HHS Public Access

Author manuscript

Neurocomputing. Author manuscript; available in PMC 2016 January 15.

Published in final edited form as:

Neurocomputing. 2016 January 15; 173(2): 317–331. doi:10.1016/j.neucom.2014.11.098.

A Learning-Based CT Prostate Segmentation Method via Joint Transductive Feature Selection and Regression

Yinghuan Shi[a,b], **Yaozong Gao**[b], **Shu Liao**[b], **Daoqiang Zhang**[b], **Yang Gao**[a], and **Dinggang Shen**[b]

[a]State Key Laboratory for Novel Software Technology, Nanjing University, China

[b]Department of Radiology and BRIC, UNC Chapel Hill, U.S.

Abstract

In[1] recent years, there has been a great interest in prostate segmentation, which is a important and challenging task for CT image guided radiotherapy. In this paper, a learning-based segmentation method via joint transductive feature selection and transductive regression is presented, which incorporates the physician's simple manual specification (only taking a few seconds), to aid accurate segmentation, especially for the case with large irregular prostate motion. More specifically, for the current treatment image, experienced physician is first allowed to manually assign the labels for a small subset of prostate and non-prostate voxels, especially in the first and last slices of the prostate regions. Then, the proposed method follows the two step: in prostate-likelihood estimation step, two novel algorithms: tLasso and wLapRLS, will be sequentially employed for transductive feature selection and transductive regression, respectively, aiming to generate the prostate-likelihood map. In multi-atlases based label fusion step, the final segmentation result will be obtained according to the corresponding prostate-likelihood map and the previous images of the same patient. The proposed method has been substantially evaluated on a real prostate CT dataset including 24 patients with 330 CT images, and compared with several state-of-the-art methods. Experimental results show that the proposed method outperforms the state-of-the-arts in terms of higher Dice ratio, higher true positive fraction, and lower centroid distances. Also, the results demonstrate that simple manual specification can help improve the segmentation performance, which is clinically feasible in real practice.

1. Introduction

Prostate cancer is currently becoming one of the most leading reasons for males' cancer-caused death worldwide. According to the report provided by the National Cancer Institute[2], prostate cancer is estimated to cause 233,000 new cases and 29,480 deaths for U.S. males in 2014. Nowadays, for prostate cancer treatment, the CT image guided radiotherapy has received many research interests, since it is able to guide the better delivery of radiation to the prostate [34].

Correspondence to: Yang Gao; Dinggang Shen.

[2]Prostate Cancer, report from National Cancer Institute. http://www.cancer.gov/cancertopics/types/prostate

For a patient, during his/her CT image guided radiotherapy, a sequence of CT scans will be acquired in several days (typically 20–40), including the planning day and the treatment days. A CT scan acquired in the planning day is called as the planning image, and the scans acquired in the subsequent treatment days are called as the treatment images. In each treatment day, the high energy X-rays will be delivered from different directions to kill the cancer tissues. Thus, the key problem is becoming how to accurately determine the location of prostate in CT images acquired from each treatment day. Normally, during different treatment days, it is usual that the locations of prostate vary in different CT scans due to the prostate motion and other factors (e.g., human body motion). Currently, determining the location of prostate is often done by the physician with slice-by-slice manual segmentation. However, manual segmentation, which can take up to 20 minutes for each treatment image, is very time-consuming, especially when lots of new CT scans come. Most importantly, manual segmentation results could be inconsistent across different treatment days due to inter- and intra- operator variability.

The major challenging issues for accurate prostate segmentation in the C-T images include: (i) the boundary between prostate region and background (non-prostate) region is usually unclear due to the low contrast in the CT images, e.g., in Fig.1(a) and Fig.1(b), the same CT image is without and with the red contour manually delineated by the physician to indicate the prostate region, respectively. (ii) The locations of the prostate regions scanned from same patient at different treatment days are often different due to the irregular and unpredictable prostate motion, e.g., in Fig.1(c) where the cyan and magenta contours denote the manual segmentations of the two bone-aligned CT images scanned from two different treatment days from the same patient. We can observe the large prostate motion even after aligning the two scans based on their bony structures. This indicates the large prostate motion relative to the bones. (iii) The prostate shape appearance varies at different treatment days, e.g., in Fig.1(d) where the cyan and magenta contours denote the manual segmentations of two bone-aligned CT images from two different treatment days from the same patient, which indicates the large shape appearance change.

In recent years, several computer-aided prostate segmentation methods [5][19][34][35][14][16][32] have been developed and achieved much success for CT image guided radiotherapy, with the common goal of segmenting prostate in the current treatment image by borrowing the knowledge learned from the planning and previous treatment images. Unfortunately, previous methods, during segmenting the prostate in the current treatment day, often has limited prior knowledge about the prostate region, e.g., the magnitude of prostate motion. Thus, it is quite possible that when large irregular motion occurs within the prostate region, previous methods might be difficult to achieve good segmentation results.

According to our experimental observation, the domain specific prior knowledge, which is usually given by the experienced physician (only taking a few seconds), is extremely helpful to guide segmentation. The specific prior knowledge includes the first and last slices of prostate regions, and confusing voxels which lie on boundary between prostate and background. The first and last slices are given to help dealing with large irregular prostate motion. The confusing voxels are assigned the labels since in some cases prostate and background voxels are quite similar.

Therefore, we want to systematically answer the question - can we use the physician's simple manual specification in the current treatment image (only taking a few seconds) to aid better segmentation? - in this paper. To this end, we present a novel learning-based method for prostate segmentation, which aims to improve the segmentation results by using additional physician's manual specified information. Specifically, when a new treatment image comes, the physician is allowed to first give some manual specification information on (i) the first and last slices of prostate regions and (ii) some voxels which lie on boundary between prostate and background if they are difficult to distinguish according to the physician's consideration. Meanwhile, the planning image and treatment images obtained in the previous treatment days from the same patient can also be used to guide accurate segmentation, with their segmentation ground truth often available. Here, physician's labeled voxels in the current treatment image are *labeled voxels*, voxels in the planning and previous treatment images are *auxiliary voxels*, and the remaining voxels in the current treatment image are *unlabeled voxels*. Our goal for segmenting prostate region is to predict the labels (prostate or background) for the unlabeled voxels, with incorporating the labeled voxels as well as the auxiliary voxels (see Fig.2 for the difference from previous supervised learning-based methods).

1.1. Framework of the Proposed Method

Main steps of the proposed method can be summarized as below: Firstly, to eliminate the whole-body patient motion which is irrelevant to the prostate segmentation, all previous and current treatment images are rigidly aligned to the planning image of the same patient based on their pelvic bone structures. Then, the ROI (Region of Interest) region is extracted according to the prostate mass center in the previous images. For one slice in every 3–5 slices within the current treatment image, the physician is allowed to label a small subset of prostate and background voxels, especially the ones which are difficult to distinguish and the ones belonging to the first and last slices of the prostate region. We illustrate typical examples of physician's labeling in Fig.3.

Secondly, combining with voxels sampled from the planning and previous treatment images, 2-D low-level features are extracted. To make full use of the labeled, auxiliary, and unlabeled voxels, for each slice, we mathematically model the prostate-likelihood estimation problem as a joint transductive feature selection and transductive regression problem. tLasso (Transductive Lasso) is first applied to select the most discriminative features slice-byslice, and wLapRLS (Weighted Laplacian regularized least squares) is further adopted to predict the 2-D prostate-likelihood map for unlabeled voxels in each slice. The predicted 2-D prostate-likelihood map of each individual slice will be merged into a 3-D prostate-likelihood map according to the order of original slices.

Finally, to use the patient-specific prostate shape information, manually segmented prostate regions in the planning and previous treatment images of the same patient will be rigidly aligned to the estimated 3-D prostate-likelihood map of the current treatment image. We then majority vote the multiple aligned results to obtain the final segmentation result. The framework of the proposed method can be referred to Fig.4.

1.2. Contributions

Our contributions are summarized into the following triple folds:

- We present a novel learning-based 3-D prostate segmentation method for CT images. Our aim is to use the physician's simple manual specification (only taking a few seconds) to guide accurate segmentation especially when large irregular prostate motion occurs in CT images.
- A transductive feature selection method, tLasso, is proposed, which mathematically extends traditional supervised Lasso [46][18] to a transductive manner. tLasso selects the most discriminative features in each slice of the current treatment image by taking not only the labeled and auxiliary voxels but also the unlabeled voxels into account, through the Laplacian regularizer.
- A prostate-likelihood predictor, wLapRLS, is proposed. wLapRLS extends LapRLS [2] by adding the prior label constraint of unlabeled voxels to the empirical loss term, whose advantage is that the results can be benefited from auxiliary voxels [53].

The rest of this paper is organized as follows: Section 2 introduces the related works. Section 3 discusses the image preprocessing used in this paper. Section 4 presents the technical details of the prostate-likelihood estimation, which mainly consist of the proposed tLasso and wLapRLS. Section 5 discusses the multi-atlases based label fusion, and Section 6 reports the experimental results. Finally, the conclusion is given in Section 7.

2. Related Works

According to the different prostate segmentation mechanisms in CT images, the previous methods can be roughly categorized into three classes: registration-based, deformable-model-based, and learning-based methods.

The registration-based methods [6][10][34][35] first warp the current treatment image onto the planning and previous treatment images. Then, their respective segmentation images are similarly warped and further combined to obtain the final segmentation result of the current treatment image. Chen et al., [6] designed a strategy that aligns the planning image to treatment image by mesh-less point set modeling and 3D non-rigid registration. Davis et al., [10] segmented the prostate by combining large deformation image registration with a bowel gas segmentation and deflation algorithm. Liao et al., [34] proposed to use anatomical feature selection and an online updating mechanism to integrate both population and patient-specific information, in order to guide the accurate prostate registration. Also, Liao et al., [35] combined the patch-based feature representation and hierarchical sparse label propagation in order to automatically localize the prostate. Several experimental results have demonstrated the robustness and effectiveness of the registration-based methods for prostate segmentation. However, the segmentation accuracy is limited in the case with inconsistent image appearance changes in CT images [19].

The deformable-model-based methods [5][14] first initialize the deformable model by using the the prostate shapes learned from the planning and previous treatment images, and then

employ the specific optimization strategies to guide prostate segmentation. Chen et al. [5] simultaneously segmented the prostate and the rectum by imposing the anatomical constraints during the model deformation procedure. Feng et al. [14] segmented the prostate by using the deformable model, which integrates the gradient profile features and the probability distribution function features. The deformable-model-based methods can combine the information from both the prostate shape and image appearance for guiding prostate segmentation. However, the good performance largely requires the good quality of the deformable model initialization. In the case where the large prostate motion happens, obtaining a good initialization for deformable model becomes a very challenging task.

The learning-based methods [32][19], which attracted lots of interests recently, first use the machine learning techniques to predict the prostate-likelihood map, and then segment the prostate in the estimated prostate-likelihood map. In Li et al.'s method [32] and Gao et al.'s [19], prostate segmentation is first formulated as a prostate-likelihood estimation problem (AdaBoost [17] in [32], sparse representation-based classification [49] in [19]) using visual features (e.g., Histogram of oriented Gradients (HoG) [9] and auto-context features [47]), and then on the obtained likelihood map, the off-the-shelf segmentation techniques (e.g., level-set) are adopted to segment the prostate. Our proposed method belongs to this class of learning-based segmentation methods.

Beyond the above three major classes, there are other methods [6][21][44] proposed for prostate segmentation in CT images. Note that, besides segmenting prostate from CT images, other segmentation methods are also developed for segmenting prostate from other imaging modalities such as Magnetic Resonance (MR) [20][30][36][37] and ultrasound [43] [54][55] images. However, they cannot be directly borrowed to segment prostate from CT images due to the aforementioned challenges.

3. Image Preprocessing

We here will first introduce the notations used in the following parts, and present the implementation details of image preprocessing, including the image pre-alignment, the ROI extraction, the manual interaction, the feature representation, as well as the training voxels sampling strategy.

3.1. Notation

In this paper, the matrices (2-D and 3-D) are denoted by bold upper case letters, the vectors are denoted by bold lower case ones, and the scalars are denoted by lower case ones.

As the CT image guided radiotherapy, for each patient, we have one planning image, several treatment images in previous treatment days, which are manually segmented by experienced physician off-line, as well as the treatment image in the current treatment day, which need to be automatically segmented. The planning image and its corresponding manual segmentation result are denoted as $\mathbf{I_p}$ and $\mathbf{G_p}$, respectively. The n^{th} treatment image, which is the current treatment image, is denoted as $\mathbf{I}_n$. The previous treatment images and their corresponding manual segmented results are denoted as $\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_{n-1}$ and $\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_{n-1}$, respectively. Also, the final 3-D prostate-likelihood map and its segmentation result

by applying our method for the current treatment image $\mathbf{I}_n$ are denoted as $\mathbf{M}_n$ and $\mathbf{S}_n$, respectively.

3.2. Aligning All Treatment Images to Planning Image

It is often inevitable that the whole-body rigid motion occurs from the patients during the scanning process across different treatment days, which is irrelevant to the prostate motion. To reduce the influence caused by the whole-body rigid motion, which is irrelevant to the prostate motion, the pelvic bone structure is first extracted from each image using threshold-based segmentation, and each treatment image ($\mathbf{I}_1$, $\mathbf{I}_2$, …, $\mathbf{I}_n$) is then rigidly aligned to the planning image ($\mathbf{I}_\mathbf{p}$) based on their corresponding pelvic bone structures.

