω , and the second regularization term imposes the smoothness constraint on the evolving surface ϕ .

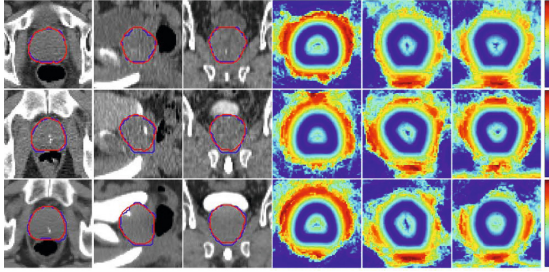


Fig. 4. Qualitative results from three planning CT images with different levels of contrast agent. Each row shows the planning CT images and their corresponding predicted distance maps in transversal, sagittal and coronal views. Red and blue contours indicate our final segmented prostate boundaries and the manually delineated boundaries, respectively.

3 Experiments

Data Descriptions: Our dataset consists of 73 planning CT images, scanned from different patients. The typical image size is $512 \times 512 \times (61 \sim 81)$ with voxel size $0.94 \times 0.94 \times 3.00 \text{ mm}^3$. The prostate in each planning CT image has been manually delineated by a radiation oncologist, which we use as ground truth. The dataset is of large appearance variability due to the uncertainty on the level of contrast agent that is present. Fig. 3 shows typical planning CT images (sagittal view) in our dataset along with the detected prostate boundaries by our boundary detection method (green) and manual rater (blue).

Parameter Setting: In the regression forest training, the number of trees is 10, the maximum tree depth is 15, the number of randomized Haar-like features is 1000 for both image appearance and context features, and the minimum sample number for each leaf node is 8. σ for controlling the size of narrowband sampling is 8. v in the level set segmentation is set to 0.01. All the parameters of regression forest is typical as adopted in other works [10]. To evaluate our segmentation method, we use four-fold cross-validation with 54 images for training and 19 images for testing. The initialization of the level set function is accomplished by using an affine transformation to transform the mean prostate shape onto the testing image. The affine transformation is estimated between six automatically detected prostate landmarks (i.e., top, base, anterior, posterior, left and right) in the testing image and their counterparts on the mean shape [2].

Qualitative Results: In addition to the boundary detection results in Fig. 3, we also plot the qualitative results for three typical planning CT images (with different levels of contrast agent) in Fig. 4. We can see that, after the proposed distance transform, prostate boundaries can be clearly seen in the predicted distance maps, and are quite consistent with the manually delineated boundaries by radiation oncologist. This demonstrates that our proposed method works very well in various planning CTs with different levels of contrast agent.

Classification versus Distance Transform: Fig. 5 (a) quantitatively compares the classification guided level set method with our proposed method (distance-map-guided level set method) on the same dataset. As aforementioned, the classification-based method

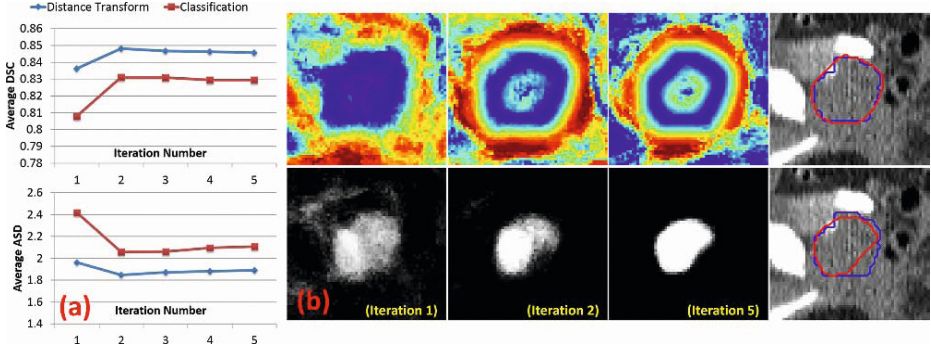


Fig. 5. (a) Quantitative comparison between classification-guided and distance-map-guided level set methods in our dataset. DSC: Dice Similarity Coefficient. ASD: Average Surface Distance. (b) Qualitative comparison between distance-map-based (first row) and classification-based (second row) auto-context refinement on a typical planning CT image. Red and blue contours indicate automatically-segmented and manually-delineated prostates, respectively.

uses a learned classifier to label the new testing image into a prostate likelihood map, which is then utilized for guiding the deformable segmentation with level sets [12]. For fair comparison, we used the classification forest with the same training parameters for the classification-guided level set method. Similarly, we also adopt the auto-context model to iteratively refine the classification response map. From Fig. 5 (a), we can clearly see that our proposed method (“distance transform”) outperforms the classification-guided level set method in all iterations (e.g., with higher Dice Similarity Coefficient (DSC) and lower Average Surface Distance (ASD)). In addition, Fig. 5 (b) gives a typical example that compares the classification-based auto-context refinement with our distance-map-based auto-context refinement. We can see that the distance-map-based refinement is able to achieve more accurate segmentation than classification-based refinement. This infers that the context features extracted from the boundary distance map are more helpful to assist the auto-context refinement than the traditional context features extracted from the classification response map.

