Published in final edited form as:

Med Image Comput Assist Interv. 2011 ; 6963: 1–9. doi:10.1007/978-3-642-23944-1_1.

A Learning based Hierarchical Framework for Automatic Prostate Localization in CT Images

Shu Liao and Dinggang Shen

Department of Radiology, Biomedical Research Imaging Center, University of North Carolina at Chapel Hill

Shu Liao: liaoshu.cse@gmail.com; Dinggang Shen: dgshen@med.unc.edu

Abstract

Accurate localization of prostate in CT images plays an important role in image guided radiation therapy. However, it has two main challenges. The first challenge is due to the low contrast in CT images. The other challenge is due to the uncertainty of the existence of bowel gas. In this paper, a learning based hierarchical framework is proposed to address these two challenges. The main contributions of the proposed framework lie in the following aspects: (1) Anatomical features are extracted from input images, and the most salient features at distinctive image regions are selected to localize the prostate. Regions with salient features but irrelevant to prostate localization are also filtered out. (2) An image similarity measure function is explicitly defined and learnt to enforce the consistency between the distance of the learnt features and the underlying prostate alignment. (3) An online learning mechanism is used to adaptively integrate both the inter-patient and patient-specific information to localize the prostate. Based on the learnt image similarity measure function, the planning image of the underlying patient is aligned to the new treatment image for segmentation. The proposed method is evaluated on 163 3D prostate CT images of 10 patients, and promising experimental results are obtained.

1 Introduction

External beam radiation therapy is a major treatment method for prostate cancer and it consists of the planning stage and treatment stage. During the planning stage, the planning image is obtained from the patient. The prostate is then manually delineated in the planning image and a specific treatment plan is designed. In the daily treatment stage, a new treatment image is obtained from the patient and the prostate is localized such that the treatment plan designed in the planning image can be adjusted accordingly and applied to the patient. Therefore, the key to the success of external beam radiation therapy lies in the accurate localization of prostates.

However, accurate localization of prostates in CT images is a challenging task due to the two main issues. The first issue is due to the low image contrast, which is illustrated by Figures 1 (a) and (b). Figure 1 (a) is a slice obtained from a 3D prostate CT image, and the prostate boundary delineated by the clinical expert is shown by the red contour in Figure 1 (b). It can be observed that the contrast between the prostate and its surrounding tissues is low and no obvious boundary can be identified. The other issue is due to the uncertainty of the existence of bowel gas. Figure 1 (c) is a slice obtained from a 3D prostate CT image of

the same patient but from a different treatment day. It can be observed that the existence of bowel gas in Figure 1 (c) changes the image appearance significantly. In order to address these two issues, many novel approaches have been proposed. For instance, profile based features and region based features are used in [1, 2] to localize the prostate, and Davis *et al.* [3] proposed a deformable image registration based prostate segmentation method by combining a deformation algorithm to deal with the bowel gas area.

In this paper, we propose a learning based framework for automatic prostate localization in CT images. The proposed method extracts anatomical features from input images, and adaptively learns the best features at distinctive image regions to localize the prostate. Regions with salient features but irrelevant to localize the prostate such as regions filled with bowel gas are automatically filtered out. An online update mechanism is adopted to adaptively combine both the inter-patient and patient-specific information during the learning process. The alignment is performed in a hierarchical manner by first estimating the translational transformation and then estimating the deformable transformation from the planning image to the treatment image. The proposed method is evaluated on 163 3D prostate CT images of 10 patients; The experimental results show that high segmentation accuracy can be achieved by the proposed method and outperforms the state-of-the-art prostate segmentation algorithms.

2 Methodology

2.1 Preprocessing

In the preprocessing step, we remove the whole-body rigid motion of each patient in the planning and treatment images. More specifically, during this step the planning and treatment images are thresholded such that only the bone structure (i.e., the pelvis, which remains fixed) is visible. Then the bone structure in each treatment image is rigidly aligned to the bone structure in the planning image. After the bone alignment process, the remaining prostate shape differences between the planning and treatment images are mainly caused by the prostate motion.