3.3. ROI Extraction

Basically, the prostates are located only in a small central part of the CT images. So the ROI extraction, which aims to extract the central part by excluding the irrelevant and redundant background voxels, is useful to alleviate the computational burden and also important for improving the final segmentation accuracy. For each patient, the mass centers of the prostate in the planning and previous treatment images are first calculated, respectively. Then, we average the calculated mass centers as the patient-specific prostate center. Finally, we extract a 3-D region centered at the patient-specific prostate center. Specifically, the extracted ROI size is 180 × 180 × 60, which is large enough to cover the whole prostate volume (i.e., the normal size of the prostate is usually from 30 × 40 × 30 mm^3 to 40 × 50 × 70 mm^3).

3.4. Physician's Simple Manual Interaction

When asking physician to label the current treatment image, we will not ask him/her to label several prostate/background voxels in each slice, which is a time-consuming task. Instead, we just ask physician to label a small amount of prostate and background voxels in the first and last slices within the prostate region, as well as one slice in every 3–5 slices within the central part of prostate region. The typical examples of the physician’s labeling are shown in Fig.3. It is noteworthy that the physician’s labeling is done after ROI extraction, so the physician is only required to label a small amount of slices in CT images.

For the slices which are not labeled by the physician, they will use the labeled voxels from the nearest slices which are labeled by the physician, since neighboring slices have similar features. An example in z-axis is illustrated in Fig.5.

3.5. Feature Extraction

We extracted three different kinds of features from 2-D slice, which include 9-dimensional histogram of oriented gradient (HoG) features [9], 30-dimensional local binary pattern (LBP) features [39][45][28] and 14-dimensional multi-resolution Haar wavelet features [38]. The window size to extract features is empirically set to 21 × 21. Specifically, HOG features are calculated within 3 × 3 cell blocks with 9 histogram bins similar to [9]. LBP features are calculated by setting the radius value as 2 and the neighboring voxel number as 8. Haar features are calculated by convolving the 14 multi-resolution wavelet basis functions with the input image similar to [38].

4. Prostate-Likelihood Estimation

Here, we first mathematically formulate the prostate-likelihood estimation task as a joint transductive feature selection and transductive regression problem, by sequentially discussing the details of the proposed tLasso and wLapRLS, respectively. Finally, we present the mechanism to generate the 3-D prostate-likelihood map.

4.1. Transductive Feature Selection by tLasso

Feature selection has been demonstrated its efficacy for learning-based prostate segmentation[32][19], since different patients, even the same patient with different slices, prefer choosing different features due to the large appearance variation across different prostate regions. Unlike previous learning-based segmentation methods[32][19] (only using the training voxels to guide supervised feature selection), in our task, the information coming from both the large amount of unlabeled voxels (with manifold assumption), and the labeled voxels (provided by the physician), along with the auxiliary voxels with the known label, can be incorporated for jointly transductive feature selection.

For the current slice, we use n^A, n^L, n^U to denote the numbers of the auxiliary, labeled and unlabeled voxels, respectively. $\mathbf{F}^A \in \mathbb{R}^{n^A \times d}$, $\mathbf{F}^L \in \mathbb{R}^{n^L \times d}$, and $\mathbf{F}^U \in \mathbb{R}^{n^U \times d}$ are the feature matrix for the auxiliary, labeled, and unlabeled voxels, respectively. Thus, $\mathbf{F} \in \mathbb{R}^{n \times d}$ denotes the row-wise combined feature matrix for the three types of voxels[3]. d is the dimension of feature space before feature selection. $\mathbf{y}^A \in \mathbb{R}^{n^A}$, $\mathbf{y}^L \in \mathbb{R}^{n^L}$, and $\mathbf{y}^U \in \mathbb{R}^{n^U}$ are the labels for the auxiliary, labeled, and unlabeled voxels, respectively. Formally, the proposed tLasso aims to (1) minimize the reconstruction error $\mathscr{R}_{\text{aux}}$ between the auxiliary voxels and the corresponding label, (2) minimize the reconstruction error $\mathscr{R}_{\text{lab}}$ between the labeled voxels and the corresponding label, (3) assume that all the auxiliary, labeled and unlabeled voxels satisfy the manifold constraint ($\mathscr{R}_{\text{lap}}$), and (4) impose the sparsity constraint to select the most significant features ($\mathscr{R}_{\text{spa}}$). Thus, we can obtain the following objective function:

$$\min(\mathscr{R}_{\text{aux}}+\mathscr{R}_{\text{lab}}+\gamma_G\mathscr{R}_{\text{lap}}+\gamma_S\mathscr{R}_{\text{spa}}), \quad (1)$$

where $\gamma_G \in \mathbb{R}$ and $\gamma_S \in \mathbb{R}$ are the two parameters to control the weights for the last two terms, respectively.

Specifically, the reconstruction error of $\mathscr{R}_{\text{aux}}$ and $\mathscr{R}_{\text{lab}}$ can be respectively written as follows:

$$\mathscr{R}_{\text{aux}}=\sum_{i=1}^{n^A}\left(\frac{y_i^A-\mathbf{x}_i^{A\top}\boldsymbol{\beta}}{n^A}\right)^2=\left\|\frac{\mathbf{y}^A-\mathbf{F}^A\boldsymbol{\beta}}{n^A}\right\|_2^2, \quad (2)$$

and

[3] $n = n^A + n^L + n^U$

$$\mathscr{R}_{\text{lab}}=\sum_{i=1}^{n^L}\left(\frac{y_i^L-\mathbf{x}_i^{L\top}\boldsymbol{\beta}}{n^L}\right)^2=\left\|\frac{\mathbf{y}^L-\mathbf{F}^L\boldsymbol{\beta}}{n^L}\right\|_2^2, \quad (3)$$

where $\boldsymbol{\beta}\in\mathbb{R}^d$ denotes the feature weight vector (the higher weight indicates the more significant feature), which is the parameter to be optimized in tLasso. $\mathbf{x}_i^A\in\mathbb{R}^d$ and $\mathbf{x}_i^L\in\mathbb{R}^d$ are the i^{th} entries of feature matrix $\mathbf{F}^A$ and $\mathbf{F}^L$, respectively. $y_i^A\in\mathbb{R}$ and $y_i^L\in\mathbb{R}$ are the i^{th} entries of labels $\mathbf{y}^A$ and $\mathbf{y}^L$, respectively.

Also, the Laplacian regularization term can be formally written as:

$$\mathscr{R}_{\text{lap}}=\frac{1}{n^2}\boldsymbol{\beta}^\top\mathbf{F}^\top\mathbf{L}\mathbf{F}\boldsymbol{\beta}, \quad (4)$$

where where $\mathbf{L}\in\mathbb{R}^{n\times n}$ is graph Laplacian matrix defined as the same in literature [2][52] [12]. Lastly, the sparsity regularization term is $\mathscr{R}_{\text{spa}}=\|\boldsymbol{\beta}\|_1$.

Mathematically, the two reconstruction terms $\mathscr{R}_{\text{aux}}$ and $\mathscr{R}_{\text{lab}}$ can be merged as the follows:

$$\mathscr{R}_{\text{aux}}+\mathscr{R}_{\text{lab}}=\|\mathbf{J}(\mathbf{y}-\mathbf{F}\boldsymbol{\beta})\|_2^2, \quad (5)$$

where $\mathbf{y}\in\mathbb{R}^n$ is the label for all the three types of voxels ($\mathbf{y}=[\mathbf{y}_L\ \mathbf{y}_A\ \mathbf{y}_U]^\top$). $\mathbf{J}\in\mathbb{R}^{n\times n}$ is a diagonal matrix used to indicate the voxels which are labeled (including the labeled and auxiliary voxels) since unlabeled voxels have no contribution on the first two term, and $\mathbf{J}$ is formally defined as

$$\mathbf{J}=\text{diag}[\overbrace{\frac{1}{n^A},\frac{1}{n^A},\ldots,\frac{1}{n^A}}^{n^A},\overbrace{\frac{1}{n^L},\frac{1}{n^L},\ldots,\frac{1}{n^L}}^{n^L},\overbrace{0,0,\ldots,0}^{n^U}] \quad (6)$$

Therefore, the objective function of tLasso can be formulated as follows:

$$\min_{\boldsymbol{\beta}}\left(\|\mathbf{J}(\mathbf{y}-\mathbf{F}\boldsymbol{\beta})\|_2^2+\frac{\gamma_G}{n^2}\boldsymbol{\beta}^\top\mathbf{F}^\top\mathbf{L}\mathbf{F}\boldsymbol{\beta}+\gamma_S\|\boldsymbol{\beta}\|_1\right). \quad (7)$$

Eq.(7) can be solved by using many existing convex optimization approaches, e.g., FISTA [31]. In this paper, we solve Eq.(7) by using the CVX toolbox [25], which is a package for specifying and solving convex programs. After obtaining $\boldsymbol{\beta}$, we can select the corresponding d' ($d'<d$) features with respect to the non-zero entries in $\boldsymbol{\beta}$. Also, we can generate a new feature matrix $\mathbf{F}'\in\mathbb{R}^{n\times d'}$, by using the selected d' features, to replace $\mathbf{F}$.

When using FISTA for solving tLasso, it has $O(n^2d+nd^2)$ in the initialization, and $O(kd^2)$ in the iteration, where k is the total iteration number. Thus, tLasso has $O(n^2d+nd^2+kd^2)$ computation complexity. Also, FISTA is proven to converge in function value as $O(1/k^2)$, thus in practice, $k<<n$. Therefore, the total computation complexity of tLasso is $O(n^2d+nd^2)$.

4.2. Prostate-Likelihood Prediction by wLapRLS

For each unlabeled voxels with the selected features in the current slice, the prostate-likelihood prediction problem is mathematically modeled as the transductive regression task, in which the auxiliary voxels (sampled from the planning and previous treatment images), the labeled voxels (specificated by the physician), and the unlabeled voxels will be integrated for the regression (different from the previous supervised learning-based prostate segmentation methods).

For the current slice, we denote $\mathbf{K}^A \in \mathbb{R}^{n^A \times n^A}$, $\mathbf{K}^L \in \mathbb{R}^{n^L \times n^L}$, and $\mathbf{K}^U \in \mathbb{R}^{n^U \times n^U}$ as the pair-wise similarity matrix for the auxiliary, labeled, and unlabeled voxels after feature selection by tLasso, respectively. Thus, we can further denote $\mathbf{K} \in \mathbb{R}^{n \times n}$ as the pair-wise similarity matrix for all the three types of voxels.

Formally, the wLapRLS aims to (1) minimize the empirical loss among the auxiliary ($\overline{\mathscr{R}_{\mathrm{aux}}}$), labeled ($\overline{\mathscr{R}_{\mathrm{lab}}}$), and even unlabeled voxels ($\overline{\mathscr{R}_{\mathrm{ulab}}}$) with their corresponding labels, (2) assume that all the auxiliary, labeled and unlabeled voxels satisfy the manifold constraint ($\overline{\mathscr{R}_{\mathrm{lap}}}$), and (3) impose the smoothness constraint ($\overline{\mathscr{R}_{\mathrm{smo}}}$) to penalize the possible solutions that might lead over-fitting. Thus, we can obtain the following objective function:

$$\min(\overline{\mathscr{R}_{\mathrm{aux}}}+\overline{\mathscr{R}_{\mathrm{lab}}}+\overline{\mathscr{R}_{\mathrm{ulab}}}+\gamma_G\overline{\mathscr{R}_{\mathrm{lap}}}+\gamma_C\overline{\mathscr{R}_{\mathrm{smo}}}), \quad (8)$$

where $\gamma_G \in \mathbb{R}$ and $\gamma_C \in \mathbb{R}$ are the two parameters to control the weights for the last two terms, respectively.

Mathematically, the empirical loss $\overline{\mathscr{R}_{\mathrm{aux}}}$ and $\overline{\mathscr{R}_{\mathrm{lab}}}$ can be respectively written as follows:

$$\overline{\mathscr{R}_{\mathrm{aux}}}=\left\|\frac{\mathbf{y}^A-\mathbf{K}^A\boldsymbol{\omega}^A}{n^A}\right\|_2^2, \quad (9)$$

and

$$\overline{\mathscr{R}_{\mathrm{lab}}}=\left\|\frac{\mathbf{y}^L-\mathbf{K}^L\boldsymbol{\omega}^L}{n^L}\right\|_2^2. \quad (10)$$

Also, we introduce the pseudo-label $\mathbf{y}^U \in \mathbb{R}^{n^U}$ for the unlabeled voxels, thus the empirical loss $\overline{\mathscr{R}_{\mathrm{ulab}}}$ can be defined as:

$$\overline{\mathscr{R}_{\mathrm{ulab}}}=\left\|\frac{\mathbf{y}^U-\mathbf{K}^U\boldsymbol{\omega}^U}{n^U}\right\|_2^2, \quad (11)$$

where $\boldsymbol{\omega}^A \in \mathbb{R}^{n^A}$, $\boldsymbol{\omega}^L \in \mathbb{R}^{n^L}$, $\boldsymbol{\omega}^U \in \mathbb{R}^{n^U}$ denote the corresponding prediction vectors of three types of voxels, which are the parameters to be optimized in wLapRLS. Note that the pseudo-label $\mathbf{y}^U$ is the classification result by using the classifier trained on all the auxiliary voxels. Here, Support Vector Machine (SVM) is adopted to train the classifier with selected

features by tLasso. The advantage of using pseudo-label $\mathbf{y}^U$ is that the results can be benefited from auxiliary voxels [53].

