



Published in final edited form as:

Med Phys. 2018 June ; 45(6): 2527–2541. doi:10.1002/mp.12898.

A semiautomatic segmentation method for prostate in CT images using local texture classification and statistical shape modeling

Maysam Shahedi,

Department of Radiology and Imaging Sciences, Emory University School of Medicine, Atlanta, GA 30322, USA

Martin Halicek,

Department of Biomedical Engineering, Emory University and Georgia Institute of Technology, Atlanta, GA 30332, USA

Rongrong Guo,

Department of Radiology and Imaging Sciences, Emory University School of Medicine, Atlanta, GA 30322, USA

Guoyi Zhang,

Department of Radiology and Imaging Sciences, Emory University School of Medicine, Atlanta, GA 30322, USA

David S. Schuster, and

Department of Radiology and Imaging Sciences, Emory University School of Medicine, Atlanta, GA 30322, USA

Baowei Fei*

Department of Radiology and Imaging Sciences, Emory University School of Medicine, Atlanta, GA 30322, USA, Department of Biomedical Engineering, Emory University and Georgia Institute of Technology, Atlanta, GA 30332, USA, Winship Cancer Institute of Emory University, Atlanta, GA 30322, USA, Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA

Abstract

Purpose: Prostate segmentation in computed tomography (CT) images is useful for treatment planning and procedure guidance such as external beam radiotherapy and brachytherapy. However, because of the low, soft tissue contrast of CT images, manual segmentation of the prostate is a time-consuming task with high interobserver variation. In this study, we proposed a semiautomated, three-dimensional (3D) segmentation for prostate CT images using shape and texture analysis and we evaluated the method against manual reference segmentations.

* **Corresponding Author:** Baowei Fei, Ph.D., Department of Radiology and Imaging Sciences, Emory University School of Medicine, 1841 Clifton Road NE, Atlanta, GA 30329, US, Telephone: 404-712-5649, Fax: 404-712-5689, bfei@emory.edu, Website: www.fei-lab.org.

The authors have no conflicts to disclose.

Each author has no conflict of interest or financial disclosure.

Methods: The prostate gland usually has a globular shape with a smoothly curved surface, and its shape could be accurately modeled or reconstructed having a limited number of well-distributed surface points. In a training data set, using the prostate gland centroid point as the origin of a coordination system, we defined an inter-subject correspondence between the prostate surface points based on the spherical coordinates. We applied this correspondence to generate a point distribution model for prostate shape using principal component analysis and to study the local texture difference between prostate and non-prostate tissue close to the different prostate surface sub-regions. We used the learned shape and texture characteristics of the prostate in CT images and then combined them with user inputs to segment a new image. We trained our segmentation algorithm using 23 CT images and tested the algorithm on two sets of 10 non-brachytherapy and 37 post-low dose rate brachytherapy CT images. We used a set of error metrics to evaluate the segmentation results using two experts' manual reference segmentations.

Results: For both non-brachytherapy and post-brachytherapy image sets, the average measured Dice similarity coefficient (DSC) was 88% and the average mean absolute distance (MAD) was 1.9 mm. The average measured differences between the two experts on both datasets were 92% (DSC) and 1.1 mm (MAD).

Conclusions: The proposed, semiautomatic segmentation algorithm showed a fast, robust, and accurate performance for 3D prostate segmentation of CT images, specifically when no previous, intra-patient information, i.e. previously segmented images, was available. The accuracy of the algorithm is comparable to the best performance results reported in the literature and approaches the inter-expert variability observed in manual segmentation.

Keywords

Computer tomography (CT); prostate; segmentation; texture features

1. Introduction

Prostate cancer (PCa) is the leading cancer diagnosed among males in the United States¹. It accounted for more than 26,000 cancer deaths in 2016¹. Currently, image-guided radiation therapy is one of the primary treatment methods for patients with localized PCa^{2, 3}. PCa radiation therapy planning is performed with the prostate border delineated on computed tomography (CT) images⁴. However, due to the low soft-tissue contrast between the prostate and surrounding tissues, manual contouring of the prostate in CT images is time-consuming⁵ and is subject to high intraobserver and interobserver variability⁶⁻⁹. Therefore, computer-assisted segmentation algorithms are being investigated and developed to perform prostate contouring more quickly and more reproducibly compared to manual segmentation.

Information regarding several automatic and semiautomatic, computer-assisted methods have been published in the literature regarding three-dimensional (3D) segmentation of the prostate on CT images. The majority of these studies were regarding learning-based segmentation techniques. Feng *et al.*¹⁰ proposed an automatic algorithm for CT prostate segmentation, based on shape and appearance modeling using population- and patient-specific statistics. Their method is more useful for radiotherapy treatment CT images when a series of previously segmented treatment images from the same patient is available. Liao *et*

*al.*¹¹ presented a multi-atlas sparse label propagation framework to estimate the prostate likelihood of image voxels using patch-based representation in a discriminative feature space for semiautomatic prostate segmentation in treatment CT images. In their method, manual segmentation labels of the planning and the first two treatment CT images were needed. After each treatment image segmentation, they used an online update mechanism with potential offline manual adjustment of the segmentation label to add the newly segmented image to the training set. Shi *et al.*^{12, 13} presented a semiautomatic prostate segmentation method that first estimated the prostate-likelihood map slice-by-slice, and then merged the 2D maps to generate a 3D prostate-likelihood map. Finally, the segmentation labels of the planning and the previous treatment images of the same test patient were rigidly aligned to the prostate-likelihood map and a majority vote was used to prepare the final segmentation label. Shi *et al.*¹⁴ also presented another segmentation algorithm in which they improved the segmentation performance by incorporating manually assigned prostate and non-prostate labels for a subset of test image voxels for training. Shao *et al.*¹⁵ presented a boundary voting technique for automatic prostate segmentation in CT images using a global regression forest for prostate boundary detection followed by a deformable prostate segmentation. They also presented a similar automatic boundary detection algorithm for segmentation of the prostate and rectum in radiotherapy planning CT images¹⁶. They estimated the prostate and rectum borders by automatic landmark detection using regression forest followed by shape modeling. Ma *et al.*¹⁷ presented a semiautomatic, learning-based segmentation method that was trained based on a combination of population and patient-specific data. They used manual segmentation of three, 2D slices of the target image as the patient-specific data. Ma *et al.* also presented an automatic, deep learning-based prostate CT segmentation algorithm¹⁸. They used the convolutional neural networks to determine the deep features of the discriminate prostate and non-prostate voxels. They refined the neural network output by applying a multi-atlas label fusion. Table V compares the segmentation accuracy of the previously presented algorithms mentioned above.

In the previously presented learning-based methods a training dataset consisting of manually segmented images was used to study a set of different characteristics of the prostate in the image. Then, during segmentation, the learned information was used to delineate the prostate border on an unseen test image set. In some of these methods the training sets consisted of CT images from other patients only, and in some others the previously segmented treatment CT images from the same target patient were also included in the training set and that helps to have a more accurate computer-assisted segmentation. However, it challenges the segmentation performance where no previously acquired CT images are available, e.g. for radiotherapy planning CT image segmentation. Moreover, manual segmentation of one or more CT images from the same patient should be available as a prerequisite for segmentation of the next treatment image. The performance of all the mentioned algorithms were evaluated against one set of manual reference segmentation. However, due to high interobserver variability in manual segmentation of the prostate in CT images, the measured error metric values are dependent on the selected manual reference segmentation label and changing the reference could change the evaluation results. Therefore, it is useful for deeper understanding of the algorithm performance to take interobserver variability into account for evaluation.

In this manuscript, we present a semiautomatic algorithm for 3D prostate segmentation on CT images, based on learned prostate shape variability and local image texture near the prostate border. The proposed learning-based method does not need to be trained on previously acquired and/or manually segmented images from the target patient. We use the prostate gland centroid point as the origin of the coordination system and define an inter-patient correspondence between prostate surface points based on the spherical coordinates of the points. During training, we apply this correspondence to generate two point distribution models (PDM) for prostate shape using principal component analysis, i.e. a low-density (LD) model with 86 surface points and a high-density (HD) model with 2056 surface points. To study the local image texture close to the prostate border, we train our algorithm on a number of rays emanating from the centroid, separately. For all the image voxels on a ray, inside and outside the prostate gland, we measure a set of local texture features using a cubic image patch centered at the voxel. For all the rays across the training set in the corresponding direction, we train a classifier using the local texture features to identify prostate versus non-prostate voxels. The segmentation algorithm is initiated with a bounding box for the gland and a number of manually selected landmark points on the prostate surface. The algorithm searches for a set of prostate border points in the 3D space by classifying the points close to the prostate border on a number of rays casted from the centroid of the gland. Then 3D point distribution models are used to regularize the segmentation and to reconstruct the 3D surface from the border points. We evaluate the segmentation algorithm against two, expert reference segmentations, and to capture different types of error, we use a set of error metrics to measure surface distances, regional overlap errors, and volume differences. We also evaluate the algorithm performance on post-low dose rate (LDR) brachytherapy CT images. The main contributions of this work include: (i) A new semi-automatic prostate segmentation algorithm was proposed and implemented for CT images, which does not use the previous images of the same patient, *i.e.*, it is independent to the planning and/or previous treatment images of the same patient. (ii) A local texture classification approach was developed to estimate the prostate border and to avoid negative effects of image texture distortion caused by necrotic tissue, tumors, or LDR brachytherapy seeds. (iii) A new shape modeling was developed for the prostate, which is based on a smooth globular shape of the gland. (iv) A comprehensive validation approach was used to evaluate the performance of the segmentation technique by capturing different performance aspects of importance to the potentially intended application of the algorithm.

2. Materials and Methods

2.A. Materials

Our CT image dataset contained 70, 3D abdominal CT scans from 70 patients. Thirty-seven of the images were acquired from post-LDR brachytherapy patients. The size of each image was $512 \times 512 \times 27$ voxels with the voxel sizes of $0.977 \times 0.977 \times 4.25$ mm. For each image two segmentation labels manually drawn by two experienced radiologists were available.

2.B. Semiautomatic Segmentation

The proposed semiautomatic segmentation algorithm consists of two main parts, *i.e.* training and segmentation. Figure 1 shows the schematic block diagram of the algorithm. The

training and segmentation components are described in detail in sections 2.B.1 and 2.B.2, respectively.

2.B.1. Training—During training, we use the training image set to extract a set of local texture features within each sector of the spherical space. The most discriminative features are then selected to train a classifier. We also use the manual segmentation labels of the training images to build an LD and an HD PDM for prostate shape. These shape models are used for shape regularization and reconstruction during segmentation. Each of the training blocks illustrated in Figure 1 is described in detail below.

