

# Prostate segmentation by sparse representation based classification

Yaozong Gao<sup>a)</sup>

Department of Computer Science, Department of Radiology, and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Shu Liao<sup>b)</sup> and Dinggang Shen<sup>c)</sup>

Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

(Received 10 June 2012; revised 15 August 2012; accepted for publication 4 September 2012; published 1 October 2012)

**Purpose:** The segmentation of prostate in CT images is of essential importance to external beam radiotherapy, which is one of the major treatments for prostate cancer nowadays. During the radiotherapy, the prostate is radiated by high-energy x rays from different directions. In order to maximize the dose to the cancer and minimize the dose to the surrounding healthy tissues (e.g., bladder and rectum), the prostate in the new treatment image needs to be accurately localized. Therefore, the effectiveness and efficiency of external beam radiotherapy highly depend on the accurate localization of the prostate. However, due to the low contrast of the prostate with its surrounding tissues (e.g., bladder), the unpredicted prostate motion, and the large appearance variations across different treatment days, it is challenging to segment the prostate in CT images. In this paper, the authors present a novel classification based segmentation method to address these problems.

**Methods:** To segment the prostate, the proposed method first uses sparse representation based classification (SRC) to enhance the prostate in CT images by pixel-wise classification, in order to overcome the limitation of poor contrast of the prostate images. Then, based on the classification results, previous segmented prostates of the same patient are used as patient-specific atlases to align onto the current treatment image and the majority voting strategy is finally adopted to segment the prostate. In order to address the limitations of the traditional SRC in pixel-wise classification, especially for the purpose of segmentation, the authors extend SRC from the following four aspects: (1) A discriminant subdictionary learning method is proposed to learn a discriminant and compact representation of training samples for each class so that the discriminant power of SRC can be increased and also SRC can be applied to the large-scale pixel-wise classification. (2) The L1 regularized sparse coding is replaced by the elastic net in order to obtain a smooth and clear prostate boundary in the classification result. (3) Residue-based linear regression is incorporated to improve the classification performance and to extend SRC from hard classification to soft classification. (4) Iterative SRC is proposed by using context information to iteratively refine the classification results.

**Results:** The proposed method has been comprehensively evaluated on a dataset consisting of 330 CT images from 24 patients. The effectiveness of the extended SRC has been validated by comparing it with the traditional SRC based on the proposed four extensions. The experimental results show that our extended SRC can obtain not only more accurate classification results but also smoother and clearer prostate boundary than the traditional SRC. Besides, the comparison with other five state-of-the-art prostate segmentation methods indicates that our method can achieve better performance than other methods under comparison.

**Conclusions:** The authors have proposed a novel prostate segmentation method based on the sparse representation based classification, which can achieve considerably accurate segmentation results in CT prostate segmentation. © 2012 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4754304>]

Key words: sparse representation, classification, segmentation, prostate CT images, image-guided radiation therapy

## I. INTRODUCTION

Prostate cancer is the second leading cause of cancer death in American men. According to Cancer Facts & Figures,<sup>1</sup> about one man in six will be diagnosed with prostate cancer during his lifetime and about one sixth of them will eventually die of prostate cancer. Needle biopsy of the prostate follow-

ing an elevated prostate specific antigen (PSA) levels or an abnormal digital rectal examination (DRE) is now a standard way for diagnosis of prostate cancer.<sup>2,3</sup> When the prostate cancer is diagnosed in its early stage, it is usually curable<sup>4</sup> and the treatment is even effective at the later stages. Currently, external beam radiotherapy is one of the major clinical treatments for prostate cancer. During the radiotherapy, the

prostate is radiated by high-energy x rays from different directions. In order to maximize the harm to the tumor and minimize the harm to the surrounding healthy tissues (e.g., bladder and rectum), the prostate needs to be accurately localized. Therefore, the effectiveness and efficiency of external beam radiation therapy highly depend on the accurate localization of the prostate.

The external beam radiation therapy basically has two stages, namely, *the planning stage* and *the treatment stage*. In the planning stage, a CT image named *planning image* is scanned from the patient. The prostate in the planning image is then manually delineated by the physician and a dose plan which encodes the intensities and shape of radiation beams to be delivered is carefully designed on the planning image space. In the treatment stage, dose is delivered in daily fractions over a period of 2–10 weeks. At each treatment day, a CT image named *treatment image* is acquired and the prostate region in this image needs to be identified so that the dose plan made in the planning image can be transformed to the current treatment image. Traditionally, the prostate is manually delineated by the physician. However, this process is laborious and time-consuming, which calls for automatic or semi-automatic prostate segmentation methods to save physician's time and efforts.

However, it is challenging to segment the prostate from CT images for three main reasons. First, the prostate is of low contrast with its surrounding tissues, as illustrated in Figs. 1(a) and 1(b). There is little intensity difference between the prostate and its surrounding tissues. Therefore, it is difficult for the intensity-based method (e.g., snake) to accurately

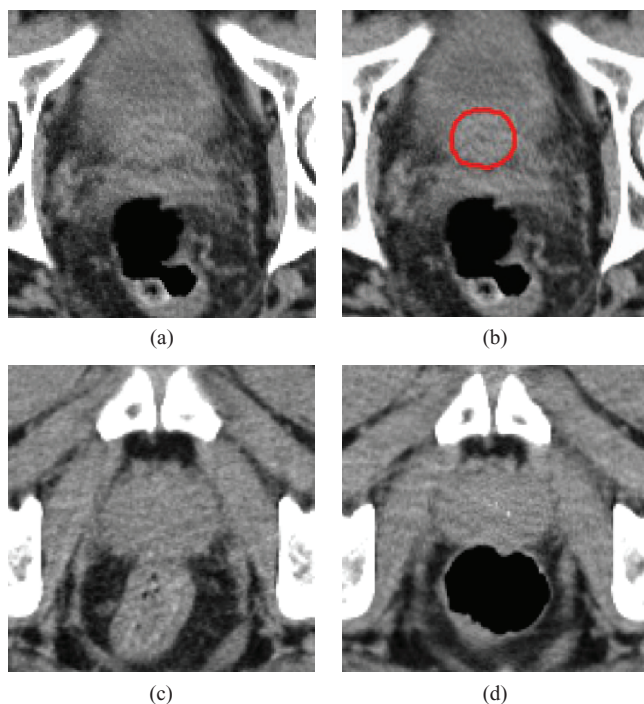


FIG. 1. (a) A typical slice of prostate CT image. (b) The same slice as (a) with prostate contour manually delineated. (c) and (d) The corresponding CT slices of different treatment images of the same patient with bowel filling and bowel gas, respectively.

localize the prostate boundary. Second, due to the uncertainty of bowel gas and filling, the image appearance can be dramatically different even for the same patient during different treatment fractions. Figures 1(c) and 1(d) show the two corresponding slices of different treatment images of the same patient, from which we can see the dramatic variations of both intensity and shape of the rectum. This causes distinct appearance in the two treatment images of the same patient. Therefore, it is not easy to segment the prostate by directly registering the previous segmented images of the same patient to the current treatment image. Third, besides the dramatic appearance variations, the unpredicted bowel gas and filling also contribute to the considerably large motion of the prostate relative to the nearby organs at different treatment days, which makes difficult to obtain a good initialization in practice.

To address the above challenges, recently many novel prostate segmentation methods have been proposed. The most popular category of prostate segmentation methods is deformable-model based method.<sup>5–9</sup> Pizer<sup>5</sup> proposed a medial shape model called "m-rep" to simultaneously segment bladder, prostate, and rectum from CT images. Feng *et al.*<sup>8</sup> proposed to combine the gradient profile features and probability distribution function features to guide the model deformation in prostate segmentation. Chen *et al.*<sup>9</sup> incorporated the anatomical constraints in the model deformation to segment the prostate and rectum. Although deformable-model based methods have demonstrated its effectiveness in prostate segmentation by considering both statistically learned shape prior and image appearance, their good performances depend on the good initialization of the deformable model, without which the deformable-model based methods may get stuck in the local minima. However, since the motion of the prostate is unpredictable, it is difficult to have a good initialization of the deformable model in practice.

Another category of methods in prostate segmentation is registration-based method.<sup>10–12</sup> Davis *et al.*<sup>10</sup> combined large deformation image registration with a bowel gas segmentation and deflation algorithm to automatically localize the prostate by registering the treatment image to the planning image. Liao and Shen<sup>12</sup> presented a way to learn an evaluation function which is used to guide the deformable CT prostate registration. Compared with deformable-model based methods, registration-based methods are more robust to the prostate motion. However, its segmentation accuracy is limited by the inconsistent image appearance caused by the uncertainty of bowel gas and filling.

Besides, Li *et al.*<sup>13</sup> formulated the prostate segmentation as a classification problem and proposed to learn the image context information to assist the pixel-wise prostate classification. Haas *et al.*<sup>14</sup> used 2D flood fill with the predefined shape guidance to segment the prostate. Ghosh *et al.*<sup>15</sup> proposed a genetic algorithm which uses prior knowledge in the form of texture and shape for prostate segmentation.

On the other hand, sparse representation as an emerging technique has become the focus of much recent research in machine learning,<sup>16,17</sup> signal processing,<sup>18</sup> and computer vision.<sup>19,20</sup> It has been successfully applied in many fields

such as compressive sensing<sup>21</sup> and face recognition,<sup>22</sup> and has achieved considerable improvements over previous methods in those fields. However, few works have been done to adapt it to solve the segmentation problems. In this paper, we propose to first use sparse representation based classification (SRC) to enhance the prostate in CT images by pixel-wise classification in order to overcome the limitation of poor contrast of the prostate images. Then based on the classification results, previous segmented prostates of the same patient are used as patient-specific<sup>26</sup> atlases to segment the prostate in a multi-atlas-based segmentation scheme. Due to the fact that the prostate shape changes slightly under radiotherapy, rigid transformation is used to align segmented prostates to the current image space based on the classification results, and finally majority voting strategy is adopted to fuse the labels from different aligned atlases. Since the proposed multi-atlas-based segmentation is guided by the classification, the segmentation accuracy highly depends on the classification performance. However, the traditional SRC suffers two main limitations when applied to the pixel-wise classification. First, the traditional SRC cannot be directly adapted to the large-scale problem where the size of training samples is huge. Second, when training samples of different classes are highly correlated, the classification performance of the traditional SRC is limited. In order to overcome these limitations, especially for the purpose of segmentation, we extend SRC from the following four aspects: (1) a discriminant subdictionary learning method is proposed to learn a discriminant and compact representation of training samples for each class in order to increase the discriminant power of SRC and adapt SRC to large-scale pixel-wise classification. (2) The L1 regularized sparse coding is replaced by the elastic net<sup>23</sup> in order to obtain a smooth and clear prostate boundary in the classification result. (3) Residue-based linear regression is incorporated to improve the classification performance and to extend SRC from hard classification to soft classification. (4) Iterative SRC is proposed by using context information to iteratively refine the classification results.