Moreover, as for the three parts of the empirical loss, we can use the prediction vector $\boldsymbol{\omega} \in \mathbb{R}^n$ to replace the $\boldsymbol{\omega}^A$, $\boldsymbol{\omega}^L$, $\boldsymbol{\omega}^U$, and formally integrate them as follows:

$$\overline{\mathscr{R}_{\text{aux}}}+\overline{\mathscr{R}_{\text{lab}}}+\overline{\mathscr{R}_{\text{ulab}}}=\|\mathbf{J}'(\mathbf{y}-\mathbf{K}\boldsymbol{\omega})\|_2^2, \quad (12)$$

where $\mathbf{J}' \in \mathbb{R}^{n\times n}$ is a diagonal matrix used to indicate the three types of voxels as

$$\mathbf{J}'=\text{diag}[\overbrace{\frac{1}{n^A},\frac{1}{n^A},\ldots,\frac{1}{n^A}}^{n^A},\overbrace{\frac{1}{n^L},\frac{1}{n^L},\ldots,\frac{1}{n^L}}^{n^L},\overbrace{\frac{1}{n^U},\frac{1}{n^U},\ldots,\frac{1}{n^U}}^{n^U}] \quad (13)$$

Furthermore, the Laplacian regularization term can be formally written as:

$$\overline{\mathscr{R}_{\text{lap}}}=\frac{1}{n^2}\boldsymbol{\omega}^\top\mathbf{K}\mathbf{L}'\mathbf{K}\boldsymbol{\omega}, \quad (14)$$

where $\mathbf{L}' \in \mathbb{R}^{n\times n}$ is graph Laplacian matrix calculated on the selected features. Also, the smoothness constraint regularization term can be written as $\overline{\mathscr{R}_{\text{smo}}}=\boldsymbol{\omega}^\top\mathbf{K}\boldsymbol{\omega}$, aiming to prevent the possible over-fitting issue [2].

Therefore, the objective function of wLapRLS can be formulated as follows:

$$\min_{\boldsymbol{\omega}}\left(\|\mathbf{J}'(\mathbf{y}'-\mathbf{K}\boldsymbol{\omega})\|_2^2+\gamma_C\boldsymbol{\omega}^\top\mathbf{K}\boldsymbol{\omega}+\frac{\gamma_G}{n^2}\boldsymbol{\omega}^\top\mathbf{K}\mathbf{L}'\mathbf{K}\boldsymbol{\omega}\right). \quad (15)$$

The $\boldsymbol{\omega}$ can be computed as a close-form solution as follows:

$$\boldsymbol{\omega}=[\mathbf{J}'\mathbf{J}'\mathbf{K}+\gamma_C\mathbf{I}+\frac{\gamma_G}{n^2}\mathbf{L}'\mathbf{K}]^{-1}\mathbf{J}'\mathbf{J}'\mathbf{y}', \quad (16)$$

where $\mathbf{I}$ is the identity matrix. Finally, by using the Representer Theorem [42], the prostate-likelihood of the unlabeled voxels in the current slice of treatment image can be calculated as the corresponding part (from $n^A + n^L + 1$ to n) of vector $\mathbf{K}_{\boldsymbol{\omega}}$. Since in some cases, obtained the prostate-likelihood might be negative or noisy, we normalize all the the prostate-likelihood into [0, 1]. Thus, 2-D prostate-likelihood map for each slice in the current treatment image can be generated with the calculated prostate-likelihood of the unlabeled voxels.

4.3. Generating the 3-D Prostate-Likelihood Map

For all slices in the current treatment image, the obtained 2-D predicted likelihood maps will be merged together to generate the 3-D prostate-likelihood map according to the order of original slices. In [32], the authors found that the combination of segmentation results along two directions (z-axis and y-axis) is helpful to improve the performance. Therefore, we generate 3-D predicted prostate-likelihood map $\mathbf{M}_n$ by averaging $\mathbf{M}_{\mathscr{X},n}$ and $\mathbf{M}_{\mathscr{Y},n}$, where

$\mathbf{M}_{\mathscr{L},n}$ and $\mathbf{M}_{\mathscr{Y},n}$ are the 3-D prostate-likelihood maps generated along z-axis and y-axis, respectively. Without loss of generality, the details of generating $\mathbf{M}_{\mathscr{L},n}$ are summarized in Algorithm.1. Since $\mathbf{M}_{\mathscr{L},n}$ is obtained from the slices cut from z-axis, similarly $\mathbf{M}_{\mathscr{Y},n}$ can also be obtained from Algorithm.1 by cutting the slices along y-axis.

Algorithm 1

Generating 3-D prostate-likelihood map $\mathbf{M}_{\mathscr{L},n}$

Input: Planning image $\mathbf{I_p}$ and ground truth $\mathbf{G_p}$, previous treatment images $\mathbf{I}_1,\ldots,\mathbf{I}_{n-1}$ and ground truth $\mathbf{G}_1,\ldots,\mathbf{G}_{n-1}$ and current treatment image $\mathbf{I}_n$

Output: 3-D prostate-likelihood maps $\mathrm{M}_{\mathscr{L},n}$

for the s^{th} slice **do**
  $(\mathbf{F}^A, \mathbf{y}^A) \leftarrow$ Sampling the auxiliary voxels.
  if the s^{th} slice is labeled by physician **then**
    $(\mathbf{F}^L, \mathbf{y}^L) \leftarrow$ Physician's labeling.
  else
    $(\mathbf{F}^L, \mathbf{y}^L) \leftarrow$ Labeled voxels in the nearest slice from s^{th} slice.
  end if
  $\boldsymbol{\beta} \leftarrow$ Solution returned by tLasso in Eq.(7).
  $\boldsymbol{\omega} \leftarrow$ Solution returned by wLapRLS in Eq.(16).
  Calculate the prostate-likelihood by Representer Theorem [42].
  Assign the calculated prostate-likelihood for the s^{th} slice in $\mathrm{M}_{\mathscr{L},n}$.
end for

5. Multi-Atlases Based Label Fusion

We now discuss the technical details of generating the final binary segmentation results according to the 3-D prostate-likelihood map. Although the prostates scanned at different days usually have large irregular motion, nonlinear transformation, and even appearance changing, the patient-specific prostate shape information can still be well used to guide final segmentation, since the shape differences among the same patient are relative smaller than that among different patients, in terms of the volume size, general shape, etc. To make full use of all the shape information from the planning and previous treatment images for guiding the segmentation, we adopt the multi-atlases based label fusion with the following steps:

Firstly, previous binary segmentation results (the ground truth of the planning and treatment images) $\mathbf{G}_1,\mathbf{G}_2,\ldots,\mathbf{G}_{n-1}$ and $\mathbf{G_p}$ will be rigidly aligned to the estimated prostate-likelihood map $\mathbf{M}_n$ by using the mutual-information based similarity metric with Powell's optimization strategy [35].

Secondly, we average all the obtained aligned results and further conduct the voxel-wise majority voting for each voxel, in order to get the final segmentation result $\mathbf{S}_n$. The role of this step in our whole segmentation method can be referred to the *multi-atlases based label fusion* step in Fig.4.

6. Experimental Results

In this section, we will validate the advantages of the proposed prostate segmentation method both qualitatively and quantitatively. The CT prostate dataset used in the experiment, and the experimental setting are first given. Then, we will systematically evaluate the proposed tLasso and wLapRLS, respectively. Also, we discuss the performance of the proposed method when the patients are with large prostate motion. Finally, the proposed method will be extensively compared with several state-of-the-arts developed in recent years.

6.1. CT Prostate Dataset

The proposed method was evaluated on a prostate 3-D CT-image dataset consisting of 24 patients with 330 CT images, and each patient has at least 9 images obtained from 1 planning day and several treatment days. The original resolution of each CT image is 512 × 512 × 60, with in-plane voxel size as 0.98 × 0.98 mm^2 and the inter-slice thickness as 3 mm. As a benchmark CT prostate segmentation dataset, the dataset is also used in [14][34][19][50]. Also in [32][33], the authors have chosen 11 and 10 typical patients for experiments respectively, which are subsets of the 24 patients. So it is fair and reasonable for the proposed method to compare with previous methods [14][32][33][34][19][50] on the same dataset. All the images of the 24 patients are manually segmented by the experienced physician, which are used as ground truth for evaluation in the experiments. For each patient, the first 3 images (i.e., the planning image and the first two treatment images) are used as training images, from which the patient-specific ROI is extracted, the auxiliary voxels are sampled, and segmentation ground truths are available.

6.2. Experimental Setting

In the experiments, we use three common evaluation metrics: the Dice ratio, the True Positive Fraction (TPF), and the Centroid Distance (CD).

The Dice ratio has been widely used to evaluate the prostate segmentation methods in previous works [19][34][35][50]. For two binary images (i.e., predicted and manual segmentation results), Dice Ratio measures the total voxel number of their overlapping part divided by the summation of voxel numbers in respective two binary images.

The TPF indicates that the percentage of corrected predicted prostate voxels in the manually segmented prostate regions. Basically, higher TPF indicates better coverage of the true prostate region, which is very helpful for the later radiotherapy because in this case the high energy X-ray can be delivered to kill the cancer tissues within whole prostate region.

The CD means the Euclidean distance between the central locations of the manual segmentation result and predicted result. The method with lower CD normally can better help locate the prostate center. Since prostate CT-images are 3-D, the CD along 3 directions: the lateral (x-axis), anterior-posterior (y-axis), and superior-inferior (z-axis) directions are needed to be calculated. It is noteworthy that in superior-inferior (z-axis) direction, the CD are calculated as 3 times of obtained values since the inter-slice voxel size is 3 mm which is approximately 3 times of that in x-axis and y-axis. Note that, the calculated CD is still with

the sign (positive or negative) whose aim is to keep the consistency with the related works [14][32][33][34]. Also, we list the standard deviation of CD to address this problem in the following sections. Good segmentation method, which perfectly matches the ground truth, will have higher Dice ratio and TPF, as well as lower CD.

6.3. Evaluation on the tLasso

To evaluate the performance of the proposed tLasso for feature selection, the experiment mainly focuses on the following two aspects: (1) to demonstrate if the transductive regularizer (Laplacian regularization term) in tLasso is useful, and (2) to demonstrate if the performance of tLasso outperforms the state-of-the-arts feature selection or dimension reduction methods in prostate-likelihood estimation. Therefore, Lasso [46] and two popular state-of-the-arts: PCA (principle component analysis) and LDA (linear discriminative analysis) are adopted for comparison. For PCA and LDA, the dimension after feature selection is determined by maximum likelihood estimation (MLE). We use the MATLAB toolbox of dimension reduction [51] for the implementation of the PCA and LDA. Since the whole dataset includes 24 patients, and more than 300 images, a subset (patient 1, 5, 10, 15 and 20), which include 63 images, is chosen to compare the performance of tLasso as well as other methods (PCA, LDA, Lasso). For different feature selection methods, we adopt the proposed wLapRLS as the regressor. Since wLapRLS is a regressor for prostate-likelihood estimation, the mean square errors (MSE) is adopted as an additional evaluation metric between the predict prostate-likelihood map and the manual segmentation result provided by the physician. Also, the Dice ratio, TPF and CD are used as the metric to validate the influence on the final segmentation result using different prostate-likelihood map obtained by different feature selection methods.

For the evaluation of MSE, since each 3-D CT image is composed by several 2-D slices, we adopt the win percentage of tLasso against other methods on both slice level and image level, respectively. For a particular slice/image, we calculate the MSE of tLasso and another method (e.g. PCA); if tLasso is with lower MSE, the winning number of slice-level/image-level will be added 1. The win percentage of slice-level/image-level is defined as the winning number of slice-level/image-level divided by the total number of slices/images. The win percentage of tLasso against PCA, LDA, and Lasso on MSE can be referred to Fig.6. It is obvious that the proposed tLasso outperforms PCA, LDA and Lasso in terms of lower MSE on both slice-level and image-level.

Moreover, for the subset (patient 1, 5, 10, 15 and 20), the results of Dice ratio, TPF and CD among different feature selection methods are listed in Table 1. The proposed tLasso outperforms PCA, LDA and Lasso only except the mean CD in z-axis which is slightly worse than that of Lasso. Due to the instability, Lasso has much larger standard deviation on Dice ratio, TPF and CD, compared with tLasso. By imposing Laplacian regularizer, the standard deviation of MSE of tLasso is much smaller than that of Lasso, which indicates that using the unlabeled voxels to guide feature selection is useful.

Furthermore, some representative examples among PCA, LDA, Lasso and the proposed tLasso are shown in Fig.7. The result of PCA is the worst since the segmented result cannot match the manual delineation in some cases, and the result of LDA is better than PCA,

however, some important boundary voxels are incorrectly segmented. The result of Lasso is very closeto the manual delineation, but some voxels around the prostate boundary are still misclassified due to its instability, which leads to poor prostatelikelihood prediction in several slices. The result of tLasso is close to the manual delineations.

For parameter setting in tLasso, the two important parameters γ_G and γ_S are chosen by leave-one-out cross validation. Specifically, for each patient, to segment the current treatment image $\mathbf{I}_n$, we will use brute-force-search strategy to estimate γ_G from 10^{-i} (i = 1, …, 10), as well as γ_S from 10^{-j} (j = 1, …, 10), using leave-one-out cross validation on the planning image $\mathbf{I_p}$ and previous treatment images $\mathbf{I}_1$, …, $\mathbf{I}_{n-1}$. The γ_G and γ_S, which obtain the lowest MSE values in cross validation, will be chosen as the final parameters. In the experiment, $\gamma_G = 10^{-6}$ and $\gamma_S = 10^{-8}$ are found to be the best. The γ_G integrates the graph Laplacian assumption on all voxels, if γ_G is set to 0, tLasso will degenerate to Lasso. The results validate γ_G (i.e., graph Laplacian term) is helpful in prostate segmentation, because the performance of tLasso is better than Lasso. The γ_S is used to control the whole number of selected features, which is commonly used in many related works [46][56][23].

In addition, we have introduced two novel methods RELIEF [29] and mRMR [40] for comparison. Under the same experimental setting, the mean Dice ratios of RELIEF and mRMR are 0.863 and 0.877, respectively, and the mean TPFs of RELIEF and mRMR are 0.839 and 0.852, respectively. Compared with these supervised feature selection methods (i.e., RELIEF, mRMR, Lasso), our proposed tLasso, which integrates the graph Laplacian assumption for transductive feature selection, can normally obtain better results.