2.2 Optimal Transformation Estimation for Training Set

After the bone alignment process, we need to obtain the optimal transformation from each training image to the planning image of the underlying patient to learn the similarity measure function in later stages. The training set contains both the inter-patient images as well as the available patient-specific treatment images. We denote each training image as I_i ($i = 1, \dots, N$). Moreover, for each training image, its corresponding segmentation groundtruth S_i ($i = 1, \dots, N$) is also available. We also denote the planning image of the current patient as I^{plan} and its segmentation groundtruth as S^{plan} . The optimal transformation to align each training image to the planning image of the current patient can be obtained by registering the binary images S_i to S^{plan} , and denote the corresponding estimated optimal transformation parameter as θ_i ($i = 1, \dots, N$). In the proposed hierarchical framework, we first estimate the optimal translational transform parameters and then estimate the optimal deformable transform parameters.

2.3 Combination of Inter-Patient and Patient-Specific Information

The contributions of the inter-patient information and patient-specific information in the training stage should be dynamically changed. Initially, when only the planning image of the current patient is available, more weights should be assigned to the inter-patient information. As more treatment images of the current patient are segmented and collected in the training set, the patient-specific information should play an increasing role as intuitively it can capture the patient-specific prostate shape variations more accurately.

In this paper, the weight of the patient-specific information ω^{intra} is determined by a modified sigmoid function based on the current number q of available patient-specific treatment images in the training set. It is defined in Equation 1:

$$\omega^{intra}(q) = \begin{cases} 0, & \text{if } q=0, \\ \frac{1}{1+\exp(-(q-5))}, & \text{if } q>0, \end{cases} \quad (1)$$

The weight for the inter-patient information ω^{inter} can be calculated as $1 - \omega^{intra}$. The determined weights ω^{intra} and ω^{inter} will be used in later steps.

2.4 Feature Extraction and Training Sample Enrichment

In this paper, we use anatomical features as signature for each voxel to characterize the image appearance. More specifically, 14 multi-resolution Haar wavelet feature [4], 9 histogram of gradient (HOG) feature [5] and 30 local binary pattern (LBP) feature [6] are extracted from a small $W \times W$ window centered at each voxel. In this paper, we set $W = 21$. Therefore, each voxel is represented by a 53 (14 + 9 + 30) dimensional feature signature.

In order to ensure the robustness of the learning process, sufficient number of training images should be used. However, on the other hand, to make our method practical in clinical application, the number of training images used should also be minimized. Therefore, in this paper, a simulation strategy in [7] is adopted to enrich the number of training images. Specifically, new training samples can be simulated based on Equation 2:

$$\hat{I}_i^j = T(I_i, \theta_i + \epsilon_i^j), \quad (i=1, \dots, N), (j=1, \dots, C_i) \quad (2)$$

where $\hat{I}_i^j$ denotes the j th simulated image generated from the i th training image I_i . θ_i is the optimal transformation parameter for I_i estimated in Section 2.2, and ϵ_i^j is the j th sample drawn from the zero mean uniform random variable ϵ_i . $T(I_i, \theta_i + \epsilon_i^j)$ denotes applying the corresponding transformation with parameter $\theta_i + \epsilon_i^j$ to image I_i . N is the number of training images available. C_i denotes the number of simulated images generated from I_i . If I_i is an inter-patient training image, then $C_i = \lceil B \cdot \omega^{inter} \rceil$, and otherwise $C_i = \lceil B \cdot \omega^{intra} \rceil$, where B is a constant and set to 40 in this paper. Equation 2 performs small perturbations on each aligned training images. The feature map calculated from each aligned training image can be permuted accordingly since all three types of features are robust against small perturbations.

It should be noted that in the deformable transformation estimation step, θ_i has huge degrees of freedom. To lower the computational cost, we need to reduce the number of degrees of

freedom in θ_i for this step. In this paper, the weighted PCA is applied to the training set based on ω^{inter} and ω^{intra} . Denote the resulting top t eigenvalues as λ_h ($h = 1, \dots, t$), with their corresponding eigenvectors e_h , the original transformation parameter vector θ_i can be approximated as:

$$\theta_i \approx \bar{\theta} + \sum_{h=1}^t c_h^i \sqrt{\lambda_h} e_h, \quad (3)$$

where $\bar{\theta}$ denotes the mean parameter vector of θ_i ($i = 1, \dots, N$), and c_h^i is the coefficient of θ_i projected on the h th eigenvector e_h . In this paper, it is found that $t = 5$ is generally sufficient to accurately approximate θ_i , which can contribute over 80% of the cumulative energy of the original deformable transformation.