The remainder of the paper is organized as follows. In Sec. II, we briefly review the sparse representation and SRC, as well as address the limitations of SRC when it is applied to pixel-wise classification. Section III introduces the extended SRC that overcomes these limitations. Section IV explains how the extended SRC can be used to guide the multi-atlas-based prostate segmentation. Section V evaluates the contributions of the extended SRC and presents our segmentation results. Finally, we conclude and discuss possible directions of the future research in Sec. VI.

## II. SPARSE REPRESENTATION AND SRC

Sparse representation models data with linear combinations of a few elements from a learned dictionary. Like the traditional data representation methods (e.g., wavelet and Fourier transform), sparse representation method has a set of basis elements, which column-wisely form a dictionary. These basis elements do not need to be orthogonal or pre-defined, which is a main difference from the traditional data

representation methods. Therefore, the dictionary for sparse representation is usually learned through a process called *dictionary learning* so that the learned dictionary can be well tailored with respect to a specific task (e.g., reconstruction and classification). Given a learned dictionary  $\mathbf{D} \in \mathbb{R}^{p \times N}$ , which has  $N$  basis elements of  $p$  dimensions, the goal of sparse representation is to select a few basis elements to best represent the input signal  $\mathbf{x} \in \mathbb{R}^p$ . Mathematically, it can be formulated as the following *sparse coding* problem:

$$\boldsymbol{\alpha}^* = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1, \quad (1)$$

where  $\boldsymbol{\alpha}^* \in \mathbb{R}^N$  is called *sparse representation* or *sparse code* of  $\mathbf{x}$  with respect to the dictionary  $\mathbf{D}$ ,  $\|\boldsymbol{\alpha}\|_1$  is the L1 norm of  $\boldsymbol{\alpha}$ , and  $\lambda$  is a parameter that controls the sparsity of  $\boldsymbol{\alpha}^*$  or the number of nonzero entries in  $\boldsymbol{\alpha}^*$ . The larger the  $\lambda$  is, the sparser the  $\boldsymbol{\alpha}^*$  is and the fewer nonzero entries the  $\boldsymbol{\alpha}^*$  has.

Based on the sparse representation, SRC (Ref. 22) was recently proposed and has achieved the state-of-the-art results in face recognition. In SRC, to classify a new sample, all training samples from different classes are used to represent it in a competitive manner, and the class label is determined by choosing the class that best reconstructs it. Specifically, the training samples belonging to the same class are first column-wisely grouped into *subdictionaries*, which are further combined to form a *global dictionary*  $\mathbf{D} \in \mathbb{R}^{p \times N}$ :

$$\mathbf{D} = [\mathbf{D}_1, \dots, \mathbf{D}_i, \dots, \mathbf{D}_K], \quad (2)$$

$$= [\mathbf{d}_{1,1}, \mathbf{d}_{1,2}, \dots, \mathbf{d}_{i,j}, \dots, \mathbf{d}_{K,N_K}], \quad (3)$$

where  $\mathbf{D}_i$  is the subdictionary of class  $i$ ,  $\mathbf{d}_{i,j}$  is the  $j$ th training sample of class  $i$ ,  $K$  is the total number of classes,  $N_K$  is the total number of training samples in class  $K$ , and  $N$  is the total number of training samples equal to  $\sum_{i=1}^K N_i$ . To classify a new sample  $\mathbf{x} \in \mathbb{R}^p$ , its sparse code  $\boldsymbol{\alpha}^* \in \mathbb{R}^N$  is first computed with respect to the global dictionary  $\mathbf{D}$  according to Eq. (1). Then the residue with respect to each class is calculated:

$$\mathbf{r}_i = \mathbf{x} - \mathbf{D}_i \boldsymbol{\alpha}_i^*, \quad i \in \{1, \dots, K\}, \quad (4)$$

where  $\mathbf{r}_i \in \mathbb{R}^p$  is the residue with respect to class  $i$ , and  $\boldsymbol{\alpha}_i^*$  carries entries of  $\boldsymbol{\alpha}^*$  corresponding to the indices of columns in  $\mathbf{D}$  belonging to  $\mathbf{D}_i$ . Finally, the signal  $\mathbf{x}$  is classified to the class with the minimum L2 residue norm.

As we can see, the global dictionary  $\mathbf{D}$  consists of subdictionaries of different classes, which are simply ensembles of training samples. Although it is effective in face recognition, it has two main limitations when applied to pixel-wise classification. First, medical images are usually 3D volumes that consist of millions of voxels. Considering that the training samples are usually drawn from several medical images and each training sample is represented by a feature vector, the size of training samples can be very large, which makes it impossible to use all of them to form the global dictionary. Second, sparse code is not stable.<sup>24</sup> When the dictionary contains strongly correlated elements, sparse representation tends to arbitrarily pick any of these elements to represent the given sample. In such cases, the performance of SRC will be limited. While it is not a problem in face recognition since faces

of different persons are assumed to be different, which indicates that the strongly correlated elements only exist within classes but not between classes, it is not the case in pixel-wise classification. In medical image segmentation, each pixel usually has to be classified into one of two classes, i.e., object or background. In order to accurately localize the boundary of object of interest, a common strategy is to draw more object and background training samples around the object boundary in the training stage. These object and background samples can be very similar and strongly correlated mainly for two reasons. First, they are spatially close and sometimes just next to each other. Second, in medical images, the object boundary is vague and not distinctive. For voxels near the boundary, usually no effective features can be used to discriminate them as object or background. As a result, the global dictionary in SRC will include strongly correlated elements from different classes, which can inevitably limit the performance of SRC.

### III. EXTENDED SRC

In order to overcome the limitations of SRC in pixel-wise classification for the purpose of image segmentation, we propose to extend SRC from the following four aspects: (1) instead of grouping all training samples into subdictionaries, a discriminant subdictionary learning method is proposed to learn a discriminant and compact representation of training samples for each class. (2) The traditional L1 regularized sparse coding is replaced by the elastic net to stabilize the sparse code and to provide a smooth and clear boundary in the classification result. (3) Linear regression is incorporated to predict the class probability based on the representation residues in order to increase the classification performance and extend SRC from hard classification to soft classifica-

tion. (4) Iterative SRC is further proposed to iteratively refine the classification results based on the context information extracted from the previous classification results. The flow chart of our extended SRC is shown in Fig. 2.

#### III.A. Discriminant subdictionary learning

In pixel-wise classification, it is common for different classes to have similar training samples. In such case, the performance of SRC is limited. Discriminant subdictionary learning aims to learn discriminant subdictionaries so that elements in different subdictionaries are as distinct as possible. In this paper, we propose to combine feature selection with dictionary learning method as a way to learn discriminant subdictionaries. First, feature selection technique is used to select discriminant features so that, after feature selection, training samples of different classes are as distinct as possible. Then dictionary learning method is adopted to learn a compact representation of these discriminant training samples in order to make the size of subdictionary practically feasible. Considering the application of our method to prostate segmentation, here we illustrate the idea only in binary classification. However, the proposed discriminant subdictionary learning can be readily extended to the case of multiclass classification when combined with multiclass feature selection techniques.<sup>25</sup>

In the context of prostate segmentation, each voxel needs to be classified to either prostate or background. Consider that each training sample is represented by a feature vector, which can be intensity or image features, the objective of feature selection is to select features discriminant between prostate and background class, so that the selected discriminant features can be used to readily distinguish prostate voxels from background voxels. In this paper, feature ranking, based

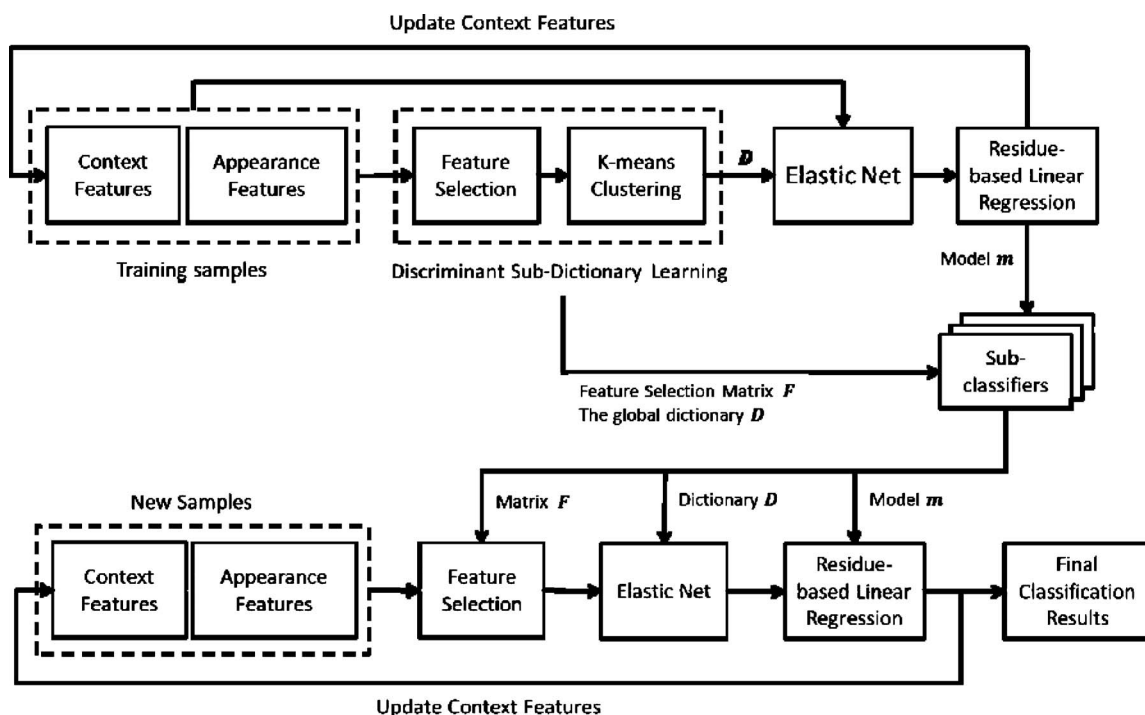


FIG. 2. The flow chart of the extended SRC.

on Fisher separation criterion (FSC),<sup>27</sup> is adopted to select those discriminant features. Specifically, for each feature  $f$  in the feature vector, its FSC score is first computed as below:

$$\text{FSC} = \frac{|\mu_1 - \mu_2|}{\sqrt{v_1 + v_2}}, \quad (5)$$

where  $\mu_1$  and  $\mu_2$  are the sample mean of feature  $f$  in prostate and background classes, respectively, and  $v_1$  and  $v_2$  are the sample variance of feature  $f$  in prostate and background classes, respectively. Features with high FSC scores are considered discriminant and thus selected while features with low FSC scores are discarded in the final feature-based representation.