2.B.1.a. Preprocessing:

Anteroposterior symmetry axis alignment: We rotated all of the training images and their manual reference segmentation labels about their inferior-superior axis in order to align the anteroposterior symmetry axes of the patients parallel to the anteroposterior axes of the images. To automatically measure a patient's anteroposterior symmetry axis angle in an image, we first roughly segmented the bones in the 3D image using a thresholding segmentation method (threshold level = 155) and kept the two largest segmented objects associated with the left and right hip bones and removed all of the other smaller objects (Figure 2a and b). We then measured the centroid of the bones to roughly estimate the prostate location (Figure 2b) and cropped the images along the x- and y-axes about the centroid so as to limit the image field of view. We selected the field of view large enough (121 pixels \times 201 pixels) to ensure accommodating the prostate and its surrounding tissue (Figure 2c). For each two-dimensional (2D) axial image slice, we flipped the slice about its anteroposterior axis (Figure 2e) and aligned the original slice to the flipped one using a 2D rigid image registration (Figure 2f). We used gradient descent optimizer and mean square error (MSE) metric to run the image registration. After registration, we measured the rotation angles of all the image slices and used their median after removing outliers divided by two as the patient's anteroposterior symmetry axis angle. Figure 2 shows the different steps of the alignment process.

Hounsfield unit range adjustment: The Hounsfield unit (HU) of prostate and its adjacent soft tissues are above -100 and below 155. Therefore, in order to reduce the inconsistency in air cavities and bones, we truncated the HU dynamic range of the images and HU values of -100 and 155 were respectively assigned to the voxels with HU values below -100 and above 155. To choose these HU threshold levels, we observed the HU range for prostate voxels across the training images. The HU values after removing the 0.1% outliers varied from -85 to 151. We chose -100 to 155 to limit the dynamic range to 256 (eight bit) levels.

Median filtering: To reduce the image noise we applied a 2D median filter as an edge-preserving, low-pass filter to each 2D axial slice. However, image filtering could disrupt the image pattern. Therefore to preserve the image texture while reducing the image noise, we applied the filter with a 3 \times 3 pixel window size which is the smallest size for median filtering.

Image resize: The images and their manual segmentation labels were up-sampled along inferior-superior axes using nearest neighbor interpolation to make the voxels isotropic. This image resampling made it easier to search in 3D space and select cubic image patches.

2.B.1.b. Shape modeling

Low-density point distribution model: To select a set of landmarks on the prostate surface and define a correspondence between the landmarks across the training set, we cast $N_I = 86$ equally spaced rays in 3D space, emanating from the centroid of the prostate gland and found the contact points between the rays and the prostate surface yielding a set of 86 landmarks. Using the centroid point as the origin of a sphere coordinate system, we defined all the landmarks with the same elevation and azimuth angles across the training set corresponding to each other. For modeling the prostate shape we defined each shape with this cloud of 86 surface points and the centroid point, yielding 87 points in total. The shapes were then aligned (translating, rotating and scaling transforms) using generalized 3D Procrustes analysis¹⁹ and the mean square distance between the points as the error metric.

High-density point distribution model: We applied the similar process used for LD PDM to build an HD PDM with $N_2 = 2056$ equally spaced casted rays. For HD PDM each shape was defined by 2057 points (2056 surface and one centroid points). To build the model the shapes were aligned (translating, rotating and scaling transforms) using generalized 3D Procrustes analysis and the mean square distance between the points as the error metric.

2.B.1.c. Feature selection

Local feature extraction: For each of the 86 rays (R_n , $n = 1, 2, 3, \dots, 86$) emanating from the prostate centroid (x_c, y_c, z_c) we selected a set of points (p_n) on the ray within a specific range around the prostate border point:

$$p_n = \{(x, y, z) | r_{\min} < r < r_{\max}, \theta = \theta_n, \varphi = \varphi_n\}, \quad (1)$$

where r , θ , and φ are, respectively, radial, elevation and azimuth coordinates of point (x, y, z) in a spherical coordinate system [Eq. (2) – (4)]. $r_{\min} = r_b - d$ and $r_{\max} = r_b + d$ are the radial coordinates of the first and the last points on the ray, respectively, where r_b is the prostate border point on the ray and d is the distance of the first and the last points from the border point. To focus on the local image textures near the prostate border, d should be small relative to the gland dimensions (in this paper $d = 5 \text{ mm}$) θ_n and φ_n are elevation and azimuth angles of points on the n th ray (R_n), respectively, and have constant values for all the points on the ray. Figure 3a illustrates the details.

$$r = \sqrt{(x - x_c)^2 + (y - y_c)^2 + (z - z_c)^2} \quad (2)$$

$$\theta = \cos^{-1} \frac{z}{r} \quad (3)$$

$$\varphi = \tan^{-1} \frac{y}{x} \quad (4)$$

We extracted all of the texture features within a 3D cubic image patch. For the point (x_p, y_p, z_p) on a 3D image, we defined the image patch P of size $W \times W \times W = (2D + 1) \times (2D + 1) \times (2D + 1)$ centered at the point (Figure 3b) as:

$$P = \mathcal{P}(x_p, y_p, z_p) = \{(x, y, z) \mid x_p - D < x < x_p + D, y_p - D < y < y_p + D, z_p - D < z < z_p + D\}. \quad (5)$$

In this paper we chose fixed value of 5 mm for parameter D to have patches that are small relative to the whole prostate gland but large enough to contain 3D image pattern for texture analysis.

For each of the cubic patches, we calculated a feature vector of size 1×67 consisting of a set of first- and second-order texture features that are listed in Table I. The patches that are centered at a point inside the prostate were labeled “prostate” or “one” and those centered at a point outside the prostate were labeled “non-prostate” or “zero”. We also measured the percentage of the patch volume within the prostate (L_p) for each patch. L_p is the number of prostate voxels in a patch divided by the total number of the patch voxels multiplied by 100. For a patch that is completely inside the prostate region $L_p=100\%$, for a patch that is completely outside the prostate $L_p=0$, and for patches that are as close to the prostate border to have an overlap with both prostate and non-prostate tissues, in which L_p is greater than zero and less than 100%.

Ray and feature selection: We collected the feature vectors from all the rays across the training images. Then for the values of each feature collected from all the corresponding rays, we applied the two-tailed, heteroscedastic t -test²⁶ to compare the values of the feature from prostate patches to the values of the same feature from non-prostate patches. In each case we tested the null hypothesis that the mean of the features measured from prostate patches is the same as the mean of the features measured from non-prostate patches. Rejection of the null hypothesis indicates a statistically significant difference between prostate and non-prostate means. For each of those features in which the null hypothesis was rejected ($\alpha=0.01$) we extracted two threshold values (T_0 and T_1); T_0 indicating that all of the feature values below/above it belong to non-prostate patches and T_1 indicates that all the feature values above/below it belong to prostate patches (see Figure 4).

We also used Spearman’s rank-order correlation (ρ) to measure the monotonic relationship between each of the features and L_p . For each ray, those features with high ($\rho > 0.6$) and

statistically significant ($p < 0.001$) correlation coefficients were selected for training a support vector machine²⁷ (SVM) classifier to classify between the prostate and non-prostate patches. Rays with no selected feature were excluded for SVM training; $M \leq N_1$ is the number of selected rays.

2.B.2. Segmentation

2.B.2.a. Preprocessing: We applied a similar preprocessing to that used for the training image to the test image and which consisted of Anteroposterior symmetry axis alignment, HU range adjustment, median filtering, and image resizing (see section 2.B.1.a).

2.B.2.b. Initialization

Operator inputs: To initialize the segmentation algorithm, a set of inputs from the operator are required, including the prostate gland bounds along the right-left (x), anteroposterior (y), and inferior-superior (z) axes, as well as a number of points on the prostate surface (anchor points) at different sub-regions, i.e. the apex, mid-gland, and base. For this purpose the operator approximated a minimum bounding box for the prostate gland that is limited inferiorly to the apex-most slice and superiorly to the base-most slice. The box included the entire prostate gland. Then for three, equally spaced slices between the defined apex and the base slices, the operator selected four anchor points on the prostate border on each slice and approximately at four, different sides, i.e. the right, left, anterior, and posterior sides, yielded total of 12 anchor points on the whole surface.

Ray casting: We used the centroid of the bounding box as an approximation for the prostate gland center and similar to the training part and casted 86, equally spaced rays emanating from the center point (Figure 5a).

Initially estimated prostate surface: We used the LD shape model generated during training to find a shape that fit within the bounding box and best matched the manually selected anchor points. For this purpose, we used the casted rays to determine the 12. corresponding points in the PDM to the anchor points. Then we used 3D thin-plate spline (TPS) analysis²⁸ to non-rigidly warp the mean shape of the model to the 12 anchor points. This helped to estimate the missing points between the anchor points. We then aligned the estimated shape to the mean shape of the model using 3D Procrustes analysis and extracted representative shape parameters from the PDM. We then restricted each parameter to the range of $[-\sqrt{3}\lambda_k, \sqrt{3}\lambda_k]$ in which λ_k is the k th eigenvalue of the shape model (corresponded to k th parameter), in order to determine the nearest shape of the model to the points.

2.B.2.c. Local classification

Classifier training: For each of the M selected rays during training, explained in section 2.B.1.c, we trained a linear kernel SVM algorithm for binary classification between prostate and non-prostate image patches, using the selected features for the ray.

Classification: We defined a search range ($[r_o - d_s, r_o + d_s]$) on each ray around the corresponding surface point from the initially estimated shape. r_o is the radius of the surface

point and d_s is the distance of the first and the last search point on the ray from the surface point. We selected a set of image patches centered at the ray points within the defined range and measured the feature vectors for each patch. We applied the threshold levels (T_0 and T_1) obtained from training to the corresponding feature values in order to classify the corresponding image patches to prostate and non-prostate. We shifted start and stop points of the range on the ray if the adjacent points to them were classified by thresholds. This could make the range narrower. Then we apply our SVM classifier to the features of the remaining unclassified patches to classify them into prostate and non-prostate. We shifted the initial surface point to the boundary of prostate and non-prostate points after removing potentially singular labels within the range. This process yielded a set of 86 candidate surface points (Figure 5b).

2.B.2.d. Shape regularization and surface reconstruction: We replaced the corresponding surface points with manually selected anchor points, in case they have shifted along the rays. Then we applied the HD PDM to the points to regularize the shape and in order have a plausible smooth shape for the prostate and to reconstruct the surface with a larger set of points. For that purpose, we first use TPS warping to non-rigidly register the 2056-point mean shape of the HD model to 86 obtained surface points, using the 86 corresponding points in the mean shape as the reference points and yielding a 2056-point candidate surface shape. We use 3D Procrustes analysis to register the shape to HD PDM and we measured the representing model parameters for the shape. We restricted each of the parameters, e.g. k th parameter, to the range of $[-\sqrt{3}\lambda_k, \sqrt{3}\lambda_k]$, in which λ_k is the corresponding (k th) eigenvalue. We then replaced the corresponding 12 points with the anchor points and also restricted the points to be within the bounding box. Finally, we used a scattered data interpolation²⁹ to generate a continuous surface out of the surface point set (Figure 5c – e), and resized the obtained label to the original image size so as to have the final segmentation results.