6.4. Evaluation on the wLapRLS

To evaluate the performance of the proposed wLapRLS, two state-of-the-arts regressors, support vector regression (SVR) and LapRLS [2] are adopted for comparison. The feature selection is conducted by the proposed tLasso. We apply the LIBSVM toolbox [4] for SVR implementation, and two different kernels (linear kernel and RBF kernel) for SVR are investigated, respectively. Also the parameters of SVR are chosen using leave-one-out cross validation on the auxiliary voxels sampled from the planning and previous treatment images for each patient. The parameters (c = 10, g = 10 in LIBSVM) which obtain the lowest MSE among all the candidate parameters are selected. The reasons for choosing SVR and LapRLS for comparison are two folds: (1) To demonstrate if using the information of the unlabeled voxels (introducing Laplacian regularizer) is helpful, because SVR is a supervised regressor; (2) To demonstrate if imposing the prior label constraint (introducing the P_A in the objective function of wLapRLS) is useful. We also choose the patient 1, 5, 10, 15 and 20 as an example. The results of the Dice ratio, TPF and CD by the SVR (linear and RBF kernel), LapRLS and proposed wLapRLS can be referred to Table 2.

It can be found that the proposed wLapRLS outperforms the SVR (linear and RBF kernel) and LapRLS except the CD in y-axis which is slightly worse than that of SVR (RBF) and the CD in z-axis which is slightly worse than that of LapRLS. It also demonstrates that (i) using the Laplacian regularizer term and (ii) imposing prior label constraint are helpful to accurately locate the prostate.

Moreover, the win percentage (with the same definition as that in the previous section) of wLapRLS against SVR (linear), SVR (RBF) and LapRLS on MSE can be referred to Fig.8. We can found that the proposed wLapRLS is usually with lower MSE compared with the related methods both on the slice-level and image-level.

It is obvious that making full use of the structure information of unlabeled voxels (wLapRLS and LapRLS) will improve the results compared with only using the label information of the labeled and auxiliary voxels (SVR). Furthermore, some representative examples among SVR, LapRLS and the proposed wLapRLS are shown in Fig.9. Although both the Dice ratios of SVR (RBF) and LapRLS are higher than 0.9, they still fail to correctly segment the slices at the top part of the prostate. i.e., the first row in Fig.9 (SVR with linear kernel even does not locate the prostate in the first slice); some boundary voxels in other slices. i.e., the second to fifth rows in Fig.9.

For parameter setting in wLapRLS, the two important parameters γ_C and γ_G are chosen by leave-one-out cross validation, which is the same with parameter selection in tLasso mentioned above. Specifically for each patient, to segment the current treatment image $\mathbf{I}_n$, we will brute-force-search γ_C from 10^{-i} ($i = 1, \ldots, 10$), as well as γ_G from 10^{-j} ($j = 1, \ldots, 10$), using leave-one-out cross validation on the planning image $\mathbf{I_p}$ and previous treatment images $\mathbf{I}_1, \ldots, \mathbf{I}_{n-1}$. The γ_C and γ_G, which obtain the lowest MSE values in cross validation, will be chosen as the final parameters. In the experiment, $\gamma_C = 10^{-6}$ and $\gamma_G = 10^{-5}$ are found to be the best.

6.5. Comparison with Previous Methods

We now extensively compare the proposed method with previous ones since the manual segmented results of the 24 patients are available by experienced physician. The mean, standard deviation, and median value of the Dice ratio, TPF, and CD (3-directions) are calculated for all the 24 patients in Table 3. Also for each patient, we list the Dice ratio in Fig.10, TPF and CD in Fig.11, respectively. According to our observation, the proposed method can achieve satisfactory results, of which Dice ratio is larger than 0.9, TPF is around 0.9, also CD in three directions are small.

Moreover, to further evaluate the performance of the proposed method, the results of several state-of-the-arts are illustrated for comparison, which include deformable-model based methods [5][14][50], registration-based methods [10][33][34][35], and learning-based method [32][19]. The best results reported in the corresponding papers are listed. The comparisons among different methods are listed in Table 4. Evaluated metrics include mean Dice ratio, median TPF (in [5][32], median TPF are evaluated instead of mean TPF), and Average Surface Distance (ASD). The ASD is adopted to discretely measure the Euclidean distance between the surface of the manual segmentation result and predicted segmentation result. Because different CT datasets are used for experiments, for Davis *et al.*'s work [10] and Chen *et al.*'s work [5], the results of [10] and [5] are listed separately for reference. For the related works [14][34][35][19][50], all the 24 patients are evaluated which is the same with ours, so we name the 24 patients CT dataset as *CT dataset 1*. For Li *et al.*'s work [32] and Liao *et al.*'s work [33], two different subsets of the 24 patients CT dataset are chosen and evaluated, which are named as *CT dataset 2* and *CT dataset 3*, respectively. For fair

comparison, we selected the same subsets with [32][33], respectively. From the results listed in Table 4, it is obvious that the proposed method outperforms the previous ones in terms of higher Dice ratio. In the mean time, higher TPF are achieved by the proposed method compared with [5] and [32]. Except the ASD is slightly worse than several related methods.

From the above observation, we can infer that physician's manual specification information is quite useful for predicting the possible central location of prostate, because we just use the low-level features without any advanced features, like auto-context features [47] used in literature [32]. Furthermore, we also illustrate in Fig.14 several typical segmented examples for the image 8 of patient 1, the image 11 of patient 4, the image 10 of patient 14, the image 8 of patient 20 and the image 7 of patient 24, with different slice indexes, respectively. In Fig.14, the red curves denote the manual segmentation results by the physician, and the yellow curves denote the segmentation results by the proposed methods. We can found that the predicted prostate boundaries are very close to the boundaries delineated by the physician. Also the proposed method can accurately separate the prostate regions and background even in the top and bottom slices, which are usually considered very difficult to segment. According to the Dice ratio reported in Fig.10, the segmentation result of patient 14 is almost the worst (mean Dice ratio is 0.859) by applying the proposed method, however, we can still found the prostate region is successfully located using our method in Fig.14.

6.6. Patients with Large Prostate Motion

Since the proposed method aims at using physician's knowledge to aid prostate segmentation, especially for the prostates with large irregular motion. In our work, we use the same CT dataset 1 as [19][35], and the same CT dataset 2 as [32], it is known that the patients 3, 10 and 15 are with larger prostate motion, which can be referred to Fig.12. In Fig. 12, we calculate the standard deviation of the prostate centers among the planning and treatment images for each patient, and it is obvious that patient 3, 10, 15 are with large prostate motion.

For patient 10, the median Dice ratio is 0.908. For patient 15, the median Dice ratio is 0.918, which are better than the corresponding results reported in [32][19][35]. Applying the proposed method to patient 3, the median Dice ratio is 0.905, which is worse than [19][35]. The comparison is listed in Fig.13. It demonstrates that the proposed method is effective when large irregular motion occurs within prostate regions.

7. Conclusion

In this paper, we presented a novel transductive learning-based method to segment the prostate in CT images. Our proposed method aims to further improve the segmentation performance by making full use of the physician's simple manual specification (only taking a few seconds). Specifically, the physician's manual specification, as well as the patient-specific information obtained from the planning and previous treatment images, will be simultaneously employed to guide the accurate segmentation. The proposed method mainly contains the two steps: in prostate-likelihood estimation step, two novel algorithms: tLasso and wLapRLS, will sequentially adopted for joint transductive feature selection and

transductive regression, respectively, aiming to generate the prostate-likelihood map. In multi-atlases based label fusion step, the final segmentation result will be obtained according to the corresponding prostate-likelihood map as well as the previous images of the same patient. We evaluated the proposed method on a real CT prostate dataset, delineated by the experienced physician, which consists of 24 patients with 330 images. Experimental results indicate that, our proposed method can obtain higher Dice ratio, higher true positive fraction, as well as lower centroid distances compared with several state-of-the-arts, especially for the cases with large irregular prostate motion. It demonstrates that physician's simple manual specification is quite useful for guiding segmentation when large irregular prostate motion occurs within the prostate regions. In our future work, more advanced features and more intelligent manual specification methods will be investigated for better segmentation results.

Acknowledgment

The work was support by NSFC (61035003, 61175042, 61321491, 61305068, 61432008), Jiangsu 973 (BK2011005), Jiangsu NSF (BK20130581), the Program for New Century Excellent Talents in University (NCET-10-0476) and Jiangsu Clinical Medicine Special Program (BL2013033). The work was also supported by the grant from National Institute of Health (1R01 CA140413).

References

1. Barbu A, Suehling M, Xu X, Liu D, Zhou SK, Comaniciu D. A probabilistic model for automatic segmentation of the esophagus in 3-D CT scans. IEEE Trans. on Medical Imaging. 2012; 31:240–250. [PubMed: 21968722]
2. Belkin M, Niyogi P, Sindhwani V. Manifold regularization: a geometric framework for learning from labeled and unlabeled examples. Journal of Machine Learning Research (JMLR). 2006; 7:2399–2434.
3. Chan TF, Esedoglu S, Nikolov M. Algorithms for finding global minimizers of image segmentation and denoising models. SIAM Journal on Applied Mathematics. 2006; 66:1632–1648.
4. Chang CC, Lin CJ. LIBSVM: A library for support vector machines. ACM Trans. on Intelligent Systems and Technology. 2011; 2:1–27.
5. Chen S, Lovelock DM, Radke RJ. Segmenting the prostate and rectum in CT imagery using anatomical constraints. Medical Image Analysis. 2011; 15:1–11. [PubMed: 20634121]
6. Chen T, Kim S, Zhou J, Metaxas D, Rajagopal G, Yue N. 3D meshless prostate segmentation and registration in image guided radiotherapy. MICCAI. 2009:43–50. [PubMed: 20425969]
7. Chowdhury N, Toth R, Chappelow J, Kim S, Motwani S, Punekar S, Lin H, Both S, Vapiwala N, Hahn S, Madabhushi A. Concurrent segmentation of the prostate on MRI and CT via linked statistical shape models for radiotherapy planning. Medical Physics. 2012; 39:2217–2228.
8. Couprie C, Grady L, Najman L, Talbot H. Power watersheds: a new image segmentation framework extending graph cuts, random walker and optimal spanning forest. ICCV. 2009:731–738.
9. Dalal N, Triggs B. Histograms of oriented gradients for human detection. CVPR. 2005:886–893.
10. Davis BC, 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; 8:442–450. [PubMed: 16685876]
11. Dehmeshki J, Amin H, Valdivieso M, Ye X. Segmentation of pulmonary nodules in thoracic CT scans: a region growing approach. IEEE Trans. on Medical Imaging. 2008; 27:467–480. [PubMed: 18390344]
12. Doquire G, Verleysen M. A graph Laplacian based approach to semi-supervised feature selection for regression problems. Neurocomputing. 2013; 121:5–13.

13. Ecabert O, Peters J, Schramm H, Lorenz C, von Berg J, Walker MJ, Vembar M, Olszewski ME, Subramanyan K, Lavi G, Weese J. Automatic model-based segmentation of the heart in CT images. IEEE Trans. on Medical Imaging. 2008; 27:1189–1201. [PubMed: 18753041]
14. Feng Q, Foskey M, Chen W, Shen D. Segmenting CT prostate images using population and patient-specific statistics for radiotherapy. Medical Physics. 2010; 37:4121–4132. [PubMed: 20879572]
15. Feulner J, Zhou SK, Hammon M, Seifert S, Huber M, Comaniciu D, Hornegger J, Cavallaro A. A probabilistic model for automatic segmentation of the esophagus in 3-D CT scans. IEEE Trans. on Medical Imaging. 2011; 30:1252–1264. [PubMed: 21303741]
16. Freedman D, Radke RJ, Zhang T, Jeong Y, Lovelock DM, Chen GT. Modelbased segmentation of medical imagery by matching distributions. IEEE Trans. on Medical Imaging. 2005; 24:281–292. [PubMed: 15754979]
17. Freund, Y.; Schapire, RE. A decision-theoretic generalization of on-line learning and an application to boosting; Proceedings of the Second European Conference on Computational Learning Theory; 1995. p. 23-37.
18. Gao G, Yang J. A novel sparse representation based framework for face image super-resolution. Neurocomputing. 2014; 134:92–99.
19. Gao Y, Liao S, Shen D. Prostate segmentation by sparse representation based classification. MICCAI. 2012:452–458.
20. Gao Y, Sandhu R, Fichtinger G, Tannenbaum AR. A coupled global registration and segmentation framework with application to magnetic resonance prostate imagery. IEEE Trans. on Medical Imaging. 2010; 29:1781–1794. [PubMed: 20529727]
21. Ghosh P, Mitchell M. Segmentation of medical images using a genetic algorithm. GECCO. 2006:1171–1178.
22. Ghose S, et al. A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images. Computer Methods and Programs in Biomedicine. 2012; 108:262–287. [PubMed: 22739209]
23. Gong P, Ye J, Zhang C. Multi-stage multi-task feature learning. Journal of Machine Learning Research (JMLR). 2013; 14:2979–3010.
24. Grady L. Random walks for image segmentation. IEEE Trans. on Pattern Analysis and Machine Intelligence. 2006; 28:1768–1783.
25. Grant M, Boyd S. CVX: Matlab software for disciplined convex programming, version 1.21. 2011 Apr. http://cvxr.com/cvx.
26. Hamamci A, Kucuk N, Karaman K, Enqin K, Unal G. Tumor-Cut: segmentation of brain tumors on contrast enhanced MR images for radiosurgery applications. IEEE Trans. on Medical Imaging. 2012; 31:790–804. [PubMed: 22207638]
27. Jenkinson M, Bannister P, Brady M, Smith S. Improve optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage. 2002; 17:825–841. [PubMed: 12377157]
28. Kang C, Liao S, Xiang S, Pan C. Kernel sparse representation with pixel-level and region-level local feature kernels for face recognition. Neurocomputing. 2014; 133:141–152.
29. Kira, K.; Rendell, L. A Practical Approach to Feature Selection; Proceedings of the Ninth International Workshop on Machine Learning; 1992. p. 249-256.
30. Langerak TR, van der Heide UA, Kotte AN, Viergever MA, van Vulpen M, Pulim JP. Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). IEEE Trans. on Medical Imaging. 2010; 29:2000–2008. [PubMed: 20667809]
31. Lee H, Battle A, Raina R, Ng AY. Efficient sparse coding algorithms. NIPS. 2007:801–808.
32. Li W, Liao S, Feng Q, Chen W, Shen D. Learning image context for segmentation of prostate in CT-guided radiotherapy. MICCAI. 2011:570–578. [PubMed: 22003745]
33. Liao S, Shen D. A learning based hierarchical framework for automatic prostate localization in CT images. Prostate Cancer Imaging MICCAI workshop. 2011:1–8.
34. Liao S, Shen D. A feature based learning framework for accurate prostate localization in CT images. IEEE Trans. on Image Processing. 2012; 21:3546–3559.