After feature selection, due to the large size of training samples in pixel-wise classification, it is practically infeasible to directly use them to form subdictionaries. For storage and computational efficiency, it is necessary to adopt a dictionary learning method to learn a compact representation of those discriminant training samples. Nowadays, many dictionary learning methods<sup>18,28,29</sup> have been proposed. However, most of them are reconstruction-oriented, which unfortunately do not consider discriminability in the dictionary optimization. As a result, elements in different independently learned subdictionaries can be similar and highly correlated. In this paper, we use K-means clustering to learn a compact representation of discriminant training samples for each class. Different from reconstruction-oriented dictionary learning methods, K-means does not learn a subdictionary to best represent the given training samples. Instead, it clusters the discriminant training samples and selects their cluster centroids as dictionary elements to form subdictionaries, which can better preserve the discriminant characteristics of training samples, compared to those learned by reconstruction-oriented methods, and thus can lead to better classification performance.

Once subdictionaries of different classes are learned, their columns are first normalized to the unit norm and then put together to form the global dictionary according to Eq. (2) for the later classification.

### III.B. Elastic net

In abdominal CT images, the prostate is of extremely low contrast and its boundary is indistinct. As far as we know, no effective features have been identified to accurately localize the prostate boundary. Therefore, even after feature selection, prostate and background training samples drawn near the prostate boundary can still be quite similar, which inevitably introduces highly correlated elements between subdictionaries. As stated previously, these highly correlated elements can cause sparse coding instable, especially when samples to classify are similar to those highly correlated elements. Specifically, during sparse coding, when several highly correlated elements are available, only one of them tends to be selected in the final sparse representation. Due to the existence of noise, this selection is quite sensitive, which means that samples with similar features can have distinct sparse codes and thus be classified into different classes due to the small noises. As

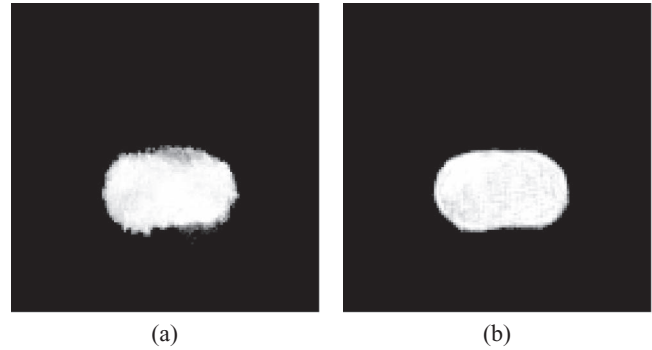


FIG. 3. (a) Zigzag prostate boundary caused by the L1 regularized sparse coding and (b) smooth prostate boundary by adopting the elastic net in SRC.

a result, the classification of prostate boundary voxels in the new treatment image is instable because their feature vectors are considered similar to the highly correlated elements between two subdictionaries. Figure 3(a) gives a typical classification result of our extended SRC using the L1-regularized sparse coding. As we can see, the prostate boundary is zigzag and unclear, which justifies our statement.

To address this problem, we replace the traditional L1-regularized sparse coding with the elastic net,<sup>23</sup> which compromises between sparsity and stability. Instead of only using the L1 constraint to regularize the least square problem, the elastic net balances between L1 constraint and L2 constraint:

$$\alpha^* = \underset{\alpha}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{D}\alpha\|_2^2 + \lambda_1 \|\alpha\|_1 + \frac{\lambda_2}{2} \|\alpha\|_2^2. \quad (6)$$

As we know that the solution of the L2-regularized least square problem is stable. Thus, adding L2 regularization helps stabilize the sparse code. Besides, the elastic net encourages a grouping effect,<sup>23</sup> where strongly correlated elements tend to be selected together during the sparse coding, which can reduce the classification error caused by the instability of the L1-regularized sparse coding when the across-subdictionary correlations are less than within-subdictionary correlations. Therefore, we propose to use the elastic net to replace the traditional L1-regularized sparse coding in pixel-wise classification where there exist highly correlated elements in subdictionaries of different classes. Practically, we found that boundary-smoothing effects can be achieved by stabilizing the sparse code. Figure 3 visually compares the classification results of the L1 regularized sparse coding and the elastic net.

### III.C. Residue-based linear regression

In the traditional SRC, residue norms with respect to each class are compared and the new sample is classified to the class with the minimum residue norm. In such cases, residues of different features are equally treated. While it is reasonable when features are of the same type and importance, it is not desirable in other cases. Usually each voxel is represented by the combination of different types of features, thus the discriminabilities of individual features are different and their contributions to classification are also different. Therefore, equally weighting them in determining the class label limits the classification performance. Besides, the traditional SRC

is a hard classification method, which only assigns a class label to the new sample. In contrast, soft classification provides more quantitative information, especially in the decision margin where the class membership is unclear. Based on these observations, we propose to learn a linear regression model to predict the class probability based on the residues, which extends SRC from hard classification to soft classification.

Specifically, in the training stage, discriminant subdictionary learning is first applied to select the  $\hat{p}$  topmost discriminant features from a set of  $N$  training samples represented by their original  $p$ -dimensional feature vectors, and to learn two discriminant subdictionaries. These learned subdictionaries are further combined into a global dictionary  $\hat{\mathbf{D}} \in \mathbb{R}^{\hat{p} \times M}$ , where  $\hat{p}$  ( $< p$ ) is the number of selected discriminant features and  $M$  ( $< N$ ) is the size of dictionary  $\hat{\mathbf{D}}$ . Then, for each sample represented by its selected discriminant features, its sparse code with respect to  $\hat{\mathbf{D}}$  can be computed using Eq. (6), and its object (prostate) residue  $\mathbf{r}_{obj}^i \in \mathbb{R}^{\hat{p}}$  and background residues  $\mathbf{r}_{bk}^i \in \mathbb{R}^{\hat{p}}$  can be calculated by following Eq. (4), respectively. Finally, these object and background residues can be used together with their class labels to learn a linear regression model by solving the regularized least square problem in Eq. (7). Since nondiscriminant features have already been filtered out during the discriminant subdictionary learning procedure, all the retained features are considered as discriminant and can have contribution in prediction of class probability. Therefore, in our linear regression model, we do not perform any feature selection and simply use L2 constraint as a regularization term:

$$\min_{\mathbf{m}} \sum_i^N \|\mathbf{l}_i - \mathbf{m}^T \mathbf{r}_i\|_2^2 + \gamma \|\mathbf{m}\|_2^2, \quad (7)$$

where  $\mathbf{l}_i \in \{-1, 1\}$  is the class label of the  $i$ th training sample,  $\mathbf{r}_i = [\mathbf{r}_{obj}^i, \mathbf{r}_{bk}^i]^T$  is the combined residue of the  $i$ th training sample,  $\mathbf{m} \in \mathbb{R}^{2\hat{p}}$  is the linear coefficient vector of the linear regression model, and  $\gamma$  is the weight for regularization.  $\gamma$  can be set to 0 when the number of training samples is sufficiently large. To classify a new signal, its combined residue  $\mathbf{r}_{new}$  is first computed by following the same pipeline in the training stage. Then, the class probability is estimated by a truncated linear mapping:

$$P(x) = \begin{cases} 0, & \mathbf{m}^T \mathbf{r}_{new} \leq -1, \\ 1, & \mathbf{m}^T \mathbf{r}_{new} \geq 1, \\ \frac{\mathbf{m}^T \mathbf{r}_{new} + 1}{2}, & \text{otherwise.} \end{cases} \quad (8)$$

By incorporating the residue-based linear regression into SRC, full residual information is used, instead of just using their norm. Besides, individual features are weighted by their contributions in predicting the class probability. Compared with the traditional SRC, we found the classification performance can be increased by using residue-based linear regression, which will be shown in Sec. V.

### III.D. Iterative SRC

Segmentation by classification is often criticized by not considering the spatial regularization because each pixel is independently classified. Due to the classification error, there can be isolated object pixels in the background or vice versa. Recently Tu proposed the autocontext model,<sup>30</sup> which uses context information to iteratively refine the classification results. Specifically, at each classification iteration, previous classification results at context locations are extracted as context features to assist the classification in the current iteration. So, for each pixel, it can be represented by a feature vector that contains both its original features and the context features, which are updated iteratively. As the classification iterates, these context features become more discriminative and thus more helpful in the classification. As a result, the classification probability map becomes clearer and clearer. Inspired by this idea, we incorporate the context information into SRC and propose the iterative SRC. In this paper, for a pixel  $(x_{pt}, y_{pt})$ , its context positions are defined as follows:

$$\begin{cases} x_{u,v} = x_{pt} + \text{sgn}(R_u \cos(\frac{\pi}{4}v)) \times \lfloor |R_u \cos(\frac{\pi}{4}v)| \rfloor \\ y_{u,v} = y_{pt} + \text{sgn}(R_u \sin(\frac{\pi}{4}v)) \times \lfloor |R_u \sin(\frac{\pi}{4}v)| \rfloor \end{cases}, \quad (9)$$

where  $(u, v)$  are two polar coordinates indexing the context locations,  $R_u$  is the radius indexed by  $u$ ,  $\text{sgn}$  is the sign function,  $|\cdot|$  gives the absolute value, and  $\lfloor \cdot \rfloor$  gives the floor of a real number. In this paper, we set  $R_u \in \{4, 5, 6, 8, 10, 12, 14, 16, 20, 25\}$  and  $v \in \{0, 1, 2, 3, 4, 5, 6, 7\}$ . Figure 4(d) gives an illustration of context locations of the center pixel in the image.

In the iterative SRC, initially we start with a uniform probability map since no classification has been performed. Due to lack of discriminability, these context features are filtered out by discriminant subdictionary learning, which means in the first classification iteration the context features are not involved. In the later classification iterations, the context features are iteratively updated and encode more and more accurate class probability information about its surrounding pixels, which can be considered as effective high-level features. As a result, more context features are identified as the topmost discriminant features in the discriminant subdictionary learning and thus selected to guide the refinement of the classification results. Figure 4 shows a typical classification probability map at different iterations, which clearly justifies the effectiveness of the iterative SRC. More results will be presented in Sec. V.

## IV. PROSTATE SEGMENTATION VIA EXTENDED SRC

In this section, we describe how the extended SRC can be applied to guide the multi-atlas-based prostate segmentation. The main motivation of using SRC is that we believe patches repeat not only spatially but also longitudinally. Therefore, in prostate segmentation, patches in the new treatment image likely have appeared in the previous treatment images or the planning image. If we build two discriminant patch-based subdictionaries for prostate and background using previous

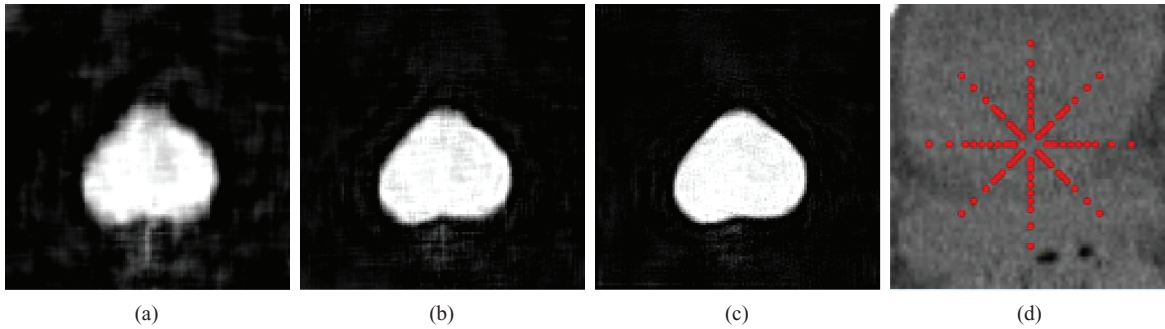


FIG. 4. (a)–(c) The classification results in the first, second, and third iteration, respectively. Points in (d) show the context locations of the center pixel in the image.

images, for a patch in the new treatment image, it tends to draw more supports from the respective subdictionary in the sparse representation. Based on the representation residues corresponding to each class, we can estimate class probability of the voxel associated with this patch.