2.C. Evaluation

To evaluate our segmentation algorithm we compared the results against an expert observer's manual segmentations as the reference segmentations using a set of different types of error metrics explained in this section. We applied our error metrics to the entire prostate gland, as well as the inferior-most third (apex region), the superior-most third (base region), and the middle third (mid-gland) of the prostate.

2.C.1. Surface disagreement measurements—Mean absolute distance (MAD) measures the average disagreement between two surfaces where each surface is defined as a set of points. We defined MAD in unilateral and bilateral modes. See reference³⁰ for more details.

Hausdorff distance (HD) measures the maximum of the shortest distance between segmentation surface and reference surface. HD is sensitive to the noisy segmentation surface.

2.C.2. Regional overlap measurements—There are several, region-based error metrics that measure the overlap between two volumes. Dice similarity coefficient³¹ (DSC) is the most commonly used and sometimes the only region-based error metric reported in the literature. However, there are other region-based methods, e.g. recall or sensitivity rate (SR), precision rate (PR), and overlap error (OE) that have also been used to evaluate the segmentation algorithms and they could explain the error type, e.g. partial overlap, over segmentation and under-segmentation, better than DSC by itself. Reporting all the listed region-based error metric values for an algorithm might seem redundant, but for comparison purposes we applied all of them to our algorithm's segmentation results. In this manuscript, we reported all of the region- based error metrics in percentages. See reference ³⁰ and ¹⁷ for more details.

We also reported the signed volume difference (ΔV) between the algorithm segmentation and the reference, defined as:

$$\Delta V(V_s, V_{ref}) = V_s - V_{ref}, \quad (6)$$

where, V_s and V_{ref} are algorithm segmentation and reference segmentation volumes, respectively. ΔV is reported in cm^3 and as a percent in this manuscript. To measure the percentage of the ΔV , it is divided by the reference volume:

$$\Delta V(V_s, V_{ref}) = \frac{V_s - V_{ref}}{V_{ref}} \times 100. \quad (7)$$

For comparison of two manual references, we reported the absolute value of the volume difference and considered the average of the volumes as the reference volume:

$$\Delta V(V_{ref1}, V_{ref2}) = |V_{ref1} - V_{ref2}|, \text{ and} \quad (8)$$

$$\Delta V(V_{ref1}, V_{ref2}) = \frac{|V_{ref1} - V_{ref2}|}{\frac{1}{2}(V_{ref1} + V_{ref2})} \times 100, \quad (9)$$

where, V_{ref1} and V_{ref2} are manual reference segmentations obtained from two experts.

3. Results

1.A. Implementation details

We implemented the proposed segmentation algorithm using MATLAB R2017a (version 9.2.0) on a 64-bit Windows 7 desktop with a 3.0 GHz Intel Xeon processor and with 64 GB memory. To speed up the algorithm execution, we developed a parallel implementation of

the algorithm using the MATLAB parallel computing toolbox and ran the code on 12 CPU cores.

To avoid interference of the brachytherapy seeds pattern on algorithm training, we trained our algorithm using non-brachytherapy images only. We used one set of the manual segmentation (reference #1) for training. We randomly selected 70% of our non-brachytherapy images, i.e. 23 images, for training the algorithm. The remaining 10 images were used for testing the algorithm. We used all of the 37 post-LDR brachytherapy images only for testing the algorithm.

For anteroposterior symmetry axis alignment during training, we selected HU of 155 as the threshold level for rough segmentation of bones. We cropped the images about the centroid to make a 121 pixels (along the x-axis) by 201 pixels (along the y-axis) 2D slices. We used the same image cropping for anteroposterior symmetry axis alignment during segmentation. We considered all of the rotation angles (α_{iT} , $i = 1, 2, 3, \dots$, *number of axial slices*) greater than 18 degrees as outliers. We set d to 5 mm and D to 5 mm ($W = 2D + 1 = 11$ mm).

During segmentation we used the same patch size used during training ($11 \times 11 \times 11$ mm) and d_s was initially set to 7 mm.

1.B. Interobserver variability in manual segmentation

We compared our two sets of manual segmentation labels using our segmentation error metrics to measure the interobserver variation in expert prostate border delineation. Table II shows the result. The mean \pm standard deviation of the prostate gland volume based on the first and the second manual segmentations were 30.1 ± 12.6 cm³ and 28.0 ± 10.0 cm³, respectively.

1.C. Segmentation algorithm accuracy and computation time

1.A.1. Single operator, single reference test on the non-brachytherapy test dataset—We applied the proposed segmentation algorithm to 10 non-brachytherapy test images using reference segmentation #1 for initializing the algorithm and evaluating the results. We defined the bounding box based on the reference label and selected 12 anchor points on the label surface, as described in 2.B.2.b. Table III shows the quantitative accuracy of the algorithm based on the error metrics. We conducted a one-tailed t -test between values of Table III and the corresponding metric values in Table II. The null hypotheses were defined regarding the relative metric values, and the values in bold show where the null hypotheses were rejected with $\alpha = 0.05$. Figure 6 and Figure 7 illustrate the non-brachytherapy image segmentation results qualitatively in 2D and 3D, respectively. The average measured segmentation computation time was 22 ± 2 s. per 3D image.

To show the effect of 12 anchor points on the performance of the algorithm, we also run the algorithm without the anchor points. Table IV shows the results. We conducted a one-tailed t -test between the metric values in the table to the corresponding metric values in Table III. The null hypotheses were defined regarding the relative metric values and the values in bold in Table IV show where the null hypotheses were rejected with $\alpha = 0.05$.

We compared the algorithm performance to some of the most recent segmentation algorithms presented in the literature in Table V using our error metric values where applicable. In those studies in which MAD was used as an error metric, but the mode of MAD calculation was not mentioned, we indicate the MAD values with an asterisk sign. We conducted on-tailed *t*-tests to compare our results with those of previous work where applicable. In these tests the null hypotheses were defined regarding the relative performance of the methods. The values in bold in Table V show where the null hypotheses were rejected with $\alpha = 0.05$.

1.A.2. Two operators, two references test on the non-brachytherapy test

dataset—To investigate the effect of using different operator on the algorithm performance, we applied the algorithm with the same configuration applied in 3.C.1 and used reference segmentation #2 for algorithm initialization. We evaluated the segmentation results by comparing the segmentation labels against reference segmentation #2 as the expected results. Figure 8 compares these results and the segmentation results obtained in 3.C.1 to the observed difference between two references for the 10 non-brachytherapy images. Box plots of Figure 8 summarize the comparison between the three groups in terms of any of the three metrics. We conducted the one-way ANOVA test followed by pairwise *t*-tests with the null hypothesis that the average of each metric for the three groups are the same. The results of the post ANOVA test are indicated in Figure 8 where ANOVA detected a statistically significant difference ($\alpha = 0.05$).

1.A.3. Single operator, single reference test on the post-LDR brachytherapy

test dataset—Table VI shows the measured segmentation error of the proposed algorithm for 37 post-LDR brachytherapy images. The algorithm was trained using the same 23 non-brachytherapy training images in section 3.C.1. The average measured computation time was 21 ± 1 s. per 3D image. We conducted one-tailed *t*-tests between each of the metric values in this experiment and the metric values obtained on non-brachytherapy (Table III). In these tests the null hypotheses were defined regarding the relative performance of the algorithm with different test data sets. The values in bold show where the null hypotheses were rejected with $\alpha = 0.05$.

1.A.4. Operator interaction time

—To record the required operator interaction time for algorithm initialization, we asked two operators (a research scholar and an MD-PhD graduate student both with experience in reading prostate CT images) to select the bounding box and the anchor points on all the non-brachytherapy and post-LDR brachytherapy test images. The average recorded time for selecting bounding box were 41 ± 9 s. (34 ± 9 s. for non-brachytherapy test images and 42 ± 8 s. for post-LDR brachytherapy test images) and 40 ± 14 s. (53 ± 15 s. for non-brachytherapy test images and 37 ± 12 s. for post-LDR brachytherapy test images) for the first and the second operators, respectively. The average recorded time for selecting 12 anchor points were 18 ± 3 s. (18 ± 1 s. for non-brachytherapy test images and 18 ± 3 s. for post-LDR brachytherapy test images) and 16 ± 2 s. (14 ± 2 s. for non-brachytherapy test images and 16 ± 2 s. for post-LDR brachytherapy test images) for the first and the second operators, respectively. The average total operator interaction time (selecting bounding box and 12 anchor points) were 59 ± 10 s. (53 ± 9 s. for non-

brachytherapy test images and 61 ± 9 s. for post-LDR brachytherapy test images) and 56 ± 14 s. (67 ± 14 s. for non-brachytherapy test images and 53 ± 12 s. for post-LDR brachytherapy test images) for the first and the second operators, respectively.

4. Discussion

4.A. Interobserver variability in manual segmentation

Pairwise comparison of our two, manual reference segmentations (Table II) shows high interobserver variation in manual segmentation of the prostate in CT images. For example, we observed that the measured disagreement between the experts for the whole prostate gland was ranging from 72% to 97% in terms of DSC, and from 0.4 mm to 3.3 mm in terms of MAD. This disagreement is slightly lower in the mid-gland region and higher in the base and apex regions, and we did not observe a meaningful difference between the base and the apex in terms of manual segmentation difficulty. This variation in manual segmentation between expert observers makes it challenging to evaluate the segmentation algorithm using a single-observer reference segmentation. Furthermore, it challenges the comparison between two, proposed algorithms based on the reported error metrics when they have been tested on different image datasets with different reference segmentations. It could be helpful for reducing the interobserver variability effects on evaluation if we evaluate our algorithm's segmentations against multiple, manual reference segmentations on the same image dataset.