35. Liao S, Gao Y, Lian J, Shen D. Sparse patch-based label propagation for accurate prostate localization in CT images. IEEE Trans. on Medical Imaging. 2012; 32:419–434. [PubMed: 23204280]
36. Liu X, Langer DL, Haider MA, Yang Y, Wernick MN, Yetik IS. Prostate cancer segmentation with simultaneous estimation of Markov random field parameters and class. IEEE Trans. on Medical Imaging. 2009; 28:906–915. [PubMed: 19164079]
37. Lu X, Li X. Group sparse reconstruction for image segmentation. Neurocomputing. 2014; 136:41–48.
38. Mallat G. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. on Pattern Analysis and Machine Intelligence. 1989; 11:674–693.
39. Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. on Pattern Analysis and Machine Intelligence. 2002; 24:971–987.
40. Peng HC, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. on Pattern Analysis and Machine Intelligence. 2005; 27:1226–1238.
41. Rousson, M.; Khamene, A.; Diallo, M.; Celi, JC.; Sauer, F. Constrained surface evolutions for prostate and bladder segmentation in CT images; Proceedings of the 1st International Conference on Computer Vision for Biomedical Image Applications (CVBIA); 2005. p. 251-260.
42. Scholkopf B, Herbrich R, Smola AJ. A generalized representer theorem. COLT. 2001:416–426.
43. Shen D, et al. Segmentation of prostate boundaries from ultrasound images using statistical shape model. IEEE Trans. on Medical Imaging. 2003; 22:539–551. [PubMed: 12774900]
44. Song Q, Wu X, Liu Y, Smith M, Buatti J, Sonka M. Optimal graph search segmentation using arc-weighted graph for simultaneous surface detection of bladder and prostate. MICCAI. 2009:827–835. [PubMed: 20426188]
45. Subrahmanyam M, Wu QM. Local ternary co-occurrence patterns: A new feature descriptor for MRI and CT image retrieval. Neurocomputing. 2013; 119:399–412.
46. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B. 1996; 58:267–288.
47. Tu Z, Bai X. Auto-context and its application to high-level vision tasks and 3D brain image segmentation. IEEE Trans. on Pattern Analysis and Machine Intelligence. 2010; 32:1744–1757.
48. Wang H, Dong L, Lii MF, Lee AL, de Crevoisier R, Mohan R, Cox JD, Kuban DA, Cheung R. Implementation and validation of a three-dimensional deformable registration algorithm for targeted prostate cancer radiotherapy. International Journal Radiation Oncology Biology Physics. 2005; 61:725–735.
49. Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y. Robust face recognition via sparse representation. IEEE Trans. on Pattern Analysis and Machine Intelligence. 2009; 31:210–227.
50. Wu Y, Liu G, Huang M, Jiang J, Yang W, Chen W, Feng Q. Prostate Segmentation based on Variant Scale Patch and Local Independent Projection. IEEE Trans. on Medical Imaging. (in press).
51. van der Maaten, LJP.; Postma, EO.; van den Herik, HJ. Technical Report TiCC-TR 2009-005. Tilburg University; 2009. Dimensionality reduction: a comparative review.
52. Yang B, Chen S. Sample-dependent graph construction with application to dimensionality reduction. Neurocomputing. 2014; 74:301–314.
53. Yang J, Yan R, Hauptmann AG. Cross-domain video concept detection using adaptive svms. ACM MM. 2007:188–197.
54. Zhan Y, Shen D. Deformable segmentation of 3-D ultrasound prostate images using statistical texture matching method. IEEE Trans. on Medical Imaging. 2006; 25:256–272. [PubMed: 16524083]
55. Zhan Y, Shen D, Zeng J, Sun L, Fichtinger G, Moul J, Davatzikos C. Targeted prostate biopsy using statistical image analysis. IEEE Trans. on Medical Imaging. 2007; 26:779–788. [PubMed: 17679329]
56. Zhou J, Liu J, Narayan VA, Ye J. Modeling disease progression via fused sparse group lasso. KDD. 2012:1095–1103. [PubMed: 25309808]

Biographies

Yinghuan Shi is currently an assistant researcher in the Department of Computer Science and Technology of Nanjing University, China. He received his Ph.D. and B.Sc. degrees from Department of Computer Science of Nanjing University, in 2013 and 2007, respectively. He was a visiting student in University of North Carolina at Chapel Hill (U.S.) from Sep. 2011 to Sep. 2012, and Massey University (New Zealand) from Nov. 2009 to Feb. 2010, respectively. His research interests include computer vision and medical image analysis. He has published more than 20 research papers in related journals and conferences such as IEEE Trans. Biomedical Engineering, CVPR and IPMI. He serves as a program committee member for several international conferences.

Yaozong Gao received the B.E. degree from the department of software engineering, Zhejiang University, China, in 2008, and the M.Sc. degree from the department of computer science, Zhejiang University, China, in 2011. He is currently pursuing the Ph.D. degree in the department of computer science, University of North Carolina at Chapel Hill (UNC-CH), North Carolina. Since fall 2011, he has been with the department of computer science, the department of radiology, and biomedical research imaging center (BRIC), UNC-CH. His research interests include machine learning, computer vision, and medical image analysis.

Shu Liao received the BEng degree (First Class Honors and Academic Achievement Awards) in computer engineering and the MPhil and PhD degrees in computer science and engineering from The Hong Kong University of Science and Technology in 2005, 2007, and 2010, respectively. He was a postdoctoral research fellow in the Department of Radiology, Biomedical Research Imaging Center, University of North Carolina at Chapel Hill from 2010 to 2012. His research interests include medical image analysis, texture analysis, and face recognition. He is currently a senior scientist in Siemens Medical Solution, US.

Daoqiang Zhang is currently a Professor of the Department of Computer Science and Engineering in Nanjing University of Aeronautics and Astronautics, China. He received his B.Sc. and Ph.D. degrees in computer science from Nanjing University of Aeronautics and Astronautics, China, in 1999 and 2004, respectively. His research interests include machine learning, pattern recognition, and biomedical image analysis. In these areas he has published more than100 technical papers in the refereed international journals and conference proceedings. He was nominated for the National Excellent Doctoral Dissertation Award of China in 2006, won the best paper award at the 9th Pacific Rim International Conference on Artificial Intelligence (PRICAI'06), and was the winner of the best paper award honorable mention of Pattern Recognition Journal 2007. He serves as an editorial board member for the Computational Intelligence and Neuroscience Journal, and also serves as a program committee member for several international conferences. He is a member of the Machine Learning Societyof the Chinese Association of Artificial Intelligence (CAAI), and the Artificial Intelligence & Pattern Recognition Society of the China Computer Federation (CCF).

Yang Gao received the Ph.D. degree from the Department of Computer Science and Technology, Nanjing University, Nanjing, China, in 2000. Currently, He is a Professor, also the Deputy Director, in the Department of Computer Science and Technology, Nanjing University. He is currently directing the Reasoning and Learning (R&L) Research Group in Nanjing University. He has published more than 100 papers in top-tired conferences and journals, e.g., IEEE TNNLS, IEEE TBME, CVIU, PR, CVPR. His current research interests include artificial intelligence and machine learning. He also serves as Program Chair, Area Chair for many international conferences.

Dinggang Shen is a Professor of Radiology, Biomedical Research Imaging Center (BRIC), Computer Science, and Biomedical Engineering in the University of North Carolina at Chapel Hill (UNC-CH). He is currently directing the Center for Image Informatics, the Image Display, Enhancement, and Analysis (IDEA) Lab in the Department of Radiology, and also the medical image analysis core in the BRIC. He was a tenure-track assistant

professor in the University of Pennsylvanian (UPenn), and a faculty member in the Johns Hopkins University. Dr. Shen's research interests include medical image analysis, computer vision, and pattern recognition. He has published more than 500 papers in the international journals and conference proceedings. He serves as an editorial board member for six international journals. He also serves in the Board of Directors, The Medical Image Computing and Computer Assisted Intervention (MICCAI) Society.

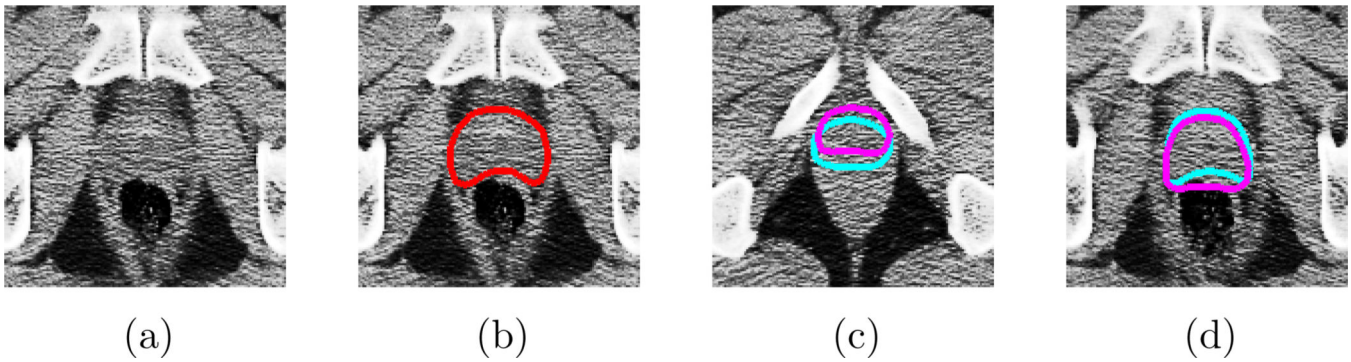

Figure 1.

Challenges in automated prostate CT segmentation. (a)(b) Low contrast in CT image (without and with the manual segmentation); (c) Large prostate motion, (d) large shape appearance change, relative to the bones, even after bone-based alignment for the two CT images.

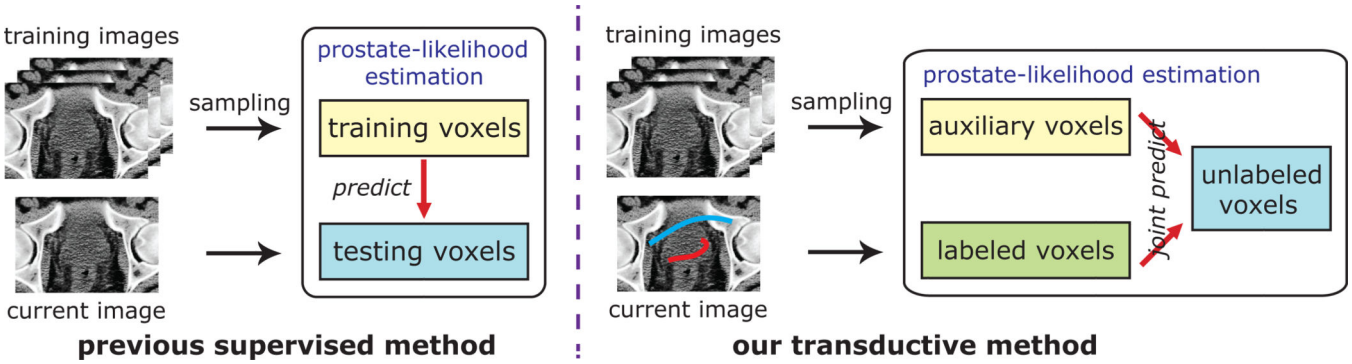

Figure 2.

The difference between previous supervised methods and the proposed transductive method. Our proposed method uses the information from both the physician's knowledge as well as the planning and previous treatment images. Red and blue lines in the current image of the proposed method mean prostate and background voxels labeled by the physician, respectively.