Our proposed method requires two treatment images to be manually segmented in order to fully capture the patient-specific appearance variations. In the later treatment days, physicians only need to manually identify the middle slices along two coordinate directions in order to initialize the method. Compared with the fact that currently physicians have to manually segment up to 40 treatment images for each patient, the effort to initialize our method is minor.

The whole section is divided into four subsections. In the Subsection IV.A, we briefly introduce the types of features used in our method. Subsection IV.B presents the preprocessing steps before prostate segmentation. Subsection IV.C explains the training stage of our method, which learns two sets of location-adaptive classifiers along two coordinate directions. The testing stage is described in the Subsection IV.D where we show how these learned location-adaptive classifiers can be applied to guide the prostate segmentation in the current treatment image. Figure 5 illustrates the overall pipeline of our prostate segmentation method.

#### IV.A. Types of features

In this paper, two types of appearance features are used, which are nine-dimensional histogram of oriented gradients (HOG) features<sup>31</sup> and 23 normalized Haar features. Each voxel is represented by the combination of local appearance features and context features. The local appearance features consist of nine-dimensional HOG features and 23 normalized Haar features computed in a  $21 \times 21$  local window. Different from the autocontext model, the context features used in this paper are not the only class probabilities extracted at context locations from previous classification results. At each context location, the same types of appearance features are also extracted in order to fully capture the context appearance information surrounding the pixel under study.

#### IV.B. Preprocessing

In the preprocessing stage, all treatment images are rigidly aligned onto the planning image space based on the pelvic bones, which are segmented by simple thresholding. This alignment process is performed by the FLIRT toolkit.<sup>32</sup> Based on the manually delineated prostate in the planning image, we compute the centroid of the prostate and define a  $128 \times 128$

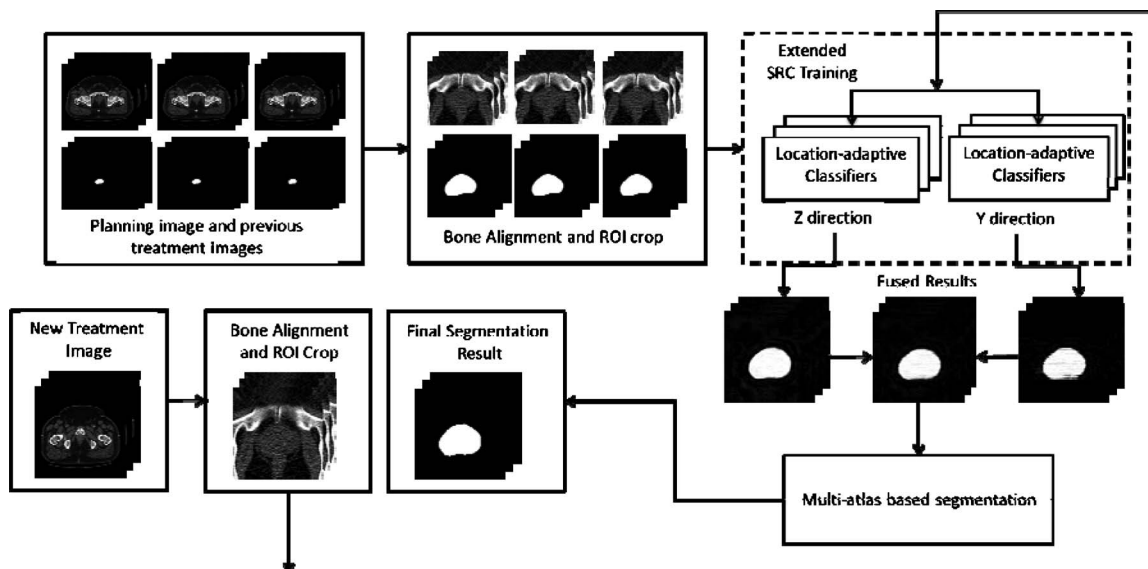


FIG. 5. The flow chart of our prostate segmentation method using the extended SRC.

$\times 30$  ROI centered at the centroid, which is used to crop all images in order to reduce the computational burden.

#### IV.C. Training stage

In the training stage, prostate and background training samples are first randomly drawn from the previous images of the same patient. Specifically, 3/5 prostate training samples and 1/2 background training samples are drawn near the prostate boundary in order for the learned classifiers to accurately localize the prostate boundary. The rest prostate training samples are taken from the prostate interior, and the rest background training samples are taken from pelvic bone, bowel gas, and other background regions. In order to utilize 3D image information and increase the robustness of classification, two sets of location-adaptive classifiers are learned along the anterior–posterior ( $y$ ) direction and the superior–inferior ( $z$ ) direction, respectively, because slices along these two directions contain richer context information (e.g., pelvic bone) than those along the lateral ( $x$ ) direction. Initially, two manually segmented treatment images as well as the planning image of the same patient are used in the training. As the number of segmented treatment images increases, only the latest five treatment images are used as the training images, which accounts for the tissue appearance change under radiation treatment.

During the training, in each direction, every three slices are grouped into sections. Based on the segmented prostate of the training images, we can know the section-to-section correspondence between training images. For each set of corresponding sections, a location-adaptive classifier is learned according to the extended SRC using the prostate and background training samples drawn from those sections. Here a location-adaptive classifier consists of a sequence of subclassifiers. Each subclassifier is a collection of a feature selection matrix obtained from discriminant subdictionary learning, a global dictionary formed by two discriminant subdictionaries, and a residue-based linear regression model. Each one is responsible for one classification iteration and takes the classification results of the previous subclassifier as input. Therefore, for each direction, we can learn a set of location-adaptive classifiers, which takes the variability of different prostate regions into account. In total, two sets of location-adaptive classifiers are learned during the training stage.

#### IV.D. Testing stage

In the testing stage, the physician first needs to manually identify the middle slices along two coordinate directions so that rough correspondences between the location-adaptive classifiers and slices can be established. Then, based on the established rough correspondences, each slice is pixel-wisely classified by its corresponding classifier. Finally, the classification results of all slices along one direction are stacked to form a 3D prostate probability map. Since the classification is performed in both anterior–posterior ( $y$ ) and superior–inferior ( $z$ ) directions, two prostate probability maps can be obtained, which are further averaged to produce a fused prostate probability map. Once we have a fused prostate probability map, all

previous segmented prostate images of the same patient can be aligned onto this map and then majority voting strategy can be adopted to obtain the final prostate segmentation result in the current treatment image.

### V. EXPERIMENTAL RESULTS

In this section, several experiments have been conducted to evaluate the contributions of different components of our method. Our dataset consists of 24 patients with total 330 CT images. Each patient has more than nine daily CT scans. The image size of axial slices is  $512 \times 512$  with voxel size  $1 \text{ mm} \times 1 \text{ mm}$ . The interslice distance is 3 mm. The manual segmentation results provided by an experienced clinical expert are available for each CT image to serve as the ground truth. Five quantitative measures are used to evaluate the performance of our method and compare with other state-of-the-art prostate segmentation methods.

*Dice similarity coefficient* (DSC) (Ref. 33) is a comprehensive set similarity measure that is widely used to evaluate the performance of segmentation methods. It is defined as  $(2 \times |V_s \cap V_g|) / (|V_s| + |V_g|)$ , where  $V_s$  and  $V_g$  are the sets of object (prostate) voxels automatically segmented by the segmentation method and manually segmented by clinical expert, respectively, and  $|\cdot|$  returns the cardinality of a set. It should be noted that the DSCs in all of our experiments are calculated based on the whole 3D prostate volume, instead of a particular slice.

*Centroid distance* (CD) measures the distance between the centroid of the automatic segmentation result and that of the ground truth. The centroid is defined as the mean position of all object (prostate) voxels. The differences along the lateral ( $x$ ) direction, the anterior–posterior ( $y$ ) direction, and the superior–inferior ( $z$ ) direction are measured independently.

*Average surface distance* (ASD) is calculated as the average distance between the surface of the automatic segmented prostate and that of the ground truth along  $360 \times 180$  rays evenly distributed in a sphere originated from the centroid of the ground truth.

*True positive rate* (TP) is the percentage of the ground truth that overlaps with the automatic segmentation result.

*False positive rate* (FP) is the percentage of the automatic segmentation result that lies outside the ground truth.

#### V.A. Parameters

To determine the values of  $\lambda_1$  and  $\lambda_2$  in the elastic net, we tested several combinations with  $\lambda_1 \in \{0.01, 0.05, 0.1, 0.2\}$  and  $\lambda_2 \in \{0.01, 0.05, 0.1, 0.2\}$ . We found that, as long as the ratio  $\lambda_1/\lambda_2$  is fixed, the classification results are not sensitive to any particular combination of  $\lambda_1$  and  $\lambda_2$ . Thus,  $\lambda_1$  and  $\lambda_2$  are both set to be 0.1 in our experiments.

The number of features selected in the discriminant subdictionary learning is 200, based on the feature dimension analysis as shown in Fig. 6. The dictionary size for each subdictionary is 800, which is the number of clusters used in the K-means clustering.

The parameter  $\gamma$  in Eq. (7) is set to be 0. Because the number of training samples used is often over 10 000, which is



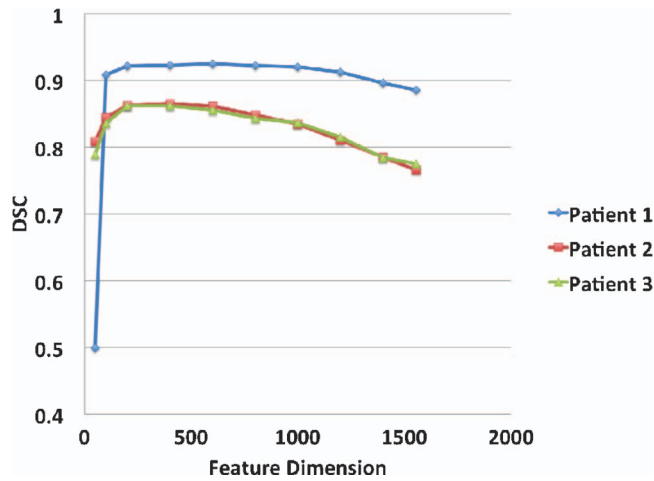


FIG. 6. The performance versus feature dimension plot. DSCs are calculated by binarizing the classification results using 0.5 threshold.

sufficient to prevent the overfitting problem. Besides, in prostate segmentation, totally three classification iterations are used to iteratively refine the classification results.