2.5 Feature Selection

The next step of our method to select the most informative features at distinctive image regions to localize the prostate. The most straight forward solution is to select the features which have the highest correlation coefficients with the Dice ratio measure based on the training set. However, the Dice ratio is an implicit function with respect to the transformation parameters and has many local maxima. In order to overcome this problem, an explicit similarity measure function which has simpler form than the implicit Dice ratio measure function yet be able to rectify the overlap degrees of the current segmented prostates is defined by Equation 4:

$$F(\theta_i, \varepsilon_i^j) = D(S^{plan}, T(S_i, \theta_i)) \cdot \exp \left(-\frac{1}{2} ((\varepsilon_i^j)^T \Sigma^{-1} \varepsilon_i^j) \right), \quad (4)$$

where the first term $D(S^{plan}, T(S_i, \theta_i))$ denotes the Dice ratio between the prostate in the planning image and the prostate in the aligned training image of I_i . The second term is an exponential function and Σ is a diagonal matrix whose diagonal elements are variances of the transformation parameters learnt from the training images. Thus the defined similarity function is a weighted Gaussian function. For simplicity, we denote the defined similarity function as the Gaussian score function in this paper.

The feature selection process is performed based on the defined Gaussian score function in Equation 4. Note that the feature map calculated for each training image in Section 2.4 encodes two pieces of information: feature value f_i of feature type i ($i = 1, \dots, 53$) and voxel location $\mathbf{x}_j$ ($j = 1, \dots, M$), where M is the number of voxels in the image. Thus the signature of each voxel is a pair $(f_i, \mathbf{x}_j)$. For a given feature type i and a given voxel location $\mathbf{x}_j$, a distance vector can be obtained for $(f_i, \mathbf{x}_j)$ based on all the simulated images:

$$\begin{aligned} V(f_i, \mathbf{x}_j) &= [d_1^1(f_i, \mathbf{x}_j), \dots, d_1^{C_1}(f_i, \mathbf{x}_j), \dots, d_t^s(f_i, \mathbf{x}_j), \dots, d_N^{C_N}(f_i, \mathbf{x}_j)], \text{ where } d_t^s(f_i, \mathbf{x}_j) \\ &= f_i(\hat{I}_t^s(\mathbf{x}_j)) \\ &\quad - f_i(I_t^{plan}(\mathbf{x}_j)) \text{ and } C_k (k=1, \dots, N) \end{aligned} \quad \text{is the}$$

number of simulated images generated from the k th training image defined in Equation 2.

The corresponding Gaussian score vector can be constructed by Equation 5:

$$\mathbf{G} = [F(\theta_1, \epsilon_1^1), \dots, F(\theta_1, \epsilon_1^{C_1}), \dots, F(\theta_t, \epsilon_t^s), \dots, F(\theta_N, \epsilon_N^{C_N})]. \quad (5)$$

Therefore, the most salient features at the most distinctive image regions can be determined by selecting the pairs $(f_i, \mathbf{x}_j)$ with the largest correlation coefficients with $\mathbf{G}$. Figure 2 shows examples of the selected distinctive regions, and it can be observed that regions with salient features but irrelevant to the localization of prostate such as regions filled with bowel gas are automatically filtered out.

2.6 Image Similarity Function Learning

Suppose there are m pairs of $(f, \mathbf{x})$ selected in Section 2.5, denote them as $(\tilde{f}_n, \tilde{\mathbf{x}}_n)$, where $n = 1, \dots, m$. For each simulated training image $\tilde{f}_i^j$, its corresponding feature distance vector is: $\mathbf{U}_i^j = [d_i^j(\tilde{f}_1, \tilde{\mathbf{x}}_1), \dots, d_i^j(\tilde{f}_m, \tilde{\mathbf{x}}_m)]$ and its corresponding Gaussian score $F(\theta_i, \epsilon_i^j)$ is also available. The next step is to learn a image similarity measure function Φ which establishes a mapping from $\mathbf{U}_i^j$ to $F(\theta_i, \epsilon_i^j)$ based on all the simulated training images. In the treatment stage, when a new treatment image comes in, Φ takes the distance vector calculated from the new treatment image and estimates the corresponding Gaussian score to guide the alignment procedure from the treatment image to the planning image. In this paper, Φ is obtained by support vector regression (SVR) with the Gaussian radial basis function (RBF) kernel based on $\mathbf{U}_i^j$ and $F(\theta_i, \epsilon_i^j)$ from all the simulated training images. Grid search is performed to determine the optimal kernel parameters in SVR.