4.B. Segmentation algorithm accuracy and computation time

With respect to the interobserver variability observed in manual segmentation, there is no gold standard defined for validation of the prostate segmentation in CT images. Therefore, the best reasonable and measurable performance compared to a manual reference segmentation which could be reached for a computer-assisted segmentation method should not be less than the highest observed variation range between each two experts in manual segmentation. For the proposed algorithm, a comparison between corresponding mean values of Table II and Table III shows that there is still a gap between the performance of our algorithm and the measured difference between the two observers. The results suggest that there is still room for improvement of the algorithm in terms of the measured error metrics. However, comparing the standard deviation of the metrics and the ranges of the metric values in Table II and Table III shows that the magnitude of this variation in terms of most of the error metrics is higher when two human experts are being compared together. It means that there are some cases for which the measured difference between the algorithm segmentation and a human expert manual segmentation is lower than the measured difference between that expert and another human expert's manual segmentation. In Figure 8, the bar graphs of DSC and MAD (Figure 8a and c) for P2 and P8 and the bar graphs of HD (Figure 8e) for P2, P6, P7, and P8 also support this finding.

For comparison with previous work published in the literature, Table V shows that the performance of the proposed method was within the reported metric values in terms of MAD and DSC and it outperformed the other methods listed in the Table in terms of SR. The better performance of those methods (references ^{10–14}) in Table V that have used the previously acquired treatment images of the same target patient for training the algorithm

seems reasonable. In general, the reported errors in terms of MAD and DSC for those segmentation methods that use prostate border information of previous planning and/or treatment images of the target patient are lower than the MAD and DSC errors we measured for our method. Our method has similar performance or outperformed the other methods in terms of MAD and DSC. We measured SR of 94% for the proposed segmentation method, and which is higher than the reported SR values of the previous method. This suggests that on average, our algorithm's segmentation label covers the reference better than the other algorithms and that is an important feature for an algorithm that may be utilized for radiation therapy. The measured PR of our algorithm, i.e. the proportion of the segmentation that covers the reference, is lower than the SR (82%). This means that 18% of the voxels are incorrectly classified as prostate. Therefore, this could cause irradiation of the healthy tissues during radiotherapy and needs to be improved.

Comparing the measured error metrics for the proposed segmentation method tested on post-LDR brachytherapy images (Table VI) to our results on non-brachytherapy images (Table III) showed no meaningful or significant difference between the corresponding metric values. This suggests that our method may be used for both non-brachytherapy and post-LDR brachytherapy images. One explanation for this finding may be that in our algorithm prostate the surface is searched locally and close to the boundary region where fewer brachytherapy seeds are located and interfere the algorithm performance. The potential interference of few brachytherapy seeds in the prostate surface point search may be compensated by using shape model and shape regularization.

The main part of the segmentation procedure of the proposed algorithm is run on a set of rays, and the execution on each ray is completely independent of the others. Therefore, the segmentation part is computationally parallelizable and it makes the algorithm highly capable for computation speed up. A parallel implementation on an unoptimized MATLAB research platform, using 12 CPU cores yields to about 20 s. of the computation time per 3D image, and which is lower than the other reported execution time in Table V. The total segmentation time including the manual operator interaction time was less than 1.5 min per 3D image which is substantially less than fully manual delineation time of 4.46 min reported in ⁵.

4.C. Limitations

The performance of our segmentation algorithm should be considered in the context of the strengths and limitations of the proposed method. To apply the fully 3D image search we resampled the images to deal with relatively large through-plane to in-plane voxel dimension ratio of 4.25:0.977 and make the voxel size isotropic. However, image resampling could affect the texture extraction part. We used nearest neighbor interpolation to reduce this effect as much as possible. Moreover, optimization on the size of the median filter for image denoising might be helpful to improve the texture-preserving noise reduction process. We also used a fixed patch size of $11 \times 11 \times 11$ voxels for this study. An optimization on patch size might help to improve the performance of the segmentation algorithm. The small size of the non-brachytherapy testing dataset (10 images) is another limitation of this study.

5. Conclusions

In this manuscript, we have presented a semiautomatic, learning-based technique for full 3D segmentation of the prostate on CT images. Our method is trained based on local texture and shape characteristics of the prostate in CT images. The algorithm is initialized with a set of operator inputs to start the search for prostate surface around an estimated prostate surface generated using a shape modeling method. A radial search is then performed to extract a set of local texture features of 3D cubic image patches centered at a set of points along a number of rays and close to the initial estimated surface. A ray-specific trained SVM classifier is then used for each ray to predict the prostate surface point on the ray. The extracted surface points are regularized and interpolated to form a smooth, plausible 3D surface for the prostate. The proposed method is not related to patient-specific data, i.e. previously acquired and segmented CT images from the same patient, for segmenting a patient's image and could, therefore, also be used for radiotherapy planning CT image segmentation. To have a better interpretation of the algorithm performance, we evaluated our segmentation method against manual reference segmentation using a set of surface-, region overlap-, and volume-based error metrics. We measured the error metrics for the whole prostate gland as well as for the apex, mid-gland, and base regions. We also used two, different sets of manual reference segmentations to evaluate our method robustness to changing the reference. The measured error for the proposed algorithm against manual segmentation shows that our segmentation method achieved a segmentation accuracy close to the variation range observed in manual segmentation and comparable to the previously presented work with a statistically significant higher sensitivity compared to the previous work.

Acknowledgments

This research was supported in part by the U.S. National Institutes of Health (NIH) grants (CA176684, CA156775, and CA204254). The work was also supported in part by the Georgia Research Alliance (GRA) Distinguished Cancer Scientist Award to BF.

References

1. Siegel RL, Miller KD, and Jemal A, Cancer statistics, 2016. CA: a cancer journal for clinicians, 2016 66(1): p. 7–30. [PubMed: 26742998]
2. Martin NE, Massey L, Stowell C, Bangma C, Briganti A, Bill-Axelsson A, Blute M, Catto J, Chen RC, and D'Amico AV, Defining a standard set of patient-centered outcomes for men with localized prostate cancer. European urology, 2015 67(3): p. 460–467. [PubMed: 25234359]
3. Fowler JE, Braswell NT, Pandey P, and Seaver L, Prostate Cancer: Experience With Radical Prostatectomy and Radiation Therapy for Localized Prostate Cancer at a Veterans Affairs Medical Center. The Journal of urology, 1995 153(3): p. 1026–1031. [PubMed: 7853550]
4. Acosta O, Dowling J, Dreon G, Simon A, De Crevoisier R, and Haigron P, Multi-atlas-based segmentation of pelvic structures from CT scans for planning in prostate cancer radiotherapy, in Abdomen and Thoracic Imaging. 2014, Springer p. 623–656.
5. Langmack KA, Perry C, Sinstead C, Mills J, and Saunders D, The utility of atlas-assisted segmentation in the male pelvis is dependent on the interobserver agreement of the structures segmented. Br J Radiol, 2014 87(1043): p. 20140299. [PubMed: 25168198]
6. Smith WL, Lewis C, Bauman G, Rodrigues G, D'Souza D, Ash R, Ho D, Venkatesan V, Downey D, and Fenster A, Prostate volume contouring: a 3D analysis of segmentation using 3DTRUS, CT, and MR. Int J Radiat Oncol Biol Phys, 2007 67(4): p. 1238–47. [PubMed: 17336224]

7. Villeirs GM, Van Vaerenbergh K, Vakaet L, Bral S, Claus F, De Neve WJ, Verstraete KL, and De Meerleer GO, Interobserver Delineation Variation Using CT versus Combined CT+ MRI in Intensity-Modulated Radiotherapy for Prostate Cancer. *Strahlentherapie und Onkologie*, 2005 181(7): p. 424–430. [PubMed: 15995835]
8. Fiorino C, Reni M, Bolognesi A, Cattaneo GM, and Calandrino R, Intra-and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning. *Radiotherapy and oncology*, 1998 47(3): p. 285–292. [PubMed: 9681892]
9. Rasch C, Barillot I, Remeijer P, Touw A, van Herk M, and Lebesque JV, Definition of the prostate in CT and MRI: a multi-observer study. *International Journal of Radiation Oncology* Biology* Physics*, 1999 43(1): p. 57–66.
10. Feng Q, Foskey M, Chen W, and Shen D, Segmenting CT prostate images using population and patient-specific statistics for radiotherapy. *Medical Physics*, 2010 37(8): p. 4121–4132. [PubMed: 20879572]
11. Liao S, Gao Y, Lian J, and Shen D, Sparse patch-based label propagation for accurate prostate localization in CT images. *IEEE transactions on medical imaging*, 2013 32(2): p. 419–434. [PubMed: 23204280]
12. Shi Y, Liao S, Gao Y, Zhang D, Gao Y, and Shen D. Prostate segmentation in CT images via spatial-constrained transductive lasso. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2013.
13. Shi Y, Gao Y, Liao S, Zhang D, Gao Y, and Shen D, Semi-Automatic Segmentation of Prostate in CT Images via Coupled Feature Representation and Spatial-Constrained Transductive Lasso. *IEEE transactions on pattern analysis and machine intelligence*, 2015 37(11): p. 2286–2303. [PubMed: 26440268]
14. Shi Y, Gao Y, Liao S, Zhang D, Gao Y, and Shen D, A learning-based CT prostate segmentation method via joint transductive feature selection and regression. *Neurocomputing*, 2016 173: p. 317–331. [PubMed: 26752809]
15. Shao Y, Gao Y, Yang X, and Shen D. CT prostate deformable segmentation by boundary regression in International MICCAI Workshop on Medical Computer Vision. 2014 Springer.
16. Shao Y, Gao Y, Wang Q, Yang X, and Shen D, Locally-constrained boundary regression for segmentation of prostate and rectum in the planning CT images. *Medical image analysis*, 2015 26(1): p. 345–356. [PubMed: 26439938]
17. Ma L, Guo R, Tian Z, Venkataraman R, Sarkar S, Liu X, Tade F, Schuster DM, and Fei B. Combining population and patient-specific characteristics for prostate segmentation on 3D CT images in SPIE Medical Imaging. 2016 International Society for Optics and Photonics.
18. Ma L, Guo R, Zhanga G, Tadea F, Schuster DM, Niehc P, Masterc V, and Fei B. Automatic segmentation of the prostate on CT images using deep learning and multi-atlas fusion in SPIE Medical Imaging. 2017 International Society for Optics and Photonics.
19. Gower JC, Generalized procrustes analysis. *Psychometrika*, 1975 40(1): p. 33–51.
20. Materka A and Strzelecki M, Texture analysis methods—a review. Technical university of lodz, institute of electronics, COST B11 report, Brussels, 1998: p. 9–11.
21. Dalal N and Triggs B. *Histograms of oriented gradients for human detection*. in *Computer Vision and Pattern Recognition*, 2005. CVPR 2005. IEEE Computer Society Conference on 2005 IEEE.
22. Ojala T, Pietikäinen M, and Harwood D, A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 1996 29(1): p. 51–59.
23. Haralick RM and Shanmugam K, Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, 1973(6): p. 610–621.
24. Albregtsen F, Statistical texture measures computed from gray level cooccurrence matrices Image processing laboratory, department of informatics, university of oslo, 2008 5.
25. Rafael Gonzalez C and Woods R, Digital image processing. Pearson Education, 2002.
26. Woolson RF and Clarke WR, Statistical methods for the analysis of biomedical data. Vol. 371 2011: John Wiley & Sons.
27. Andrew AM, An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods by Nello Christianini and John Shawe-Taylor, Cambridge University Press, Cambridge, 2000, *xiii+ 189 pp., ISBN 0–521-78019–5 (Hbk, £ 27.50)*. 2000, Cambridge Univ Press.