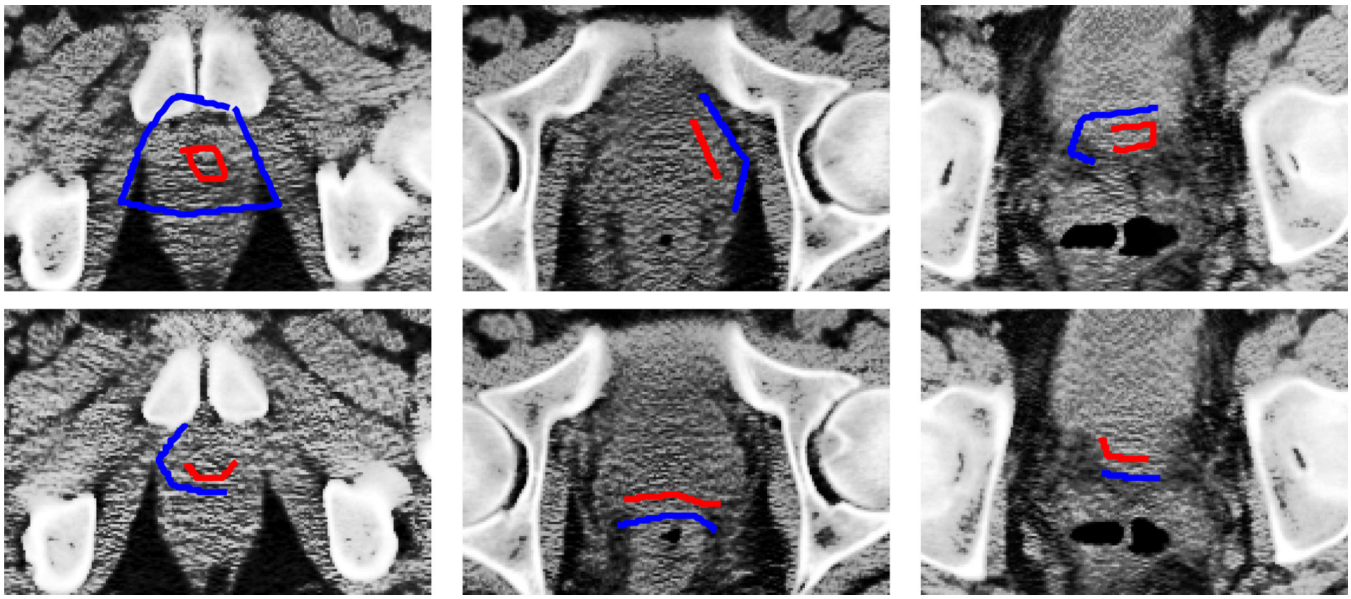

Figure 3.

Typical examples of physician's labeling. Red curve denotes the labeled prostate voxels, and blue curve denotes the labeled background voxels. The labeled voxels usually lie on the boundary of the prostate region and background, which are difficult to distinguish.

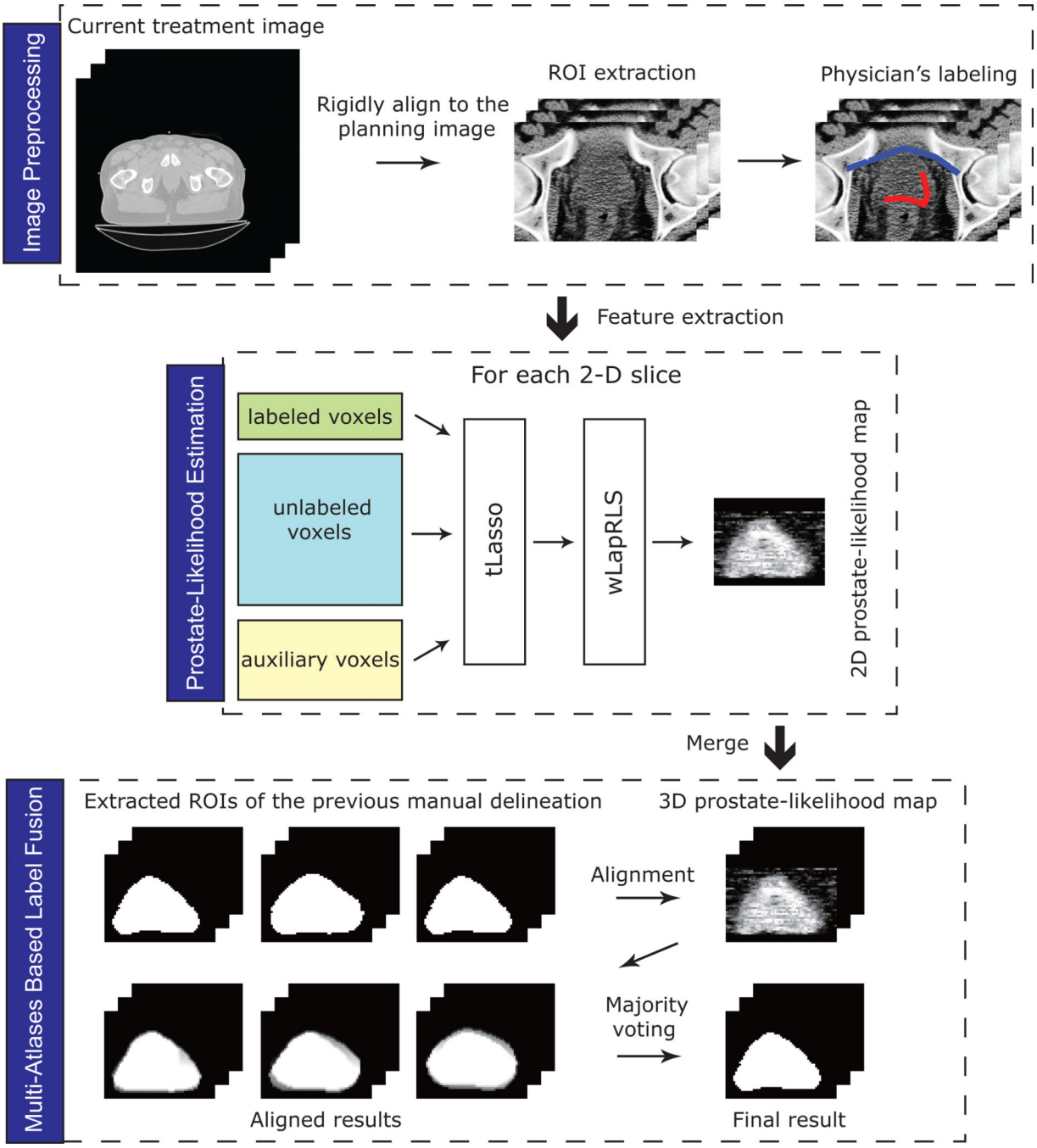

Figure 4.
The flowchart of the proposed transductive prostate segmentation method.

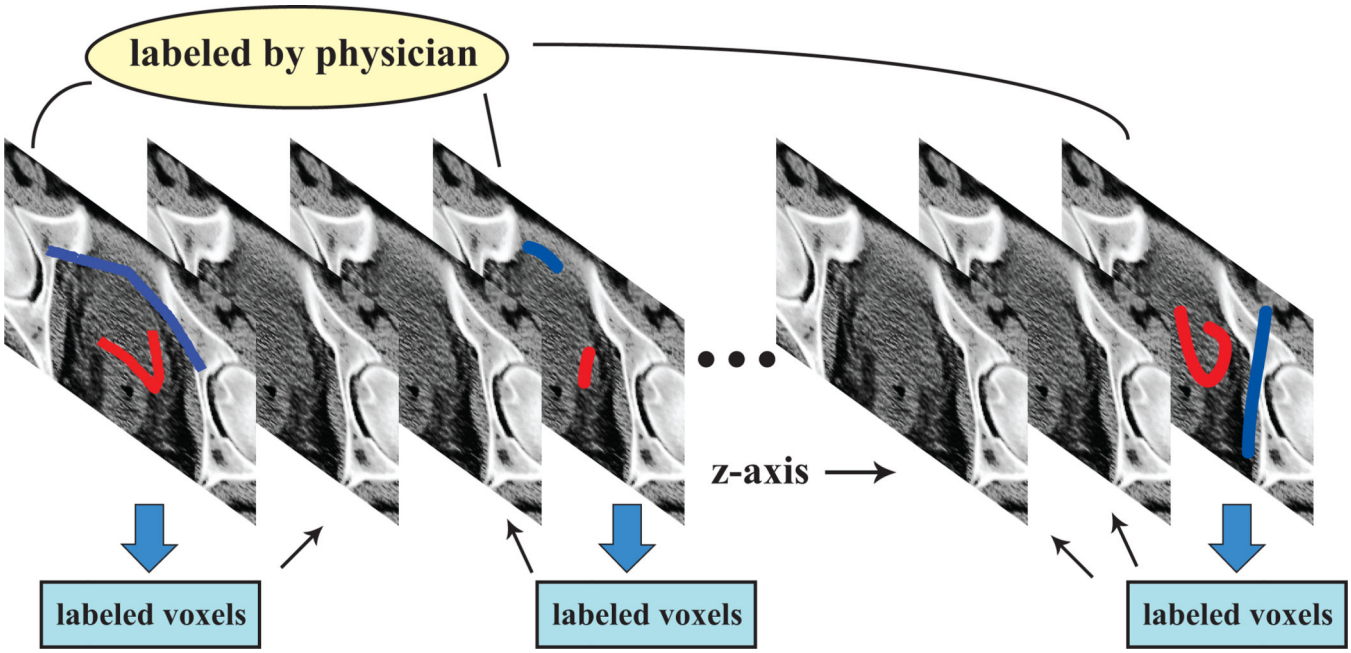

Figure 5.

An example in z-axis, labeled voxels are shared for the slices which are not labeled by the physician. The four slices which are not labeled will share the same labeled voxels of their nearest slices, respectively.

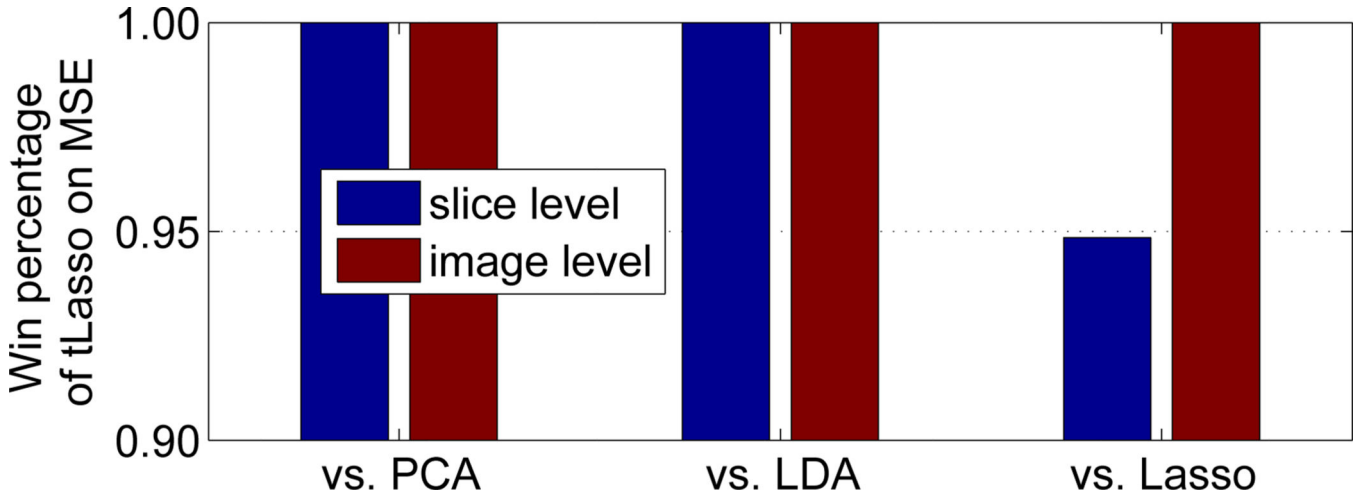

Figure 6.

The win percentage of tLasso against other methods (PCA, LDA and Lasso) on MSE in slice level and image level, respectively. Slice level means that the comparison is on each corresponding individual slice, while image level means that the comparison is on each corresponding individual image.

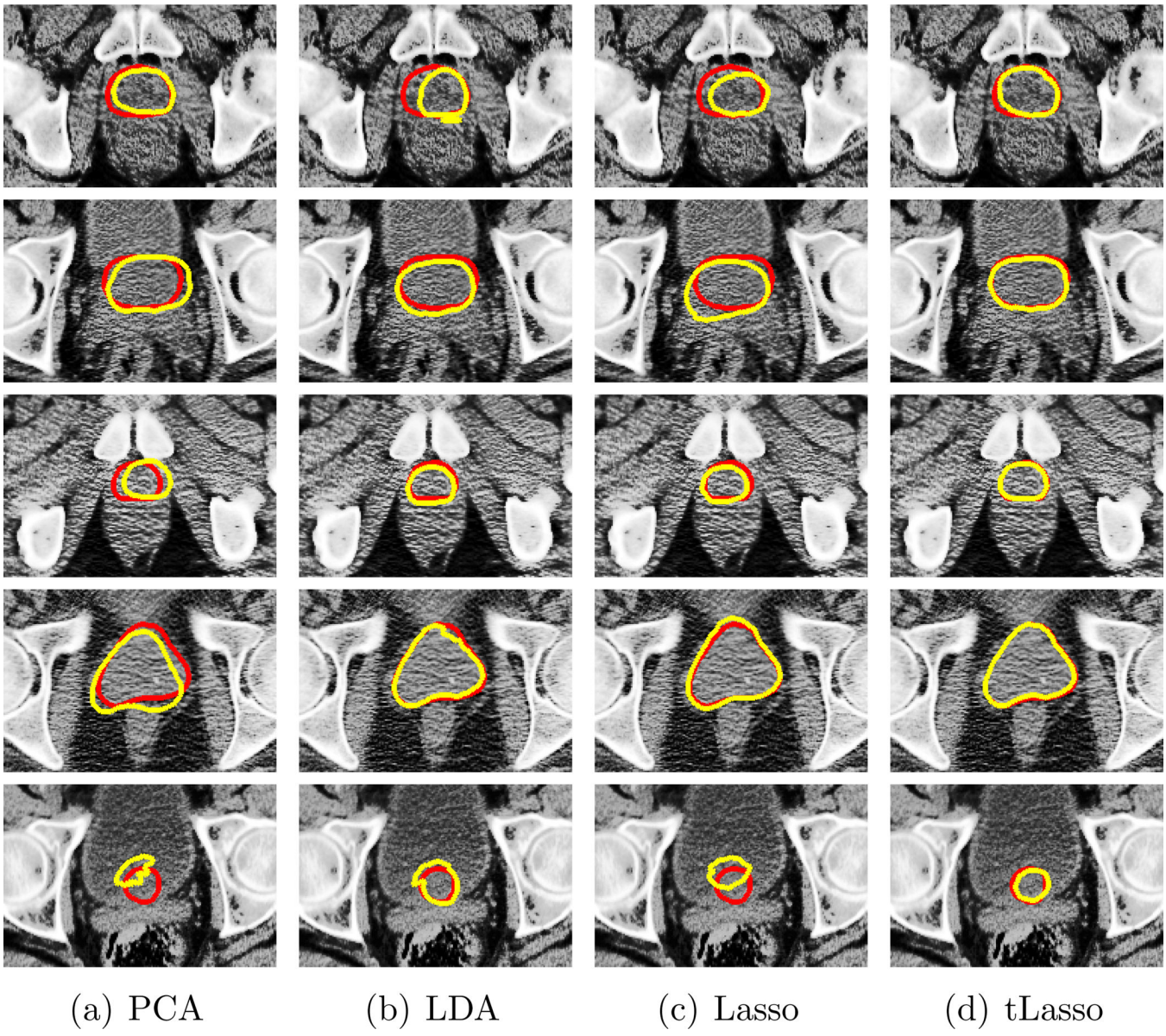

Figure 7.