2.7 Predict Optimal Transformation for New Treatment Images

After learning the image similarity measure function Φ in Section 2.6, we can use Φ to estimate the optimal transformation to align the new incoming treatment image I^{new} to the planning image of the current patient. Similar to Section 2.6, we can calculate the feature distance vector from the transformed treatment image $T(I^{new}, \theta)$ with the current transformation parameter θ based on the selected features, denoted it as $\mathbf{U}_\theta^{new}$. Then the corresponding Gaussian score can be calculated as $\Phi(\mathbf{U}_\theta^{new})$; the higher the score, the better the alignment is expected. Therefore, the parameter θ which can maximize $\Phi(\mathbf{U}_\theta^{new})$ is determined as the optimal parameter, denoted as θ_{opt}^{new} . The grid search optimization method is adopted in this paper to estimate θ_{opt}^{new} . After getting θ_{opt}^{new} , we can localize the prostate in the new treatment image by transforming the segmented prostate in the planning image via the inverse transformation based on θ_{opt}^{new} .

3 Experimental Results

In this section, we evaluate the proposed method on a 3D CT prostate dataset consisting of 10 patients. Each patient has more than 10 CT scans. There are 163 images in total. Each image has in-plane resolution of 512×512 , with voxel size $0.98 \times 0.98 \text{ mm}^2$ and the inter-slice voxel size 3 mm . The manual segmentation results provided by clinical experts are also available for each image.

The segmentation accuracy is evaluated based on two quantitative measures: The centroid distance and the Dice ratio between the estimated prostate and the manual segmented prostate. The Dice ratio between two segmented volumes S_1 and S_2 is defined as

$$DSC(S_1, S_2) = \frac{2|S_1 \cap S_2|}{|S_1| + |S_2|}.$$

The box and whisker plots of the centroid distances with and without using the online update mechanism are given in Figure 3. Without using the online updated mechanism means only the inter-patient information is adopted to segment each treatment image of the underlying patient, which is the most challenging case. It can be observed in Figure 3 (a) that even in the most challenging case, the proposed method can still achieve high prostate localization accuracy (i.e., for all the patients, the 25th and 75th percentiles of the centroid distances are within 1 mm error). Moreover, by using the online update mechanism, the centroid distance can be further minimized as illustrated in Figure 3 (b).

The average values and standard deviations of the Dice ratios for each patient are listed in Table 1. It can be observed that the proposed method achieves high segmentation accuracies (i.e., above 85% of the Dice ratio for each patient) *even in the case that only the planning image of the current patient is available to segment each treatment image*, and the segmentation accuracy can be further improved with the online update mechanism. It is also compared with two state-of-the-art prostate segmentation algorithms proposed by Davis *et al.* [3] and Feng *et al.* [8] and the average Dice ratios of different approaches are listed in Table 2. It can be observed that the Dice ratio of the proposed method is significantly higher than Davis's method [3], and the standard deviation of the Dice ratios of the proposed method is lower than Feng's approach [8].

To visualize the segmentation accuracy of the proposed method, Figure 4 shows a typical segmentation result of the proposed method. In Figure 4, the black contours and the white contours are the prostate boundaries determined by the clinical expert and the proposed method respectively. It can be observed that the prostate boundaries estimated by the proposed method are very close to those delineated by the clinical expert even in the cases of the presence of bowel gas, which implies the robustness of the proposed method.