28. Bookstein FL, Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on pattern analysis and machine intelligence*, 1989 11(6): p. 567–585.
29. Amidror I, Scattered data interpolation methods for electronic imaging systems: a survey. *Journal of electronic imaging*, 2002 11(2): p. 157–176.
30. Shahedi M, Cool DW, Romagnoli C, Bauman GS, Bastian-Jordan M, Gibson E, Rodrigues G, Ahmad B, Lock M, Fenster A, and Ward AD, Spatially varying accuracy and reproducibility of prostate segmentation in magnetic resonance images using manual and semiautomated methods. *Med Phys*, 2014 41(11): p. 113503. [PubMed: 25370674]
31. Dice LR, Measures of the amount of ecologic association between species. *Ecology*, 1945 26(3): p. 297–302.

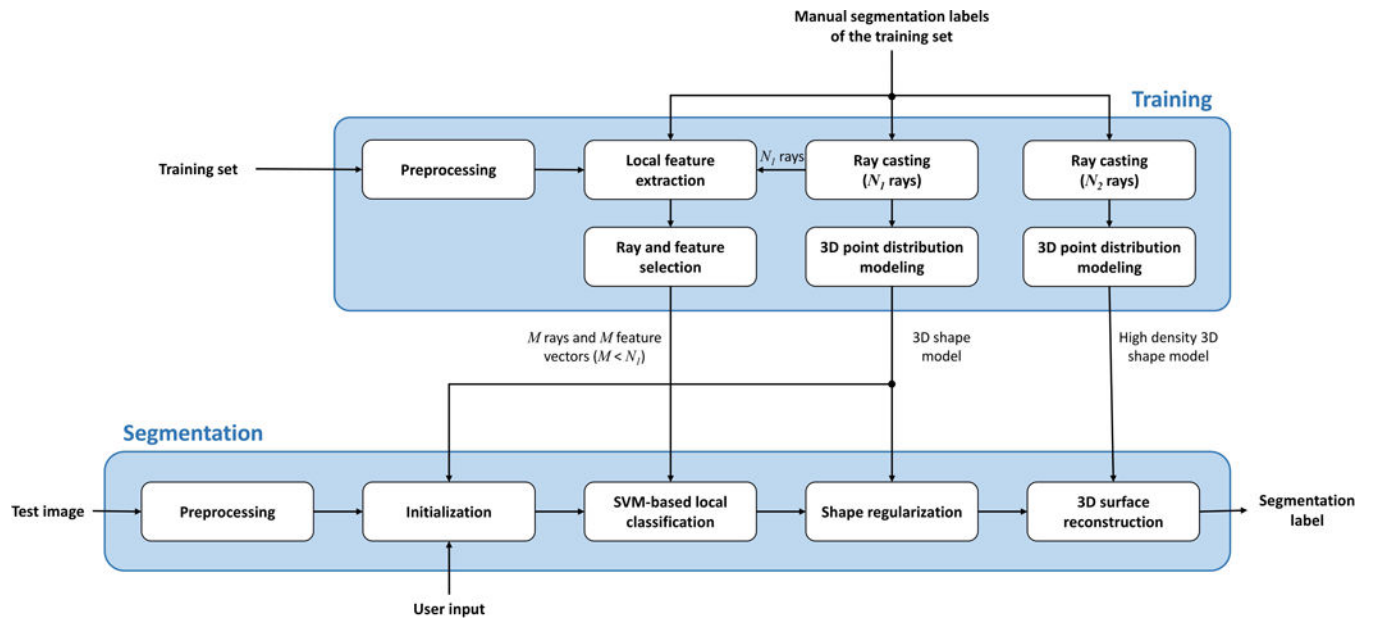


Figure 1.
The general framework of the proposed segmentation method.

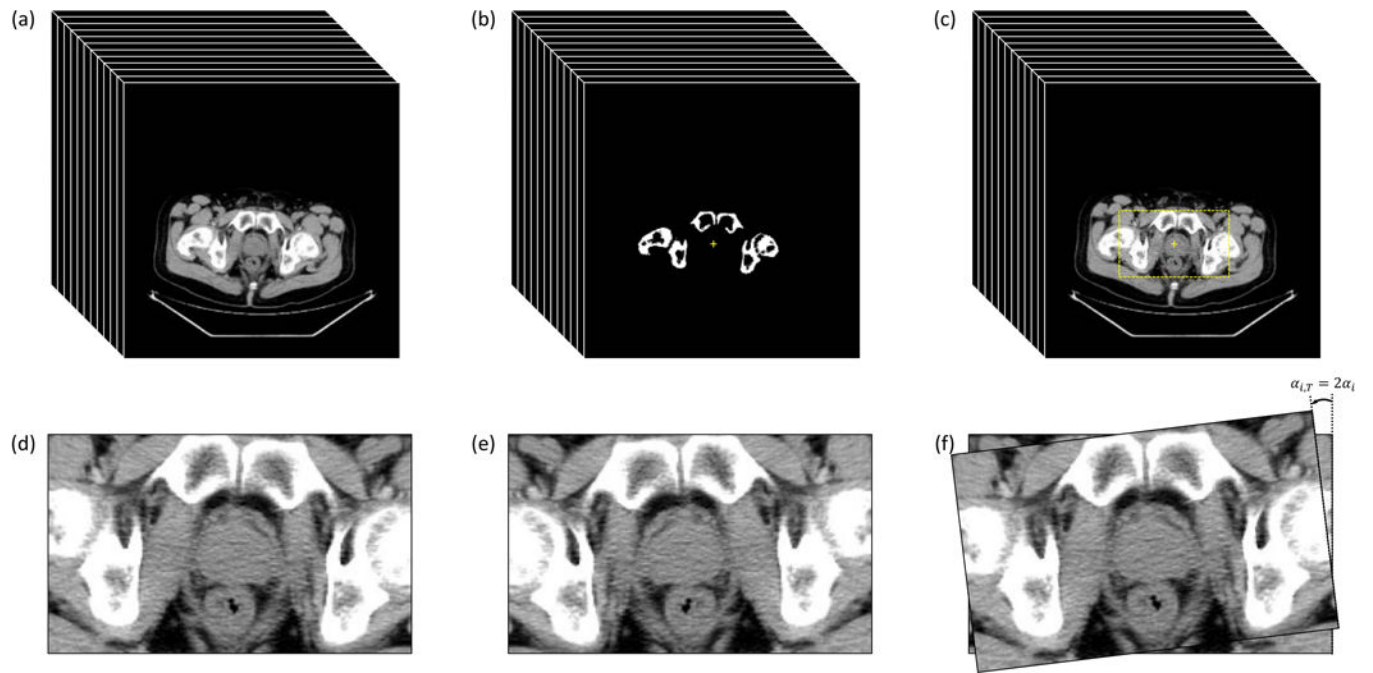


Figure 2.

Anteroposterior symmetry axis angle measurement. (a) Original 3D CT image. (b) Rough hip bone segmentation label after thresholding. The yellow cross shows the centroid of the bones. (c) The cropped region (dashed box) about the centroid. (d) i th 2D slice of the image and (e) its flipped version. (f) Rigid registration between (d) and (e). $\alpha_{i,T}$ is the rotation angle after rigid registration and α_i is the measured angle between the patient anteroposterior symmetry axis on i th slice and the anteroposterior axis of the image.

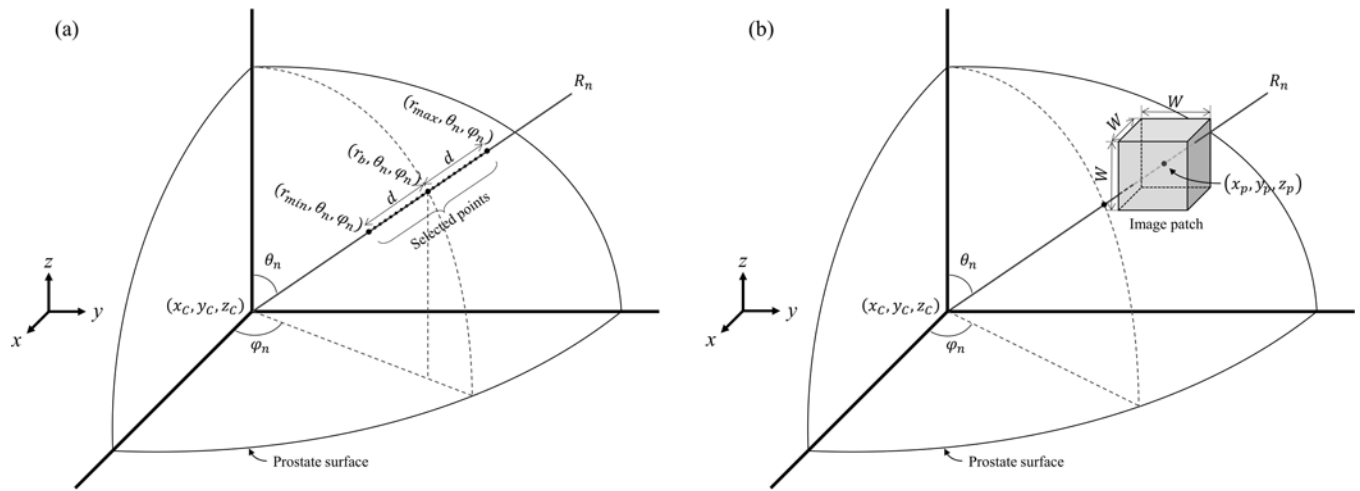


Figure 3.

A schematic illustration of the prostate surface, (a) the selected points on a sample ray (R_n) used for feature extraction, and (b) a selected 3D cubic image patch centered at a sample point, (x_p, y_p, z_p) , on ray R_n .