### V.B. Importance of using our classification method in prostate segmentation and its sensitivity to the middle slice identification

Before we start to evaluate our proposed extended SRC, we first studied the importance of using our classification method in prostate segmentation and also its sensitivity to the middle slice identification. To evaluate the importance of using our classification method in prostate segmentation, we compared it with a simple translation-based method, which translates the patient-specific atlases to match the two manually identified middle slices and adopts the majority voting strategy to do the label fusion. The mean DSC obtained by the simple translation-based method is  $0.685 \pm 0.204$ , which is significantly ( $p < 0.0001$ ) worse than our method that can obtain  $0.913 \pm 0.045$  mean DSC in our dataset. Besides, the simple translation-based method heavily relies on the accuracies of two identified middle slices. In practice, physicians may introduce errors in identifying the middle slices, which may affect the segmentation performance. But, as we will show below, our method is quite robust to the middle slice identification.

To test the sensitivity of our method to the manually identified middle slices, we run our method using different middle slice positions. Specifically, the middle slices along  $y$  direction and  $z$  direction both vary within six slices from the true middle slices identified by the ground truth. In total, 169 ( $13 \times 13$ ) combinations of different middle slice positions have been tested. The experiment is performed using the 12 CT images of patient 1, and the mean DSCs of different middle slice configurations are shown in Fig. 7. It should be noted that the slice thickness along the superior–inferior ( $z$ ) direction is 3 mm and the number of total slices in patient 1 along  $z$  direction is 17. From Fig. 7, we can see that our method can achieve above 0.90 mean DSC as long as the error of middle slice identification is within two slices (6 mm) in  $z$  direction.

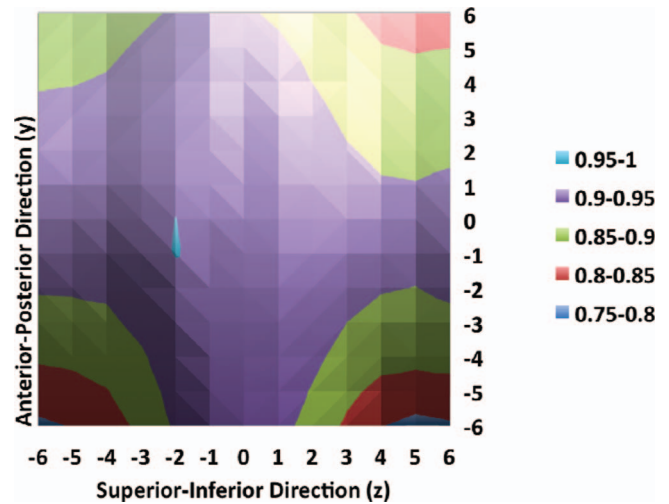


FIG. 7. The mean DSCs of patient 1 using different middle slice configurations.

Even in the extreme case where the error of the middle slice identification in  $z$  direction is 18 mm, equal to the six-slice thickness, our method can still achieve over 0.90 mean DSC as long as the error in the other direction is not large. This is because we use the classification fusion strategy to guide the segmentation, which can incorporate the complementary information from both directions. That is, in case that the classification fails in one direction, the classification in the other direction can assist and help improve the classification quality. Therefore, our method is quite robust to the middle slice identification.

### V.C. Role of K-means clustering in discriminant subdictionary learning

To justify our statement that K-means performs better than reconstruction-oriented dictionary learning methods in discriminant subdictionary learning, we compared it with the K-SVD algorithm,<sup>18</sup> which is a popular reconstruction-oriented dictionary learning method that has been successfully applied in many fields.<sup>34,35</sup> It learns a dictionary to best represent a set of training samples by minimizing the following energy function:

$$\min_{D, A} \|X - DA\|_F^2, \quad \text{subject to } \forall i, \|\alpha_i\|_0 \leq T_0, \quad (10)$$

where  $D$  is the dictionary to learn,  $X$  is the training sample matrix where the  $i$ th column corresponds to the  $i$ th training sample,  $A$  is the sparse code matrix where the  $i$ th column  $\alpha_i$  denotes the sparse code of the  $i$ th training sample,  $\|\cdot\|_F^2$  is the Frobenius norm of the matrix,  $\|\cdot\|_0$  is the informal L0 norm which counts the number of nonzero entries, and  $T_0$  is the sparsity which restricts the maximum number of nonzero entries in the sparse code.

In this experiment, K-SVD and K-means are separately used as the dictionary learning method in our discriminant subdictionary learning. Their classification results along the superior–inferior ( $z$ ) direction are compared based on DSC, to see which performs better in the terms of discriminant subdictionary learning. The experiment was conducted

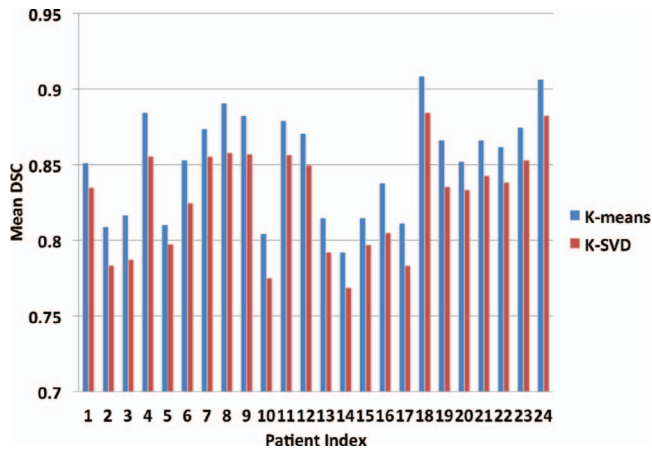


FIG. 8. The mean DSCs of 24 patients obtained using K-means and K-SVD, respectively.

on 330 images of all 24 patients. Except the discriminant subdictionary learning, all other parts are the same with the traditional SRC so that the influence from other contributions of our method can be excluded. Specifically, in our method, both the elastic net and residue-based linear regression are not included to replace the corresponding parts in the traditional SRC. Besides, the classification only takes one iteration so that the context information does not affect the comparison. To maximize the performance of K-SVD, orthogonal matching pursuit (OMP) algorithm is adopted to solve the sparse coding problem in SRC since it is internally used by K-SVD as the sparse coding solver. The sparsity of OMP (the maximum number of nonzero entries allowed) is chosen to be 20, which is empirically proven to be sufficiently good, considering both classification performance and computational efficiency. The overall mean DSC obtained by K-SVD and K-means are  $0.825 \pm 0.050$  and  $0.850 \pm 0.057$ , respectively. The mean DSCs of all 24 patients using K-SVD and K-means are shown in Fig. 8. As mentioned previously, K-means can preserve the discriminability of the training samples during the dictionary learning process, while K-SVD cannot. This renders the classification performance using K-means in the discriminant subdictionary learning is significantly ( $p < 0.0001$ ) better than that using K-SVD.

#### V.D. Elastic net versus L1 regularized sparse coding

In this subsection, we justify our statement that the elastic net outperforms the traditional L1 regularized sparse coding when subdictionaries of different classes contain highly correlated elements. Figure 13(a) shows a typical dictionary element correlation matrix between prostate and background subdictionaries learned in the first iteration. The entry in the  $i$ th row and the  $j$ th column corresponds to the correlation between the  $i$ th column in background subdictionary and the  $j$ th column in prostate subdictionary, which is calculated as their dot product. It can be clearly seen that even after discriminant dictionary learning, there are still many highly correlated elements between two subdictionaries. As explained previously, in such case, the performance of SRC is

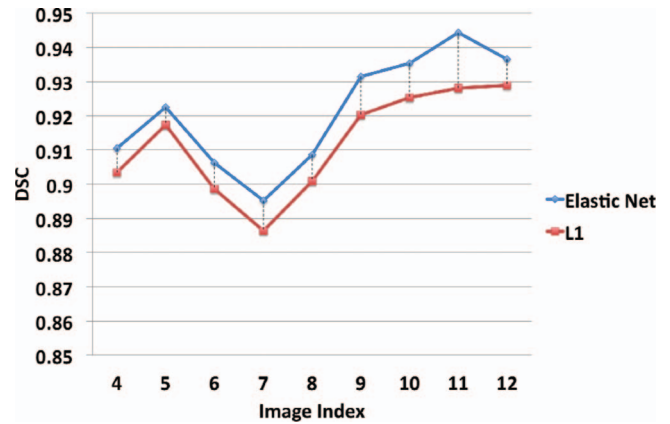


FIG. 9. The DSCs of final classification results of patient 1 using the elastic net and the L1 regularized sparse coding, respectively.

limited. The elastic net can stabilize the sparse code, which selects and deselects highly correlated elements together in the sparse representation. So when a voxel to classify finds its similar elements in both prostate and background subdictionaries, instead of selecting one of them in the sparse representation, which may cause serious classification error due to the existence of noises, the elastic net tends to be cautious and selects all of them, which results in an informative classification probability that is useful for later refinements.

In this experiment, we evaluate the contribution of the elastic net in the context of iterative SRC since we believe that the prostate probability map obtained by the elastic net can better guide the classification refinement than that by the L1 regularized sparse coding. The elastic net and the L1 regularized sparse coding are separately used in our extended SRC. Their final classification results along the superior–inferior ( $z$ ) direction are compared based on DSC. SPARSE Modeling Software (SPAMS) toolbox<sup>36</sup> is adopted to solve both the L1 regularized sparse coding and the elastic net. Figure 9 shows the DSCs of final classification results of patient 1 using the elastic net and the L1 regularized sparse coding, respectively. From the figure, we can see that better classification results are obtained using the elastic net. Figure 10 visually compares several typical classification results between the elastic net and the L1 regularized sparse coding, from which we can see that, by adopting the elastic net, smoother and clearer boundaries can be achieved in the final classification results for different prostate regions (e.g., apex, central, and base regions).

#### V.E. Residue-based linear regression versus residue-norm-based classification

In this experiment, we show that by adopting residue-based linear regression the classification performance can be improved. Since the classification results of residue-based linear regression are probabilistic, we first binarize the results using 0.5 threshold and then compare them with residue-norm-based classification results. Only one classification iteration is used to exclude the influence of context information.