Representative segmentation results for patient 1, 5, 10, 15 and 20 shown in different rows, by different feature selection methods shown in different columns. Red curve indicates manual segmentation results by physician and yellow curve indicates automated segmentation results.

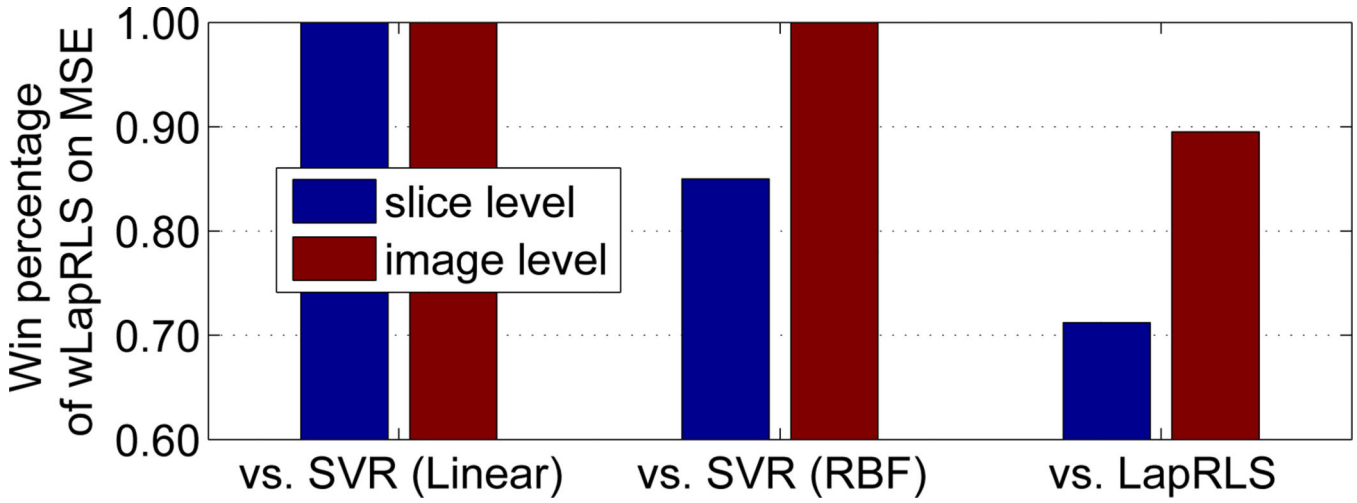

Figure 8.

The win percentage of wLapRLS against SVR (linear), SVR (RBF) and LapRLS on MSE in slice level and image level, respectively. Slice level means that the comparison is on each corresponding individual slice, while image level means that the comparison is on each corresponding individual image.

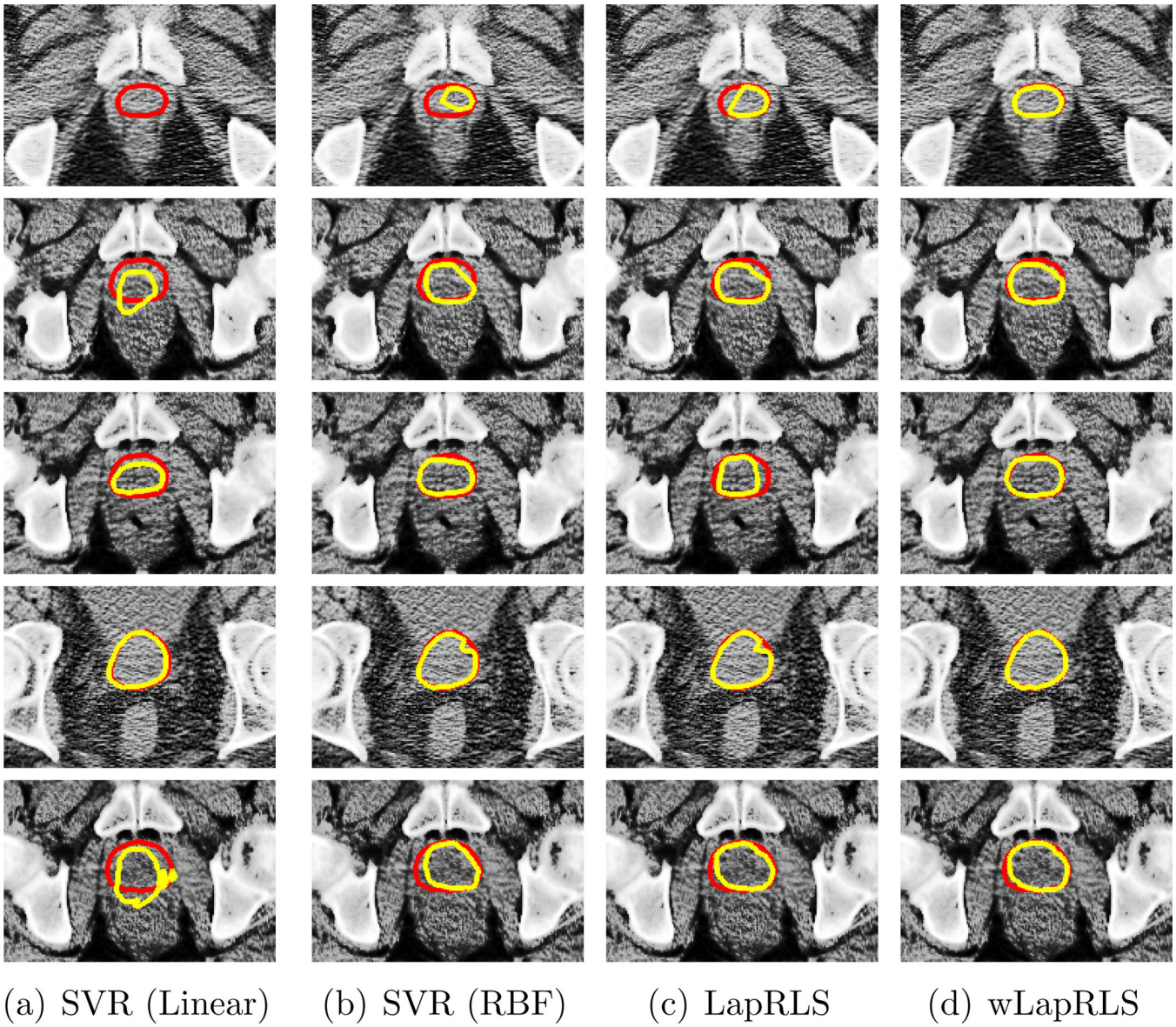

Figure 9.

Representative segmentation results for patient 1, 5, 10, 15 and 20 shown in different rows by different regression methods shown in different columns. Red curve indicates manual segmentation results by physician and yellow curve indicates automated segmentation results by different regression methods. Feature selection is conducted by tLasso.

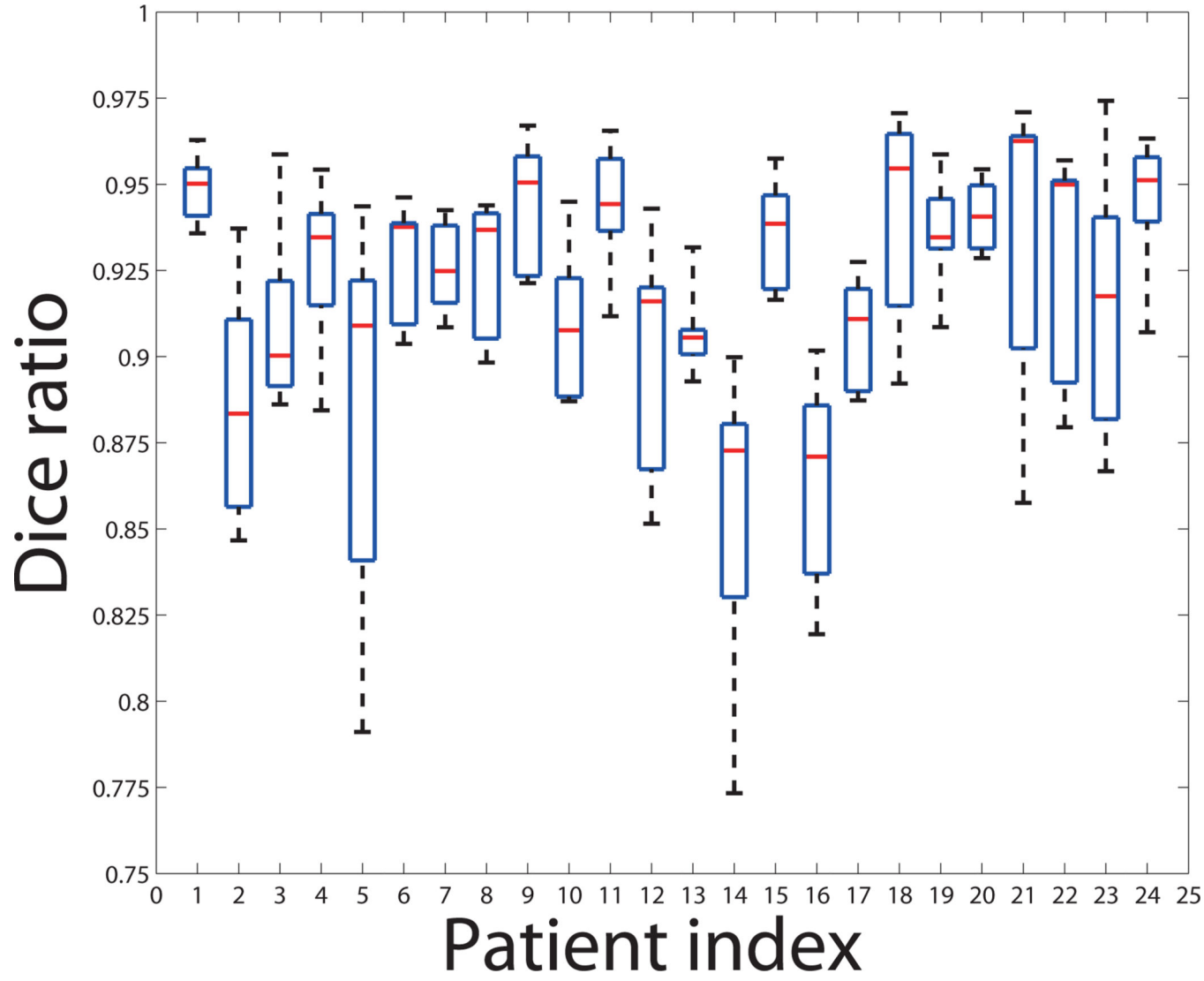

Figure 10.

The Dice ratio of the 24 patients. Five horizontal lines (ascending order in values) means the min, 25% percentile, median, 75% percentile, and the max value, respectively.

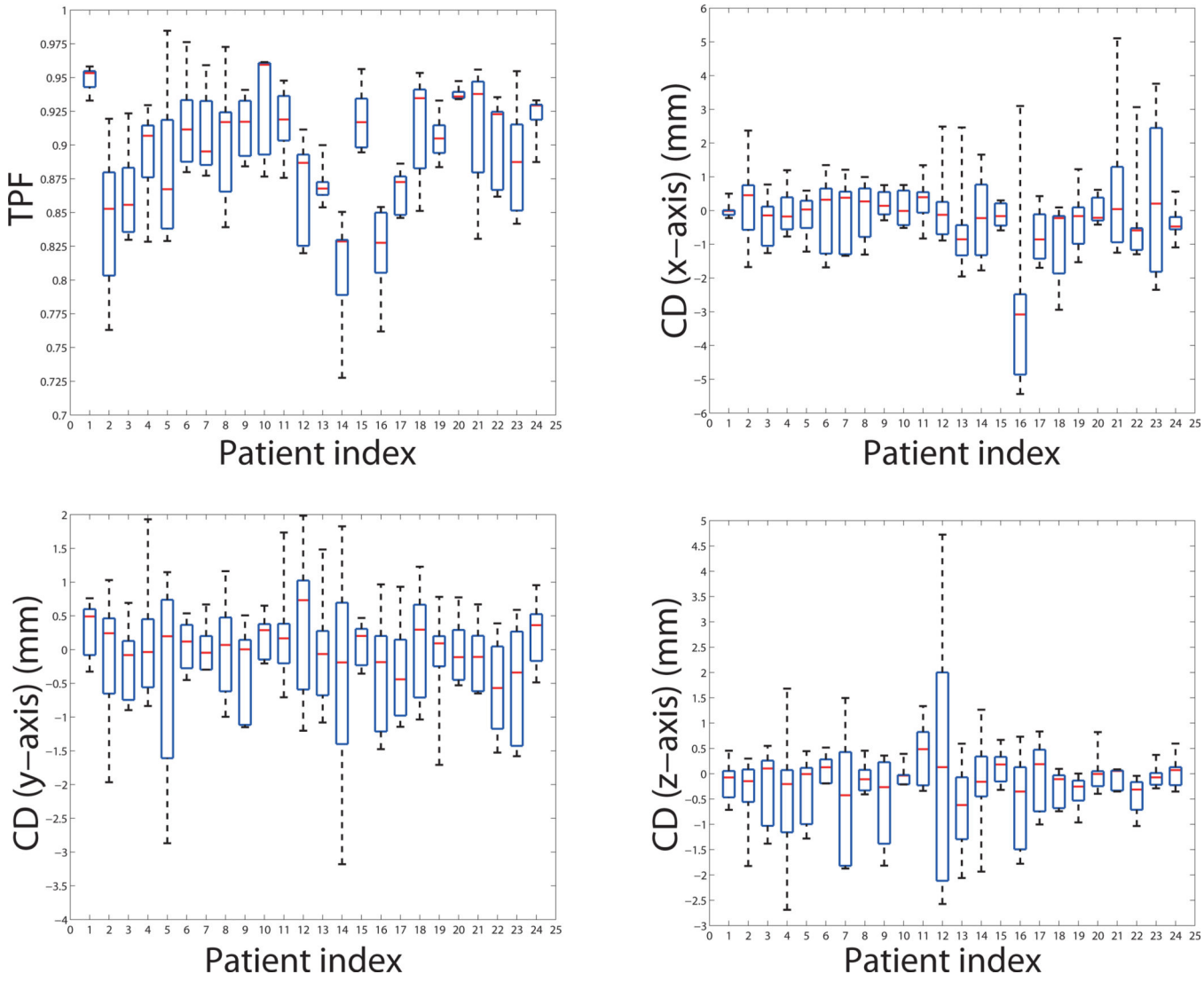

Figure 11.