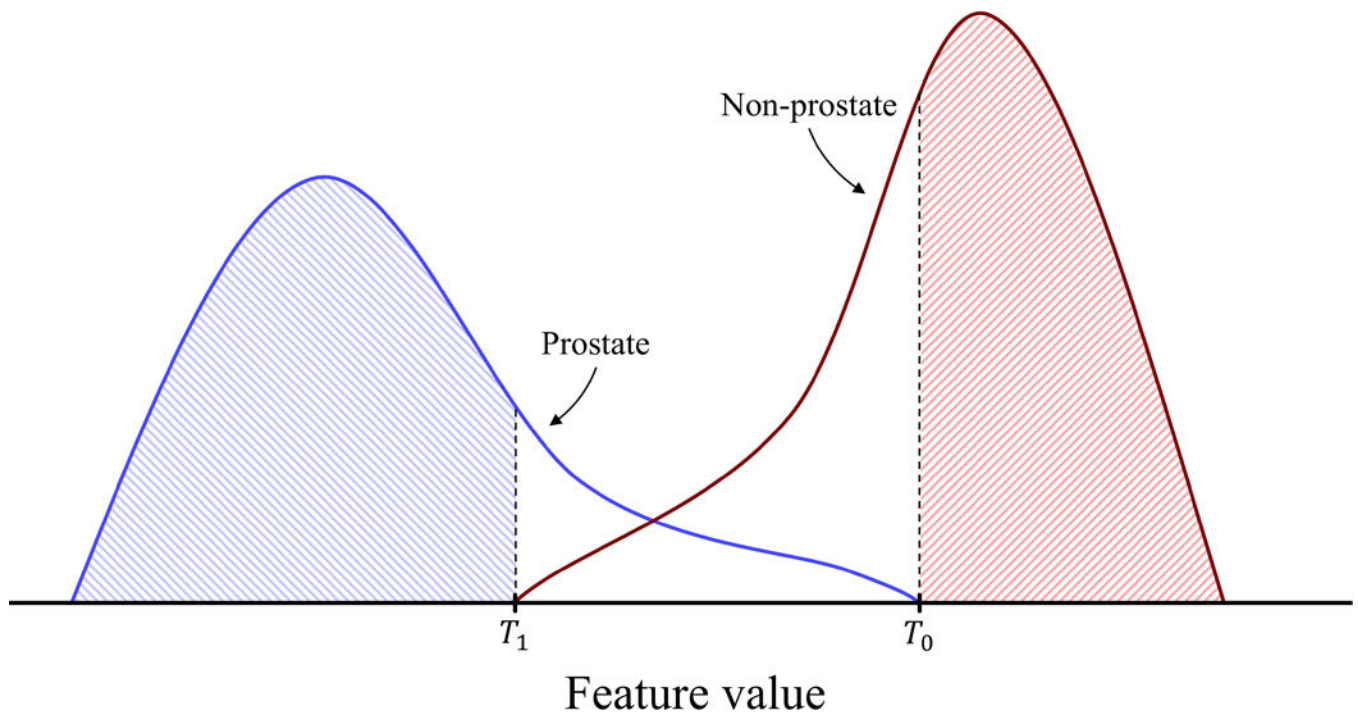


Figure 4.

Histograms of feature values measured from prostate and non-prostate image patches. T_0 is the threshold level in which all of the feature values above that level belong to non-prostate image patches, and T_1 is the threshold level in which the feature values below that level belong to prostate image patches. The feature values between T_0 and T_1 belong to either prostate or non-prostate image patches.

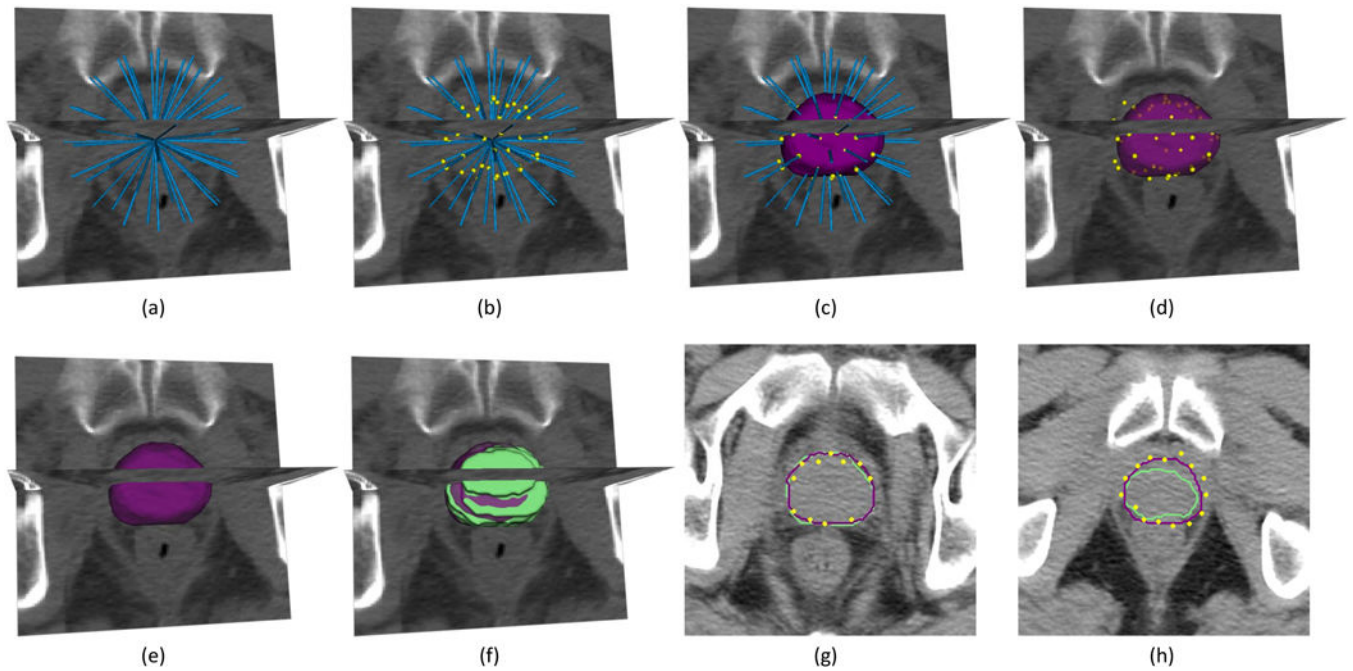


Figure 5.

Segmentation procedure steps. (a) 86 rays cast from a center point of the gland. (b) Yellow dots show the candidate surface points. (c) – (e) Purple surface shows the algorithm results after shape regularization and reconstruction. (f) Comparison between the algorithm segmentation in purple and the reference in green. (g) and (h) show algorithm results in purple, reference in green, and surface candidate points (yellow dots). Note that yellow and green appear bright and purple appears dark in grayscale print.

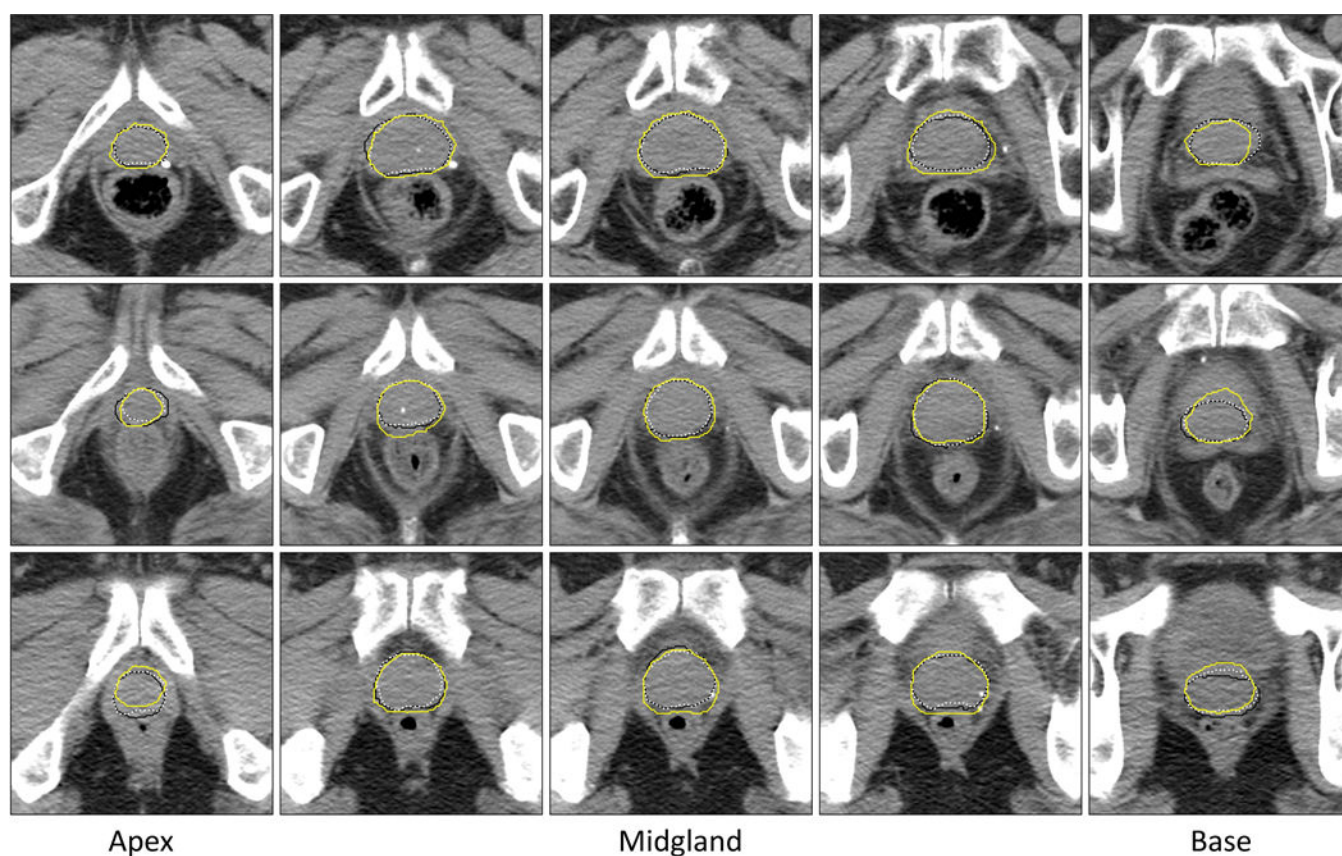


Figure 6.

Qualitative segmentation results on five, 2D axial slices for three sample prostates. Each row shows the results for one patient. The left column shows the apex slices, and the right column shows the base slices. The algorithm segmentation is shown with yellow (or bright in grayscale print) contours, the first reference is shown with black contours, and the second reference is shown with white dotted contours.