Figure 11 displays the mean DSCs of all 24 patients obtained using residue-norm-based classification (RN) and

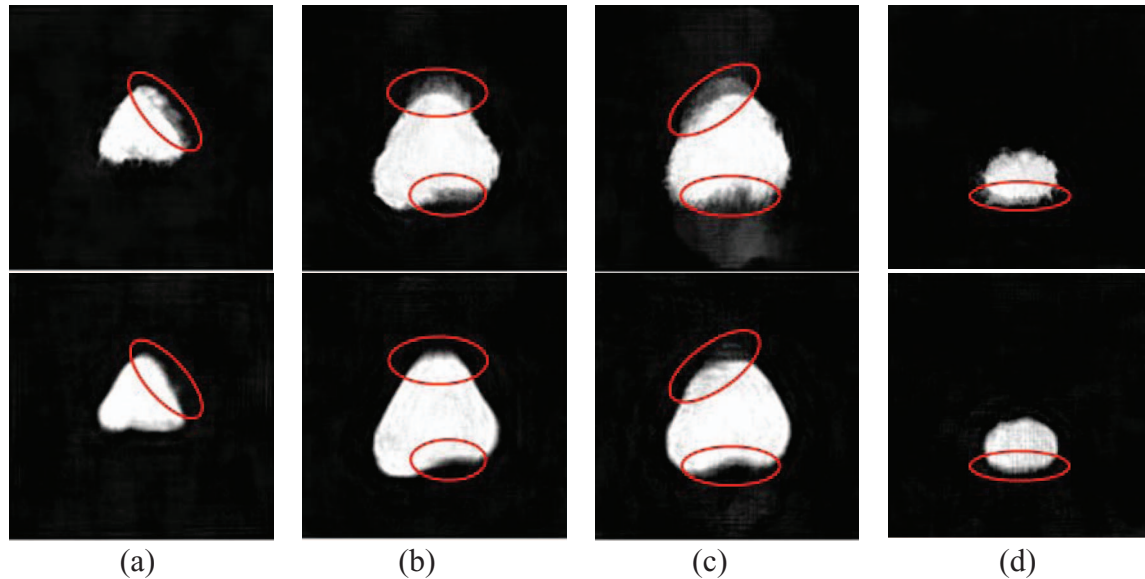


FIG. 10. The first row shows several typical classification results using the L1 regularized sparse coding in different prostate regions. The second row shows the corresponding classification results using the elastic net. The major differences between the classification results of the L1 regularized sparse coding and the elastic net are highlighted by circles. (a) Apex slice; (b) central slice; (c) central slice; (d) base slice.

residue-based linear regression (RBLR), respectively. From the figure, we can find that the classification performance is increased in almost all patients by adopting residue-based linear regression. Only in patient 5, the classification performance shows a noticeable decrease. That is because our classification method fails in one treatment image of patient 5. In such case, residue-based linear regression can lead to worse classification performance than residue-norm-based classification. Figure 12 visually compares several final classification results in different prostate regions (e.g., apex, base, and central part) using residue-based linear regression and residue-norm-based classification, respectively, from which we can see that by considering different contributions of features in regression and weighting their residues accordingly, better segmentation results can be achieved, especially at voxels near the prostate boundary which are considered the most difficult to segment. Statistically, we find the classification

performance by adopting residue-based linear regression is significantly ( $p < 0.0001$ ) better than that using residue-norm-based classification.

#### V.F. Iterative SRC versus single-iteration SRC

This subsection evaluates the iterative classification scheme in the extended SRC and shows how the classification results get refined during iterations. In the iterative SRC, the context features extracted from previous classification results are incorporated to gradually increase the discriminatory power of two subdictionaries. Figure 13 shows typical dictionary element correlation matrices in successive three iterations, from which we can see that the correlations between elements of prostate and background subdictionaries considerably decrease during iterations. This indicates that two subdictionaries become more and more discriminant. As the across-subdictionary correlations decrease, the performance of SRC increases.

Figure 14 gives two additional examples to illustrate the classification refinements during three iterations in the apex and central regions of the prostate. In the first iteration, without the assistance of context information, it is difficult to accurately classify all voxels, especially those located near the prostate boundary. So the classification maps contain many misclassifications and the prostate boundaries are also not clear. However, as more context information becomes available, the classification results get obviously refined and the prostate boundary becomes clearer and clearer. This shows the effectiveness of context information in classification. Figure 15 quantitatively compares the classification performance using one-iteration and three-iteration classifications. As we can see, classification performance can be improved by taking into account the context information.

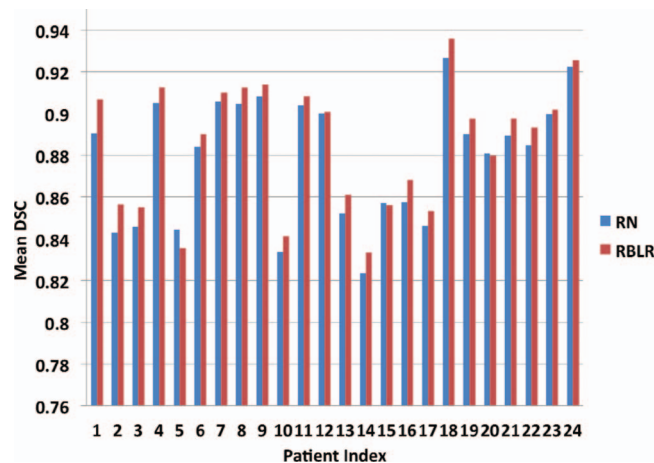


FIG. 11. The mean DSCs of 24 patients after one classification iteration using RN and RBLR, respectively.

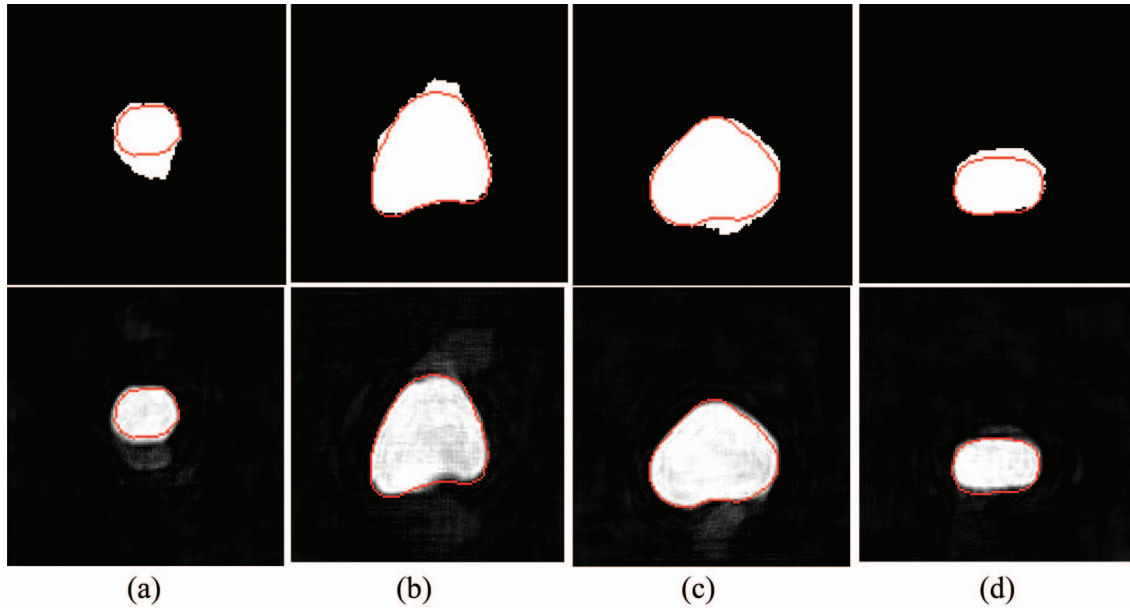


FIG. 12. The first and second rows show several final classification results in different prostate regions overlaid with manually delineated prostate contours using residue norm and residue-based linear regression, respectively. (a) Apex slice; (b) central slice; (c) central slice; (d) base slice.

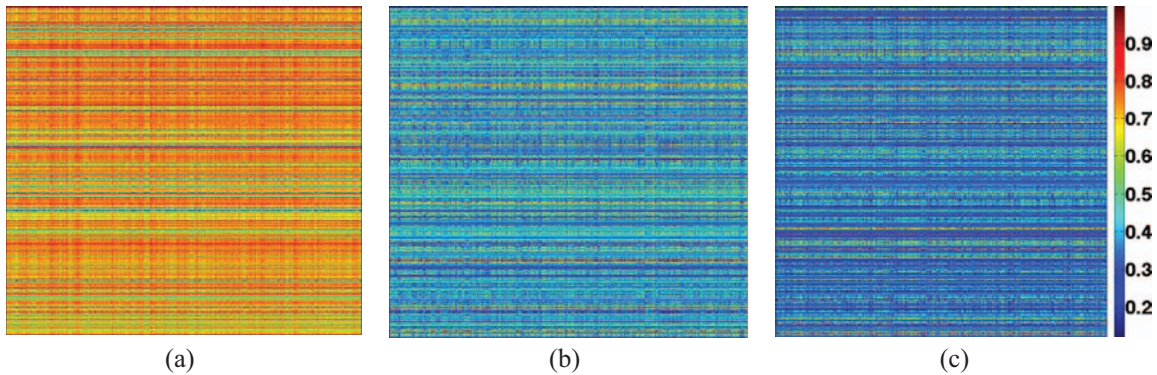


FIG. 13. The dictionary element correlation matrices between prostate and background subdictionaries in successive three classification iterations.

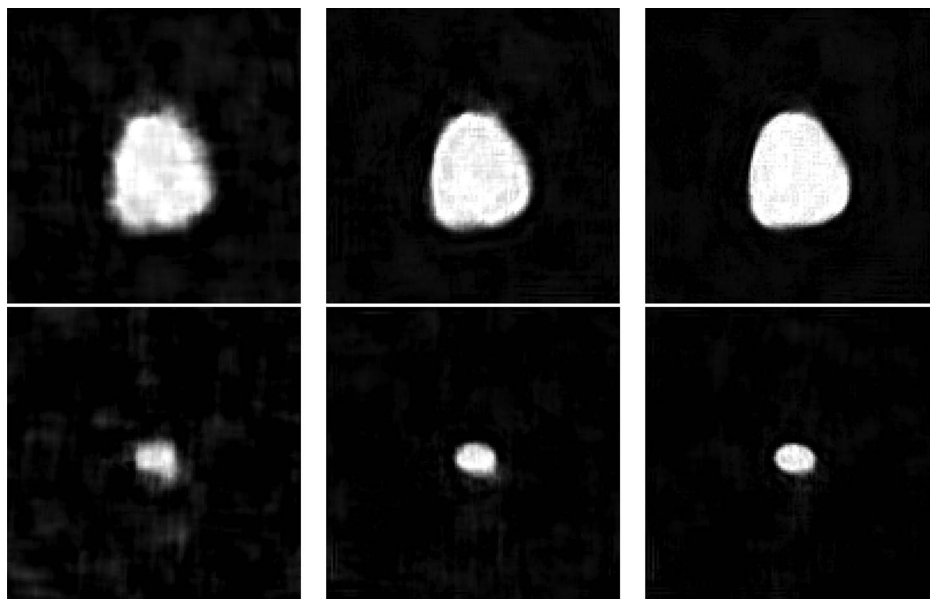


FIG. 14. Three columns show classification results in the first, second, and third iterations, respectively. The first and the second rows show typical classification refinements in the central slice and the apex slice of the prostate, respectively.