4 Conclusion

In this paper, we propose a new learning based framework for accurate localization of prostate in CT images. In the proposed framework, each input image is represented by anatomical feature maps, and a feature selection process is performed to determine the most salient features at distinctive image regions to localize the prostate. The underlying feature selection process also automatically filters the regions with salient features but irrelevant to the localization of prostates such as regions filled with bowel gas. An explicit image similarity measure is defined and learnt to avoid the local maxima problem. An online update mechanism is also adopted to adaptively combine the inter-patient information and patient-specific information in the learning stage. It is demonstrated that the proposed method can achieve high prostate segmentation accuracies in CT images and outperforms state-of-the-art prostate segmentation algorithms. Most importantly, the proposed method is

highly exible in clinical application as high segmentation accuracies can be achieved even in the case that only the planning image of the underlying patient is available.

References

1. Cosio F. Automatic initialization of an active shape model of the prostate. *Medical Image Analysis*. 2008; 12:469–483. [PubMed: 18346931]
2. Stough J, Broadhurst R, Pizer S, Chaney E. Regional appearance in deformable model segmentation. *IPMI*. 2007:532–543.
3. Davis B, Foskey M, Rosenman J, Goyal L, Chang S, Joshi S. Automatic segmentation of intra-treatment ct images for adaptive radiation therapy of the prostate. *MICCAI*. 2005:442–450. [PubMed: 16685876]
4. Mallat S. A theory for multiresolution signal decomposition: The wavelet representation. *PAMI*. 1989; 11:674–693.
5. Dalal N, Triggs B. Histograms of oriented gradients for human detection. *CVPR*. 2005:886–893.
6. Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *PAMI*. 2002; 24:971–987.
7. Zhou L, Liao S, Shen D. Learning-based prostate localization for image guided radiation therapy. *ISBI*. 2011:2103–2106.
8. Feng Q, Foskey M, Chen W, Shen D. Segmenting ct prostate images using population and patient-specific statistics for radiotherapy. *Medical Physics*. 2010; 37

Fig. 1.

(a) An image slice obtained from a 3D prostate CT image. (b) The prostate boundary highlighted by the red manually-delineated contour superimposed on the image in (a). (c) An image slice with bowel gas obtained from a different 3D prostate CT image of the same patient as in (a).

Fig. 2.

Examples showing some selected distinctive regions. The selected regions are highlighted in green. The prostate boundary determined by the clinical expert is delineated in red.

Fig. 3.

Centroid distance between the estimated prostate volume and the manually segmented prostate volume (a) without using the online update mechanism, and (b) with the online update mechanism. The horizontal lines in each box represent the 25th percentile, median, and 75th percentile respectively. The whiskers extend to the most extreme data points.

Fig. 4.

typical performance of the proposed method on six slices of the 14th treatment image of the 8th patient. The Dice ratio of this treatment image is 90.42%. The black contours show the results of manual segmentation, and the white contours show the results obtained by the proposed method.

Table 1

Mean and standard deviations of the Dice ratios of each patient. PA i denotes the i th patient. BA denotes before alignment. WU and WOU denote the proposed method with and without using the online update mechanism respectively. The highest value of the Dice ratio is bolded.

	PA 1	PA 2	PA 3	PA 4	PA 5
BA	0.64±0.06	0.46±0.13	0.71±0.08	0.69±0.07	0.40±0.05
WOU	0.88±0.01	0.87±0.01	0.88±0.02	0.85±0.07	0.88±0.02
WU	0.89 ± 0.02	0.89 ± 0.01	0.90 ± 0.01	0.87 ± 0.08	0.90 ± 0.02

	PA 6	PA 7	PA 8	PA 9	PA 10
BA	0.78±0.10	0.79±0.02	0.80±0.07	0.47±0.10	0.73±0.08
WOU	0.85±0.01	0.86±0.02	0.87±0.03	0.86±0.02	0.88±0.01
WU	0.88 ± 0.01	0.88 ± 0.02	0.90 ± 0.01	0.89 ± 0.02	0.89 ± 0.01

Table 2

Comparisons of the proposed method with two state-of-the-art deformable model based prostate segmentation algorithms in CT images based on the Dice ratios. WU denotes the proposed framework using the online update mechanism.

Methods	Mean	Median	Std. Dev.
Davis [3]	0.82	0.84	0.06
Feng [8]	0.89	0.90	0.05
WU	0.89	0.90	0.02