The results of TPF and CD along the lateral (x-axis), anterior-posterior (y-axis), and superior-inferior (z-axis) directions, respectively. Five horizontal lines (ascending order in values) means the min, 25% percentile, median, 75% percentile, and the max value, respectively.

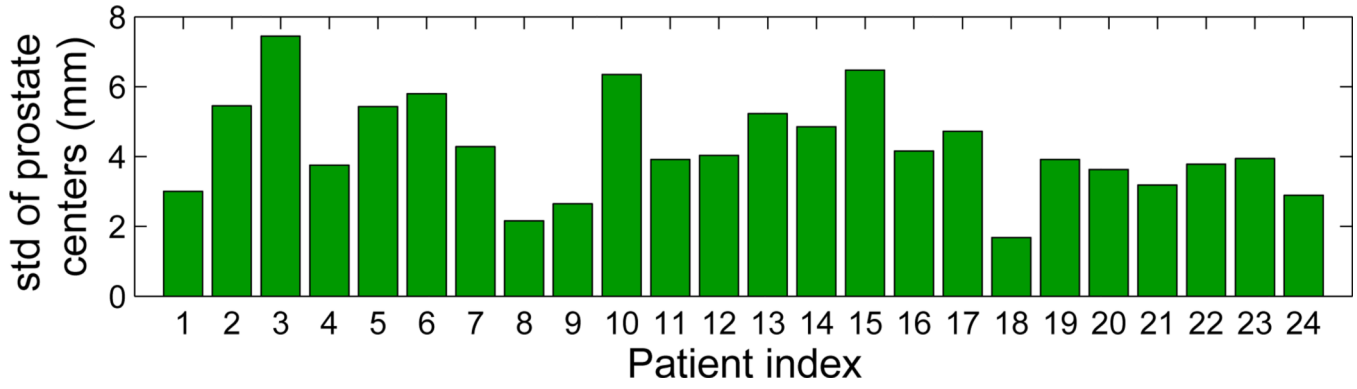

Figure 12.

The standard deviation of prostate centers for each patients. Patient 3, 10, 15 are with large prostate motion according to the observation

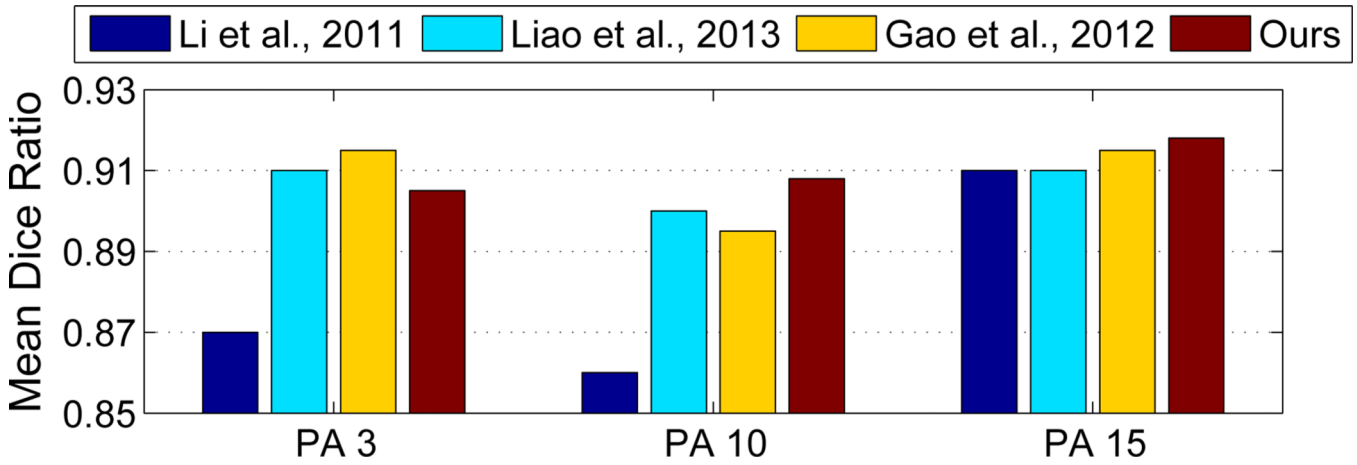

Figure 13.

The median Dice ratio of patient 3, 10 and 15 compared with [32][19][35]. The patient 3, 10 and 15 are considered with large irregular prostate motion.

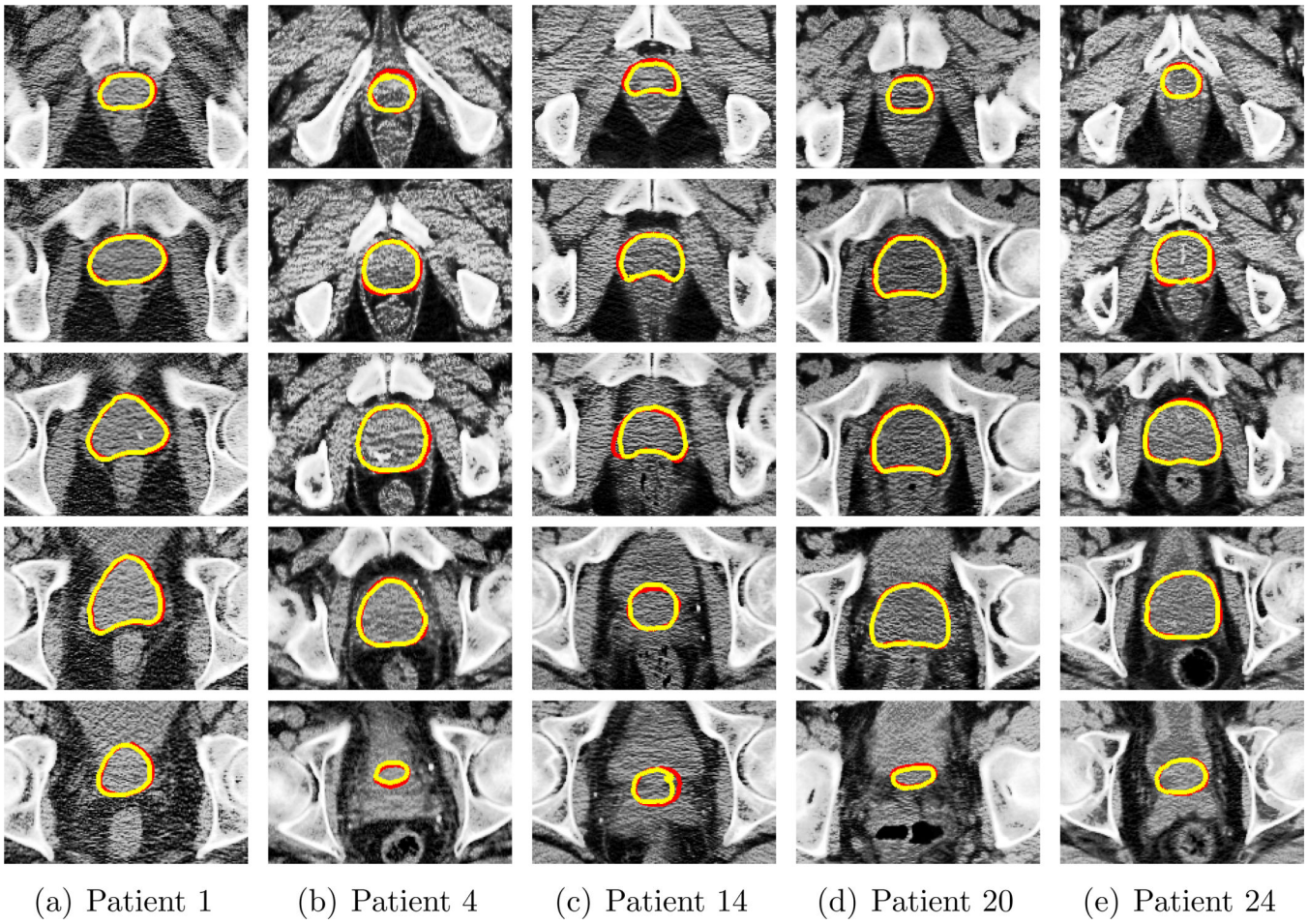

Figure 14.
Typical segmentation results for Patients 1, 4, 14, 20, 24 with different slice indexes in different rows. Red curve indicates manual segmentation results by physician and yellow curve indicates segmentation results by using our proposed method.

Table 1

Comparison on the mean and standard deviation on the Dice ratio, TPF, and CD among different feature selection methods for patient 1, 5, 10, 15 and 20. The regressor is the proposed wLapRLS. The best results are marked by bold font.

Methods	PCA	LDA	Lasso	tLasso
Dice ratio	0.768 ± 0.048	0.878 ± 0.045	0.919 ± 0.060	**0.936** ± **0.039**
TPF	0.703 ± 0.065	0.830 ± 0.057	0.868 ± 0.083	**0.877** ± **0.050**
CD (x-axis) (mm)	0.917 ± 3.139	−0.258 ± 0.948	−0.109 ± 2.320	−**0.038** ± **0.227**
CD (y-axis) (mm)	1.805 ± 3.715	−0.401 ± 1.730	0.193 ± 2.022	**0.171** ± **0.656**
CD (z-axis) (mm)	−0.216 ± 0.664	−0.097 ± 0.330	**0.033** ± 0.476	−0.038 ± **0.127**

Table 2

Comparison on the mean and standard deviation on the Dice ratio, TPF, and CD among different regressors (SVR with linear kernel, SVR with RBF kernel, LapRLS, wLapRLS) for patient 1, 5, 10, 15 and 20. Feature selction is conducted by tLasso. The best results are marked by bold font.

Methods	SVR (linear)	SVR (RBF)	LapRLS	wLapRLS
Dice ratio	0.855 ± 0.057	0.909 ± 0.043	0.918 ± 0.042	**0.936** ± **0.039**
TPF	0.782 ± 0.073	0.843 ± 0.063	0.861 ± 0.063	**0.877** ± **0.050**
CD (x-axis) (mm)	−0.144 ± 0.992	−0.085 ± 0.255	−0.065 ± 0.240	−**0.038** ± **0.227**
CD (y-axis) (mm)	−0.594 ± 2.136	**0.115** ± **0.497**	0.202 ± 0.435	0.171 ± 0.656
CD (z-axis) (mm)	−0.118 ± 0.359	−0.025 ± 0.146	−**0.019** ± **0.099**	−0.038 ± 0.127

Table 3

The Dice ratio, TPF and CD for the 24 patients by using the proposed method.

	Mean ± Std	Median
Dice ratio	0.920 ± 0.036	0.926
TPF	0.901 ± 0.052	0.901
CD (x-axis) (mm)	−0.08 ± 1.27	−0.12
CD (y-axis) (mm)	0.02 ± 0.80	0.04
CD (z-axis) (mm)	−0.12 ± 0.83	−0.09

Table 4

Comparison of different performance measurements (mean Dice ratio, median TPF, and mean ASD) with other related methods, with the best performance marked by bold font.

	# Patients	# Images	Methods	Mean Dice	Median TPF	ASD (mm)
Other datasets	3	40	Davis *et al.*[10]	0.820	N/A	N/A
	13	185	Chen *et al.*[5]	N/A	0.840	1.10± N/A
CT dataset 1	24	330	Feng *et al.*[14]	0.893	N/A	2.08 ± 0.79
			Liao *et al.*[34]	0.899	N/A	1.08± N/A
			Gao *et al.*[19]	0.913	N/A	1.24 ± 0.77
			Liao *et al.*[35]	0.909	N/A	**0.97**± N/A
			Wu *et al.*[50]	0.914	N/A	1.30 ± 0.49
			our method	**0.920**	**0.901**	1.26 ± 0.78
CT dataset 2	11	164	Li *et al.*[32]	0.908	0.900	1.40± N/A
			our method	**0.923**	**0.911**	1.27 ± 0.80
CT dataset 3	10	176	Liao *et al.*[34]	0.896	N/A	N/A
			our method	**0.922**	**0.907**	1.28 ± 0.80