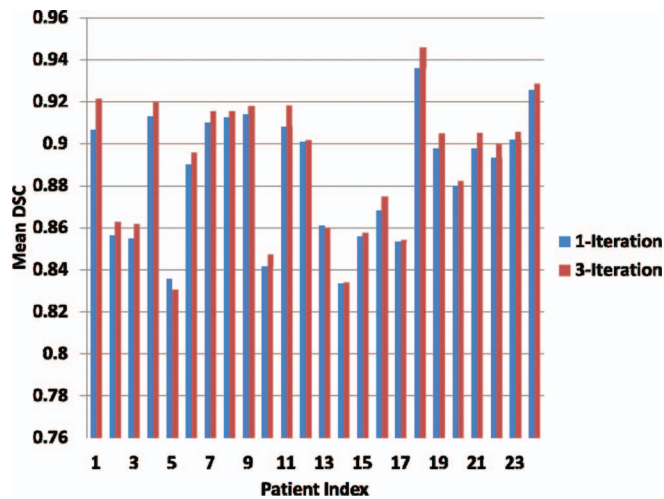


FIG. 15. The mean DSCs of 24 patients using one-iteration and three-iteration classifications, respectively.

### V.G. Multi-atlas-based segmentation

In this subsection, we evaluate multi-atlas-based segmentation and purely classification-based segmentation. In multi-atlas-based segmentation, we first use the proposed extended SRC to do the classification, then rigidly align the patient-specific atlases to the target image based on the matching with the classification result, and finally adopt the majority voting strategy to fuse the labels from different aligned atlases. In the purely classification-based segmentation, we simply use 0.5 as a threshold to binarize the classification map for segmentation. We compared two strategies using the classification map obtained along the superior–inferior ( $z$ ) direction, and used DSC to measure their performances. Figure 16 displays the mean DSCs of all 24 patients between these two strategies. Obviously, by using multi-atlas-based segmentation, which takes patient-specific shape priors into account, irregular prostate shapes can be avoided in the final segmentation result, and thus the segmentation accuracy can be largely improved, especially for the cases where the classification does not do a good job (e.g., patients 5, 13, 14, and 17).

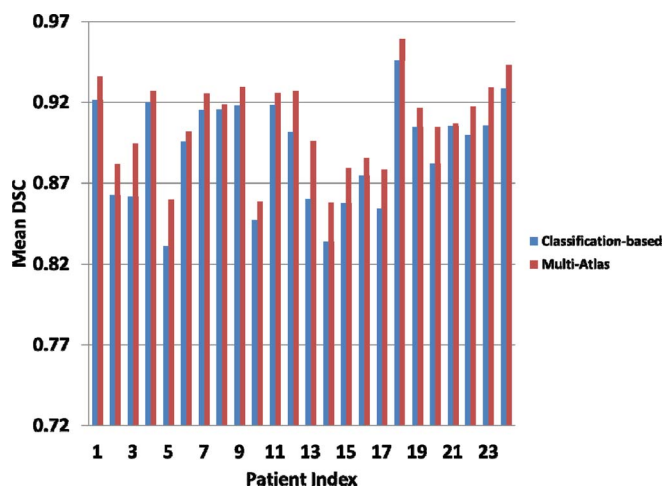
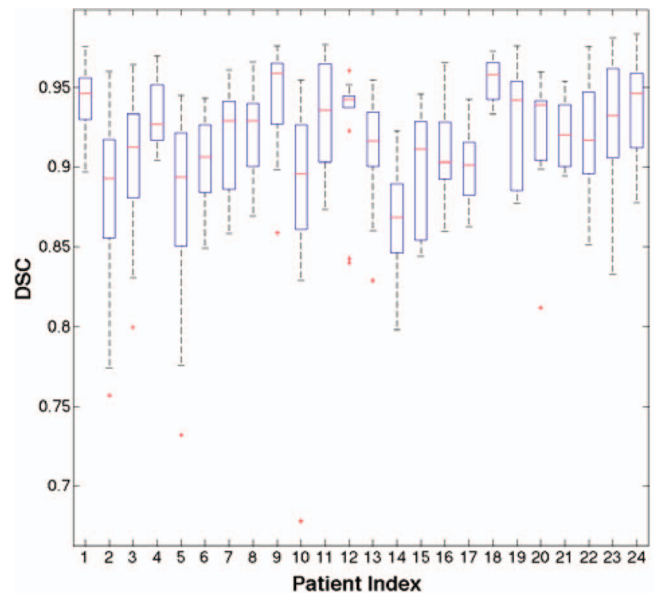


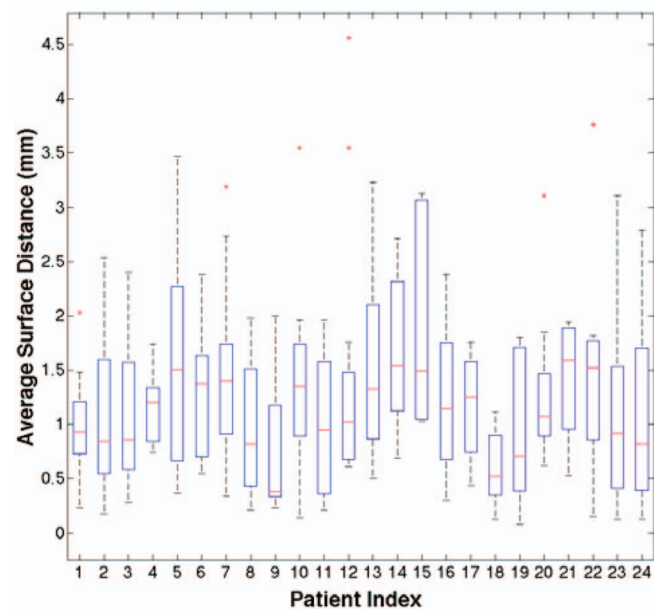
FIG. 16. The mean DSCs of 24 patients using purely classification-based segmentation and multi-atlas-based segmentation, respectively.

### V.H. Comparison with other state-of-the-art methods

To demonstrate the effectiveness of our method in prostate segmentation, five state-of-the-art prostate segmentation methods are used to compare with our method, including two deformable-model based methods,<sup>8,9</sup> two registration based methods,<sup>10,12</sup> and one classification based method.<sup>13</sup> The best quantitative results reported in their works are adopted in the comparison. Figure 17 shows the box-and-whisker diagram of DSC and ASD measurements of our



(a)



(b)

FIG. 17. (a) and (b) The box-and-whisker diagram of DSC and ASD measurements of our method on 24 patients, respectively. The bottom and top of each box are the 25th and 75th percentile, respectively. The band near the middle is the 50th percentile. The whiskers extend to the most extreme data points not considered as outliers. Outliers (marked as crosses) are defined as data measurements outside the range  $[q_1 - 1.5(q_3 - q_1), q_3 + 1.5(q_3 - q_1)]$ , where  $q_1$  and  $q_3$  are the 25th and 75th percentiles, respectively.

TABLE I. The quantitative comparison of our method with other five state-of-the-art prostate segmentation methods based on various measurements. NA in the table means the corresponding measurement was not reported. The best performance of each measurement among comparison is shown in bold letter.

Category Method	Deformable models		Registration methods		Classification methods	
	Feng <i>et al.</i> (Ref. 8)	Chen <i>et al.</i> (Ref. 9)	Davis <i>et al.</i> (Ref. 10)	Liao <i>et al.</i> (Ref. 12)	Li <i>et al.</i> (Ref. 13)	Our method
Subject no.	24	13	3	10	11	24
Image no.	330	185	40	163	161	330
Mean DSC (%)	89.3 ± 5.0	NA	82.0 ± 6.0	89.0 ± 2.0	90.8 ± NA	<b>91.3 ± 4.5</b>
Median DSC (%)	90.6	NA	84.0	90.0	NA	<b>92.1</b>
Median TP	NA	0.84	NA	NA	0.90	<b>0.92</b>
Median FP	NA	0.13	NA	NA	0.10	<b>0.08</b>
Mean ASD (mm)	2.08 ± 0.79	NA	NA	NA	1.4 ± NA	<b>1.24 ± 0.77</b>
Median ASD (mm)	1.87	<b>1.10</b>	NA	NA	NA	1.14
Mean CD, <i>x</i> (mm)	NA	NA	-0.26 ± 0.6	NA	0.18 ± NA	<b>0.02 ± 0.61</b>
Mean CD, <i>y</i> (mm)	NA	NA	0.35 ± 1.4	NA	<b>0.02 ± NA</b>	-0.05 ± 1.53
Mean CD, <i>z</i> (mm)	NA	NA	0.22 ± 2.4	NA	0.57 ± NA	<b>0.14 ± 1.84</b>

method on 24 patients. In most cases, our method achieves over 0.9 DSC and about 1 mm ASD, which is one pixel width. Table I quantitatively compares our method with other state-of-the-art methods based on various measurements. From the table, we can observe that our method achieves the best performance on almost all measurements. Although the median ASD of our method is slightly larger than that of Chen *et al.*,<sup>9</sup> the median TP and FP of our method are much better than theirs. In the mean centroid comparison, the mean CD along

*y* direction of our method is slightly worse than that of Li *et al.*,<sup>13</sup> but in other two directions our method achieves more accurate estimation. Moreover, compared with Li's method, our method is more extensively evaluated on a larger dataset that consists of more than 300 CT images. It should be further pointed out that we use the same dataset as Feng *et al.*,<sup>8</sup> Liao and Shen,<sup>12</sup> and Li *et al.*<sup>13</sup> The datasets of Liao's method<sup>12</sup> and Li's method<sup>13</sup> are subsets of our dataset. Therefore, the comparison with their methods clearly reflects the