Comparison with other CT Prostate Segmentation Methods: Our method obtains an average surface distance (ASD) 1.85 ± 0.87 mm on our dataset. Due to the fact that neither the executables nor the datasets of other works are publicly available, it is difficult for us to directly compare our method with other CT prostate segmentation methods. Thus, we only cite the results reported in their publications for reference. The comparison shows that our method achieves more accurate segmentations than [8] (ASD 4.09 ± 0.90 mm), [2] (ASD 3.35 ± 1.40 mm), and the current state-of-the-art method [7] (ASD 2.37 ± 0.89 mm). Besides, it is worth noting that most existing methods were evaluated only on the datasets without contrast agent. It is not clear whether these methods can be applied to the mixed datasets which contain planning CTs with different levels of contrast agent. Actually, this aspect is very important in the clinical application. A desired prostate segmentation method should be able to deal with various kinds of planning CTs obtained with different contrasts and scanning protocols, as it is never pre-known

which type of an unseen image would need to be segmented. Clearly, our method wins at this point, since it has been evaluated with good performance on the dataset with planning CTs of different contrasts. Additionally, the comparison with inter-rater variability of manual prostate delineations ($ASD\ 3.03 \pm 1.15\text{mm}$ [2]) indicates that our method is also able to obtain more consistent segmentations than the human raters.

4 Conclusion

In this paper, we propose to predict the boundary distance transform for anatomical boundary detection and deformable segmentation. It is applied to segment the prostate from CT images. Validated on 73 planning CT images with various contrasts, our proposed distance transform learning method shows better performance than prostate classification method for guiding the deformable segmentation of the prostate. Moreover, the comparisons with other CT prostate segmentation methods indicate that our method can be more adaptive to different datasets with various contrasts. Also, compared to manual prostate delineations, our method can achieve more consistent segmentations, since there often exists large inter-rater variability for manual delineations.

References

1. Foskey, M., Davis, B., et al.: Large deformation three-dimensional image registration in image-guided radiation therapy. *Phy. Med. Biol.* 50(24), 5869 (2005)
2. Lay, N., Birkbeck, N., Zhang, J., Zhou, S.K.: Rapid multi-organ segmentation using context integration and discriminative models. In: Gee, J.C., Joshi, S., Pohl, K.M., Wells, W.M., Zöllei, L. (eds.) *IPMI 2013. LNCS*, vol. 7917, pp. 450–462. Springer, Heidelberg (2013)
3. Feng, Q., Foskey, M., Tang, S., Chen, W., Shen, D.: Segmenting CT prostate images using population and patient-specific statistics for radiotherapy. *Med. Phys.* 37(8), 4121–4132 (2010)
4. Gao, Y., Liao, S., Shen, D.: Prostate segmentation by sparse representation based classification. *Med. Phys.* 39(10), 6372–6387 (2012)
5. Costa, M.J., Delingette, H., Novellas, S., Ayache, N.: Automatic segmentation of bladder and prostate using coupled 3D deformable models. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part I. LNCS*, vol. 4791, pp. 252–260. Springer, Heidelberg (2007)
6. Chen, S., Lovelock, D.M., Radke, R.J.: Segmenting the prostate and rectum in CT imagery using anatomical constraints. *Med. Ima. Anal.* 15(1), 1–11 (2011)
7. Lu, C., et al.: Precise segmentation of multiple organs in CT volumes using learning-based approach and information theory. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part II. LNCS*, vol. 7511, pp. 462–469. Springer, Heidelberg (2012)
8. Rousson, M., Khamene, A., Diallo, M., Celi, J.C., Sauer, F.: Constrained surface evolutions for prostate and bladder segmentation in CT images. In: Liu, Y., Jiang, T.-Z., Zhang, C. (eds.) *CVBIA 2005. LNCS*, vol. 3765, pp. 251–260. Springer, Heidelberg (2005)
9. Tu, Z., Bai, X.: Auto-context and its application to high-level vision tasks and 3D brain image segmentation. *PAMI* 32(10), 1744–1757 (2010)
10. Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E.: Regression forests for efficient anatomy detection and localization in CT studies. In: Menze, B., Langs, G., Tu, Z., Criminisi, A. (eds.) *MICCAI 2010. LNCS*, vol. 6533, pp. 106–117. Springer, Heidelberg (2011)
11. Canny, J.: A computational approach to edge detection. *PAMI* 8(6), 679–698 (1986)
12. Chan, T., Vese, L.: Active contours without edges. *TIP* 10, 266–277 (2001)