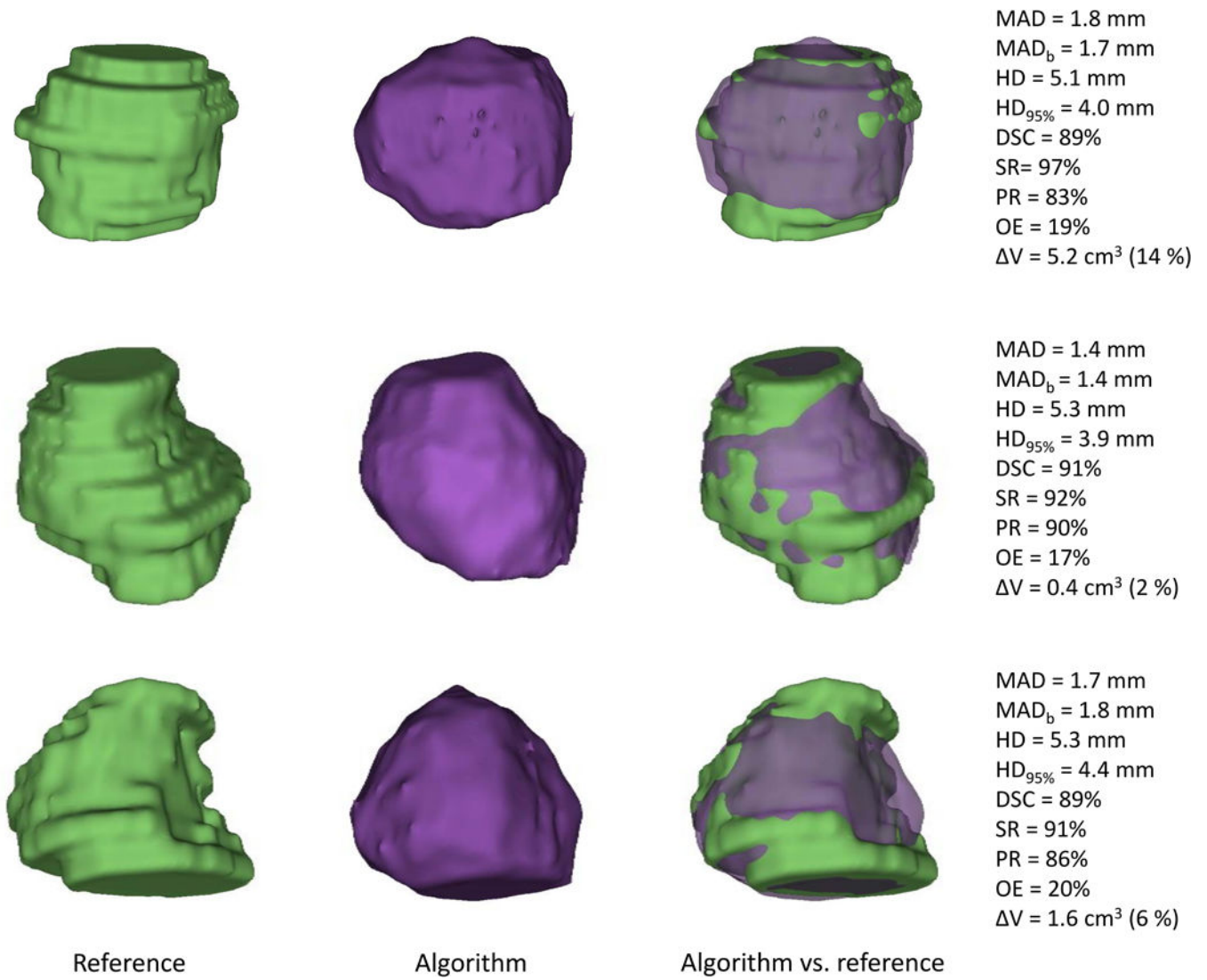


Figure 7.
Qualitative and quantitative segmentation results in 3D for the three, sample prostates shown in Figure 2.

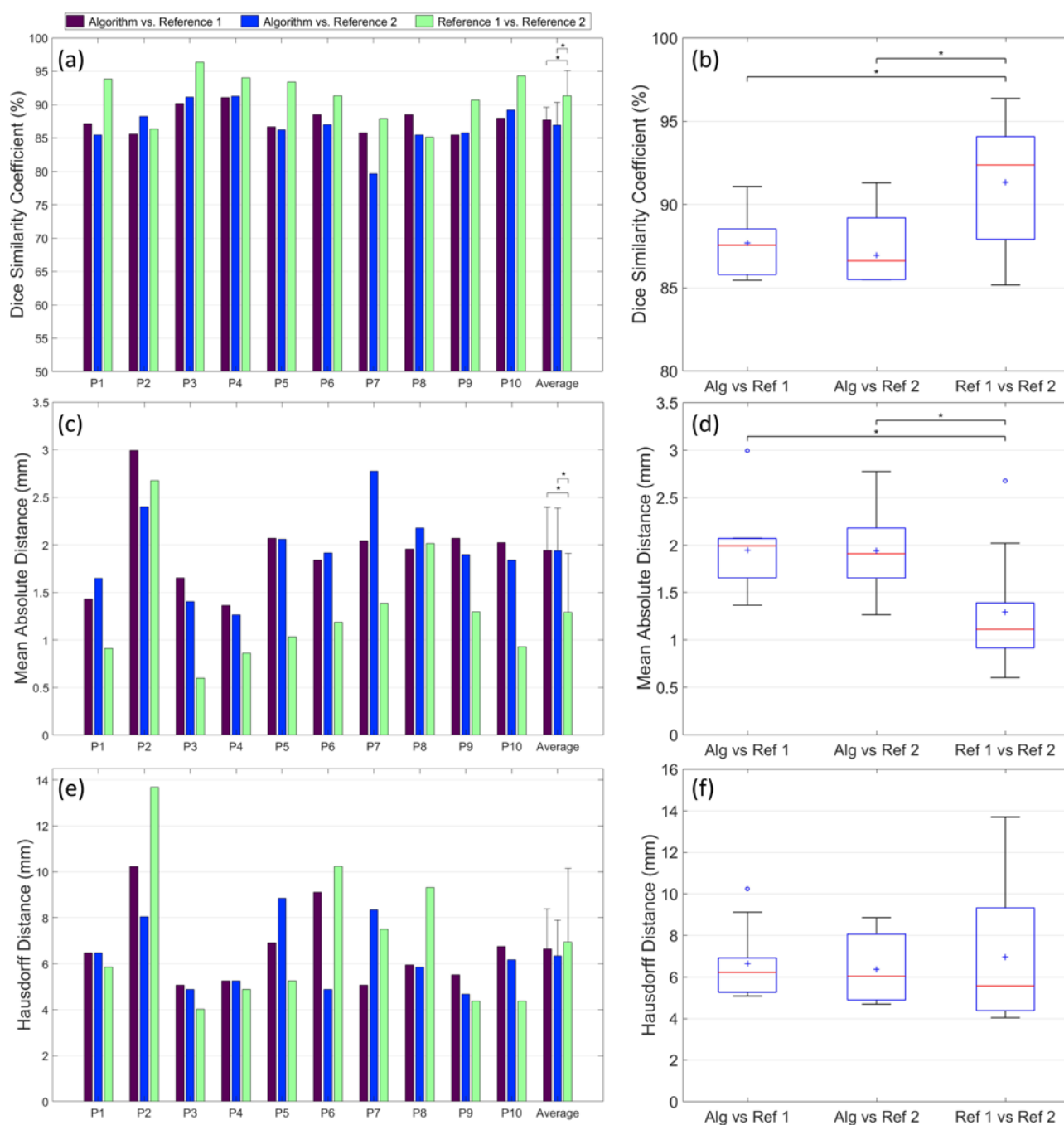


Figure 8.

Performance of the proposed segmentation algorithm based on two operators and two references vs interobserver variability in manual segmentation for 10 non-brachytherapy test images in terms of (a) – (b) DSC, (c) – (d) MAD (for reference 1 to reference 2 comparison MAD_b was measured) and (e) – (f) HD error metrics. Box plots (b), (d) and (f) show the minimum (lower bar), 25 percentile to 75 percentile (blue box), maximum (upper bar), mean (blue cross symbol), median (red line segment) values of the metrics, and the outliers (blue

‘o’ symbols) where they exist. Asterisk symbols indicate statistically significant differences between the averages of the metrics ($p < 0.05$).

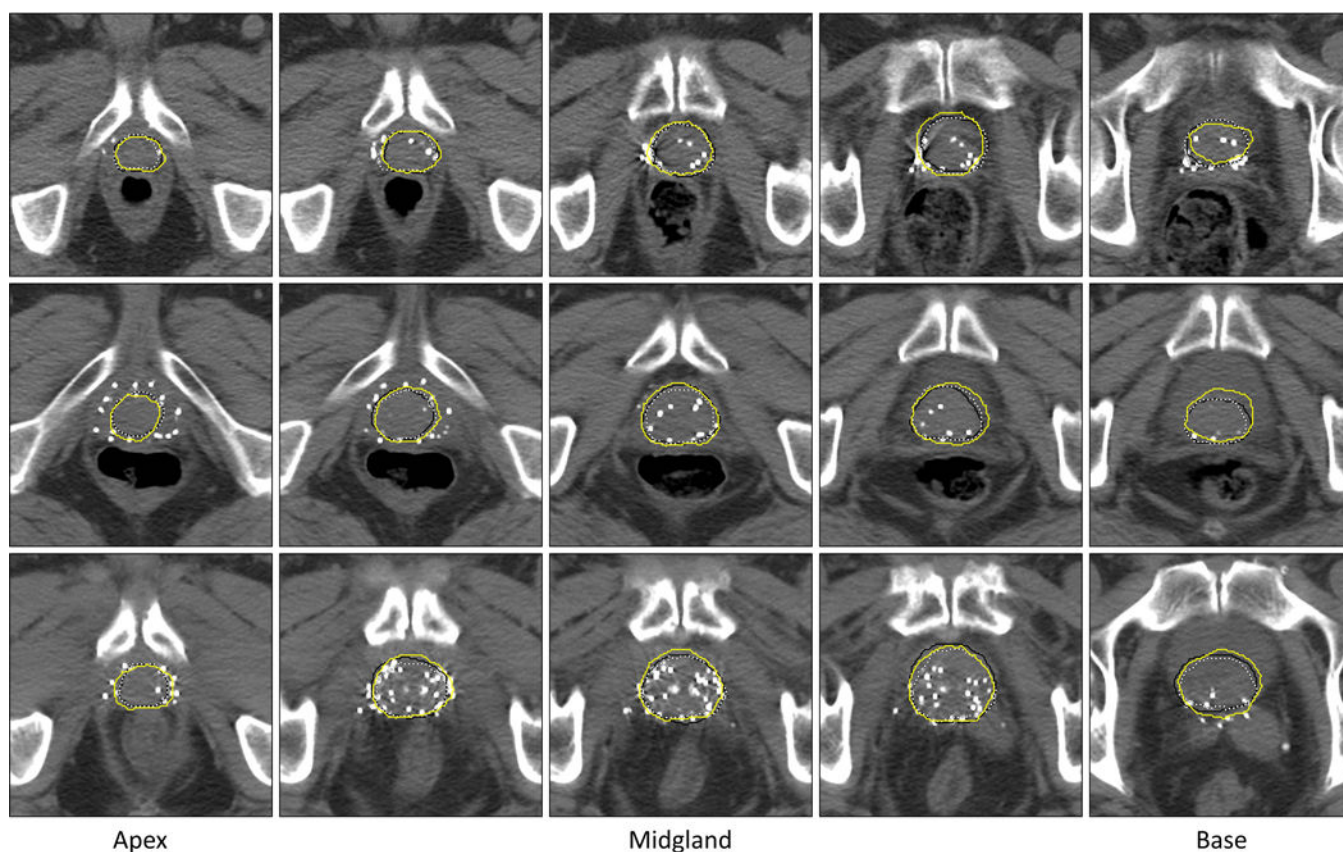


Figure 9. Qualitative segmentation results on five, 2D axial slices for three sample post-LDR brachytherapy prostate images. Each row shows the results for one patient. The left column shows the apex slices, and the right column shows the base slices. The algorithm segmentation is shown with yellow (or bright in grayscale print) contours, the first reference is shown with black contours, and the second reference is shown with white dotted contours.