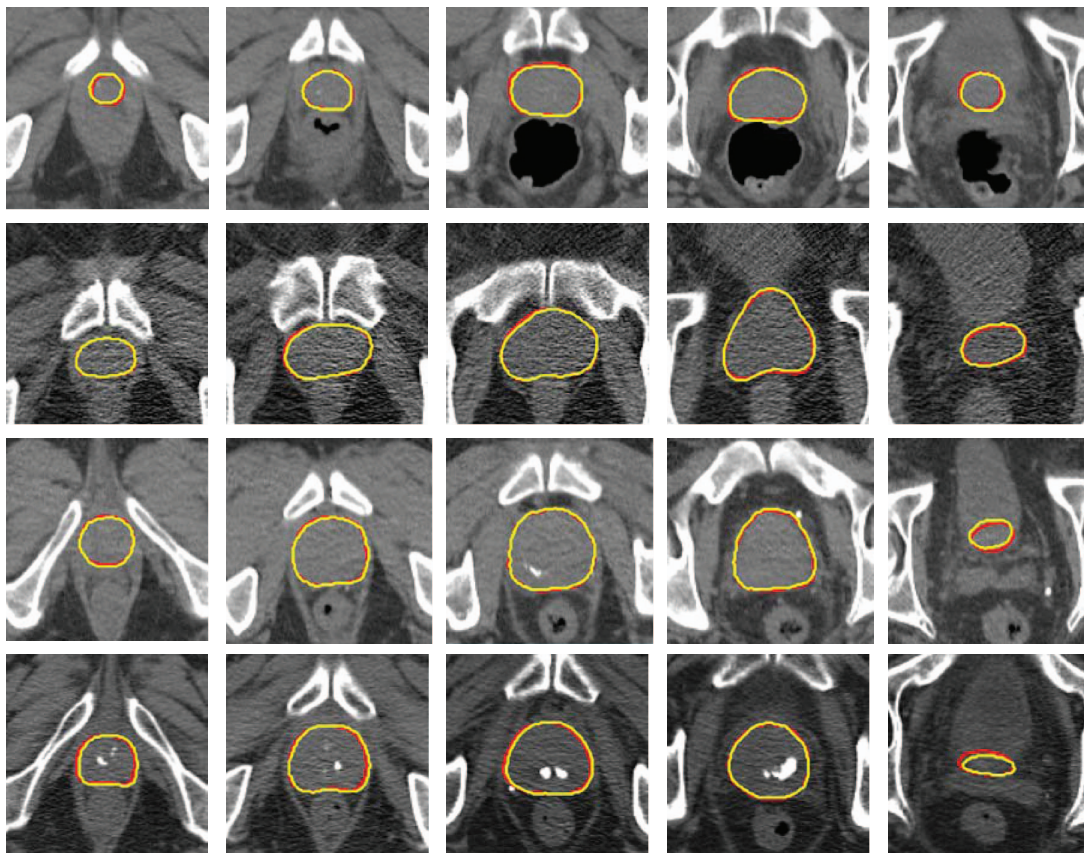


FIG. 18. Several typical segmentation results by our method. Each row corresponds to different slices of one CT image. The dark lines indicate the prostate contours manually delineated by expert, and the light lines indicate our segmentation results.

effectiveness of our methods in prostate segmentation. Figure 18 gives several visual segmentation results achieved by our method. It can be seen that even in the existence of bowel gas or heterogeneous prostate region, our method can still achieve accurate segmentation results.

## VI. CONCLUSION AND DISCUSSION

In this paper, we propose to first use sparse representation based classification to enhance the prostate by pixel-wise classification in order to address the poor contrast of CT prostate images. Then, based on the classification results, previous segmented prostates of the same patient can be used as patient-specific atlases to align onto the current treatment image space and finally the majority voting strategy can be adopted to segment the prostate. In order to overcome the limitation of the traditional SRC in pixel-wise classification, especially for the purpose of image segmentation, we further propose to extend SRC from four aspects. Five state-of-the-art prostate segmentation methods have also been compared with our method using various measurements. The experimental results show that our method can achieve better segmentation accuracy than others under comparison.

The main limitation of our method is the use of semiautomatic procedure. The future direction includes automatic estimation of middle slices and automatic segmentation of the first two treatment images. One possible direction is to learn the appearance information of distinctive prostate landmarks and train several landmark detectors, which can be used to detect the landmarks in the new treatment image. These detected landmarks can be used to robustly estimate the middle slice positions. To automatically segment the first two treatment images, we will consider using the transfer-learning methods to borrow the appearance information from the population data for guiding the segmentation. As more patient-specific treatment data are collected, the influence of population data will be gradually reduced and replaced by the current patient's data. Thus, the classification performance of learned classifiers can be gradually improved.

## ACKNOWLEDGMENTS

This work was supported in part by NIH grant CA140413, by National Science Foundation of China under Grant No. 61075010, and also by The National Basic Research Program of China (973 Program) Grant No. 2010CB732505.

<sup>a)</sup>Electronic mail: yzgao@cs.unc.edu

<sup>b)</sup>Electronic mail: liaoshu.cse@gmail.com

<sup>c)</sup>Author to whom correspondence should be addressed. Electronic mail: dgshen@med.unc.edu; Telephone: 919-966-3535; Fax: 919-843-2641.

<sup>1</sup>*Cancer Facts & Figures* (American Cancer Society, Atlanta, 2012).

<sup>2</sup>Y. Zhan, D. Shen, J. Zeng, L. Sun, G. Fichtinger, J. Moul, and C. Davatzikos, "Targeted prostate biopsy using statistical image analysis," *Medical Imaging, IEEE Transactions on* **26**(6), 779–788 (2007).

<sup>3</sup>Y. Zhan, Y. Ou, M. Feldman, J. Tomaszewski, C. Davatzikos, and D. Shen, "Registering histologic and MR images of prostate for image-based cancer detection," *Academic radiology* **14**(11), 1367–1381 (2007).

<sup>4</sup>D. Shen, Z. Lao, J. Zeng, W. Zhang, I. A. Sesterhenn, L. Sun, J. W. Moul, E. H. Herskovits, G. Fichtinger, and C. Davatzikos, "Optimized prostate biopsy via a statistical atlas of cancer spatial distribution," *Medical Image Analysis* **8**(2), 139–150 (2004).

<sup>5</sup>S. M. Pizer *et al.*, "A method and software for segmentation of anatomic object ensembles by deformable m-reps," *Med. Phys.* **32**(5), 1335–1345 (2005).

<sup>6</sup>D. Freedman *et al.*, "Model-based segmentation of medical imagery by matching distributions," *IEEE Trans. Med. Imaging* **24**(3), 281–292 (2005).

<sup>7</sup>M. J. Costa *et al.*, "Automatic segmentation of bladder and prostate using coupled 3D deformable models," *Med. Image Comput. Comput. Assist. Interv.* **10**(Pt 1), 252–260 (2007).

<sup>8</sup>Q. Feng *et al.*, "Segmenting CT prostate images using population and patient-specific statistics for radiotherapy," in *Proceedings of the Sixth IEEE International Conference on Symposium on Biomedical Imaging: From Nano to Macro* (IEEE, Boston, MA, 2009), pp. 282–285.

<sup>9</sup>S. Chen, D. M. Lovelock, and R. J. Radke, "Segmenting the prostate and rectum in CT imagery using anatomical constraints," *Med. Image Anal.* **15**(1), 1–11 (2011).

<sup>10</sup>B. C. Davis *et al.*, "Automatic segmentation of intra-treatment CT images for adaptive radiation therapy of the prostate," *Med. Image Comput. Comput. Assist. Interv.* **8**(Pt 1), 442–450 (2005).

<sup>11</sup>U. Malsch *et al.*, "An enhanced block matching algorithm for fast elastic registration in adaptive radiotherapy," *Phys. Med. Biol.* **51**(19), 4789–4806 (2006).

<sup>12</sup>S. Liao and D. Shen, "A learning based hierarchical framework for automatic prostate localization in CT images," in *Proceedings of Prostate Cancer Imaging: Image Analysis and Image-Guided Interventions*, edited by A. Madabhushi *et al.* (Springer, Berlin, 2011), pp. 1–9.

<sup>13</sup>W. Li *et al.*, "Learning image context for segmentation of prostate in CT-guided radiotherapy," in *Proceedings of the 14th International Conference on Medical Image Computing and Computer-Assisted Intervention—Volume Part III* (Springer-Verlag, Toronto, Canada, 2011), pp. 570–578.

<sup>14</sup>B. Haas *et al.*, "Automatic segmentation of thoracic and pelvic CT images for radiotherapy planning using implicit anatomic knowledge and organ-specific segmentation strategies," *Phys. Med. Biol.* **53**(6), 1751 (2008).

<sup>15</sup>P. Ghosh and M. Mitchell, "Segmentation of medical images using a genetic algorithm," in *Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation* (ACM, Seattle, WA, 2006), pp. 1171–1178.

<sup>16</sup>P. Garrigues and B. Olshausen, "Group sparse coding with a Laplacian scale mixture prior," *Adv. Neural Inf. Process. Syst.* **23**, 1–9 (2010).

<sup>17</sup>A. Krause and V. Cevher, "Submodular dictionary selection for sparse representation," in *ICML 2010: Proceedings of the 27th International Conference on Machine Learning*, 2010.

<sup>18</sup>M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.* **54**(11), 4311–4322 (2006).

<sup>19</sup>H. Jia-Bin and Y. Ming-Hsuan, "Fast sparse representation with prototypes," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010.

<sup>20</sup>J. Zhuolin, L. Zhe, and L. S. Davis, "Learning a discriminative dictionary for sparse coding via label consistent K-SVD," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

<sup>21</sup>R. Baraniuk *et al.*, "Applications of sparse representation and compressive sensing," *Proc. IEEE* **98**(6), 906–909 (2010).

<sup>22</sup>J. Wright *et al.*, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(2), 210–227 (2009).

<sup>23</sup>H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. R. Stat. Soc. Ser. B* **67**, 301–320 (2005).

<sup>24</sup>H. Xu, C. Caramanis, and S. Mannor, "Sparse algorithms are not stable: A no-free-lunch theorem," *IEEE Trans. Pattern Anal. Mach. Intell.* **99**, 187–193 (2011).

<sup>25</sup>T. Li, C. Zhang, and M. Oghara, "A comparative study of feature selection and multiclass classification methods for tissue classification based on gene expression," *Bioinformatics* **20**(15), 2429–2437 (2004).

<sup>26</sup>Y. Shi, F. Qi, Z. Xue, L. Chen, K. Ito, H. Matsuo, and D. Shen, "Segmenting lung fields in serial chest radiographs using both population-based and patient-specific shape statistics," *Medical Imaging, IEEE Transactions* **27**(4), 481–494 (2008).

- <sup>27</sup>I. G. A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.* **3**, 1157–1182 (2003).
- <sup>28</sup>H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient Sparse Coding Algorithms," in *Advances in Neural Information Processing Systems (NIPS) 2007*, Vol. 19, pp. 801–808.
- <sup>29</sup>K. Kreutz-Delgado *et al.*, "Dictionary learning algorithms for sparse representation," *Neural Comput.* **15**(2), 349–396 (2003).
- <sup>30</sup>Z. Tu and X. Bai, "Auto-context and its application to high-level vision tasks and 3D brain image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(10), 1744–1757 (2010).
- <sup>31</sup>N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *IEEE Computer Science Conference on Computer Vision and Pattern Recognition (CVPR) 2005*, Vol. 1, pp. 886–893.
- <sup>32</sup>B. Fischer and J. Modersitzki, *FLIRT: A Flexible Image Registration Toolbox Biomedical Image Registration*, edited by J. Gee, J. Maintz, and M. Vannier (Springer, Berlin, 2003), pp. 261–270.
- <sup>33</sup>L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology* **26**(3), 297–302 (1945).
- <sup>34</sup>M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.* **15**(12), 3736–3745 (2006).
- <sup>35</sup>J. Mairal *et al.*, "Non-local sparse models for image restoration," in *IEEE 12th International Conference on Computer Vision (ICCV) 2009*, Kyoto, Japan, pp. 2272–2279.
- <sup>36</sup>J. Mairal *et al.*, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.* **11**, 19–60 (2010).