Table I.

List of features and their descriptions.

Feature	# of features	Description
First-order features	10	Intensity of the patch center voxel, mean (μ_P), standard deviation (σ_P), median (m_P), minimum intensity, maximum intensity, skewness ²⁰ , kurtosis ²⁰ , entropy ²⁰ , and energy ²⁰ of a 3D patch intensity histogram. Since the first-order texture features are measured based on the image histogram, they could be easily implemented for 3D patches.
Histogram of oriented gradients (HOG) ²¹	8	HOG describes the distribution of intensity gradients in an image. Eight, 2D orientation bins are defined as forming the HOG. In this study we calculated one histogram per 2D axial patch slice and obtained the average of all the corresponding bins of the HOGs to form one, eight-bin HOG for a 3D image patch.
Histogram of Local binary patterns (LBP) ²²	8	LBP is calculated in 3D within a 3×3×3 mask. One, eight-bin histogram is formed per 3D image patch.
Grey-level co-occurrence, matrices-based ²³ (GLCM)-based features	32	For each pixel spatial relationship (pixel neighboring) in 2D (2D offset), GLCM for a 3D patch is defined as the average of the GLCMs of all the 2D axial patch slices related to the same pixel neighboring. GLCM-based features were measured based on four, different 2D offsets, i.e. (-1,0), (0,-1), (1,-1), (-1,-1), and consist of energy, entropy, contrast ²⁰ , homogeneity ²⁴ , inverse different moment ²⁰ , correlation ²⁰ , cluster shade ²⁴ , and cluster prominence ²⁴ of the four GLCMs.
Mean gradient angle	1	Mean gradient angle of a 3D patch is the average of mean gradient angles of all the 2D axial slices.
Edge-based features	8	The eight-bin histogram of edge directions in a 3D patch. The edges are detected using Sobel operator ²⁵ for edge detection for all the 2D slices.
Total	67	

Table II.

The average and range of interobserver variability in manual prostate segmentation across the dataset. Mean \pm standard deviation [range] observed a difference between two experts based on our error metrics. As both segmentation labels in each pairwise comparison were from experts, the bilateral MAD (MAD_b) and $|V|$ are reported in this Table. N_{pat} and N_{img} are the number of patients and the number of images, respectively.

Region of Interest	N_{pat}	N_{img}	MAD (mm)	HD (mm)	DSC (%)	SR (%)	PR (%)	OE (%)	$ V $ (cm^3)
Whole Gland	70	70	1.1 ± 0.5	6.2 ± 2.4	92 ± 4	89 ± 5	95 ± 5	15 ± 7	2.4 ± 3.3
Apex			1.1 ± 0.7	4.7 ± 2.3	91 ± 9	88 ± 8	94 ± 11	17 ± 12	0.9 ± 1.3
Midgland			1.1 ± 0.5	4.3 ± 1.6	93 ± 3	89 ± 5	98 ± 2	12 ± 5	1.2 ± 1.4
Base			1.2 ± 0.7	4.8 ± 2.0	91 ± 7	91 ± 6	92 ± 12	17 ± 11	1.2 ± 1.1

Quantitative accuracy of the segmentation methods on non-brachytherapy images. Mean \pm standard deviation [range] of the error metrics for different regions of interest. N_{pat} and N_{img} are the number of patients and the number of images, respectively. Twenty-three non-brachytherapy images were used for training the algorithm. There are statistically significant differences detected between those metric values in bold and the corresponding metric values in Table VI, where applicable ($p < 0.05$).

Table III.

Region of Interest	N_{Pat}	N_{Img}	MAD (mm)	MAD _b (mm)	HD (mm)	DSC (%)	SR (%)	PR (%)	OE (%)	V (cm ³)
Whole Gland	10	10	1.9 \pm 0.5	2.0 \pm 0.5	6.2 \pm 1.6	88 \pm 2	94 \pm 3	82 \pm 4	22 \pm 3	5.3 \pm 5.0
Apex			1.9 \pm 0.6	1.9 \pm 0.6	5.6 \pm 1.5	85 \pm 5	90 \pm 5	82 \pm 7	25 \pm 7	1.0 \pm 1.4
Midgland			1.8 \pm 0.4	1.8 \pm 0.4	5.1 \pm 0.8	90 \pm 2	97 \pm 2	84 \pm 4	18 \pm 4	2.2 \pm 1.1
Base			2.2 \pm 0.8	2.3 \pm 0.8	5.9 \pm 1.8	85 \pm 5	93 \pm 4	80 \pm 8	25 \pm 7	2.1 \pm 2.8

Quantitative accuracy of the segmentation methods on non-brachytherapy images using no manually selected anchor points. Mean \pm standard deviation [range] of the error metrics for different regions of interest. N_{pat} and N_{img} are the number of patients and the number of images, respectively. Twenty-three non-brachytherapy images were used for training the algorithm. There are statistically significant differences detected between those metric values in bold and the corresponding metric values in Table IV, where applicable ($p < 0.05$).

Table IV.

Region of Interest	N_{pat}	N_{img}	MAD (mm)	MAD _b (mm)	HD (mm)	DSC (%)	SR (%)	PR (%)	OE (%)	V (cm ³)
Whole Gland	10	10	2.1 \pm 0.3	2.2 \pm 0.3	7.4 \pm 1.7	85 \pm 1	82 \pm 3	89 \pm 3	26 \pm 2	-2.4 \pm 2.0
Apex			2.5 \pm 0.6	2.7 \pm 0.5	6.8 \pm 1.9	78 \pm 5	71 \pm 7	88 \pm 11	36 \pm 7	-1.6 \pm 2.0
Midgland			1.8 \pm 0.4	1.8 \pm 0.4	5.1 \pm 1.1	89 \pm 2	89 \pm 5	90 \pm 5	19 \pm 4	-0.4 \pm 1.3
Base			2.3 \pm 0.7	2.5 \pm 0.7	6.4 \pm 1.6	83 \pm 5	80 \pm 5	87 \pm 8	29 \pm 7	-0.4 \pm 1.3

Table V.

Comparison of the proposed method to previous work where applicable. Mean \pm standard deviation of the error metrics for the whole prostate gland. The MAD mode (unilateral or bilateral) is not defined for those methods in which their MAD values are indicated with asterisks. There are significant differences detected between those metric values in bold and the corresponding metric values of our methods ($p < 0.05$).

Method, Year	N _{pat}	N _{img}	MAD (mm)	MAD (mm)	DSC (%)	SR (%)	PR (%)	OE (%)	Execution Time (s)
Proposed algorithm	10	10	1.9 \pm 0.5	2.0 \pm 0.5	88 \pm 2	94 \pm 3	82 \pm 4	22 \pm 3	22 \pm 2
Ma <i>et al.</i> ¹⁸ , 2017	92	92	-	-	84	-	-	-	-
Ma <i>et al.</i> ¹⁷ , 2016	15	15	-	-	85 \pm 3	83 \pm 1	-	26 \pm 1	-
Shi <i>et al.</i> ¹⁴ , 2016	24	330	1.3 \pm 0.8*	92 \pm 4	90 \pm 5	-	-	-	-
Shi <i>et al.</i> ¹³ , 2015	24	330	1.1 \pm 0.6*	95 \pm 3	93 \pm 4	-	-	-	-
Shao <i>et al.</i> ¹⁶ , 2015	70	70	-	1.9 \pm 0.2	88 \pm 2	84	86	-	-
Shao <i>et al.</i> ¹⁵ , 2014	70	70	-	2.0 \pm 0.8*	85 \pm 6	84	86	-	-
Shi <i>et al.</i> ¹² , 2013	24	330	1.1*	94 \pm 3	92 \pm 4	-	-	-	-
Liao <i>et al.</i> ¹¹ , 2013	24	330	1.0	-	91	[†]	-	-	156
Feng <i>et al.</i> ¹⁰ , 2010	24	306	2.1 \pm 0.8*	89 \pm 5	-	-	-	-	95.5

[†]The overall average was not reported

Quantitative accuracy of the segmentation methods on post-LDR brachytherapy images. Mean \pm standard deviation [range] of the error metrics for different regions of interest. N_{pat} and N_{img} are the number of patients and the number of images, respectively. 23 non-brachytherapy images were used for training the algorithm. The error values in bold show a statistically significant difference with the corresponding error value in Table (p < 0.05).

Table VI.

Region of Interest	N_{pat}	N_{img}	MAD (mm)	MAD _b (mm)	HD (mm)	DSC (%)	SR (%)	PR (%)	OE (%)	V (cm ³)
Whole Gland	37	37	1.9 \pm 0.3	1.9 \pm 0.3	6.3 \pm 1.7	88 \pm 2	92 \pm 3	83 \pm 4	22 \pm 4	3.0 \pm 2.2
Apex			1.7 \pm 0.5	1.7 \pm 0.5	4.9 \pm 1.0	88 \pm 4	91 \pm 4	85 \pm 6	22 \pm 6	0.5 \pm 0.7
Midgland			1.8 \pm 0.5	1.8 \pm 0.5	5.2 \pm 1.0	89 \pm 3	94 \pm 3	85 \pm 6	19 \pm 5	1.2 \pm 1.0
Base			2.3 \pm 0.6	2.3 \pm 0.6	6.2 \pm 1.8	85 \pm 4	91 \pm 6	80 \pm 7	26 \pm 6	1.3 \pm 1.3