Prostate Segmentation Based on Variant Scale Patch and Local Independent Projection

Yao Wu, Guoqing Liu, Meiyan Huang, Jiacheng Guo, Jun Jiang, Wei Yang, Wufan Chen*, *Senior Member, IEEE*, and Qianjin Feng*, *Member, IEEE*

Abstract—Accurate segmentation of the prostate in computed tomography (CT) images is important in image-guided radiotherapy; however, difficulties remain associated with this task. In this study, an automatic framework is designed for prostate segmentation in CT images. We propose a novel image feature extraction method, namely, variant scale patch, which can provide rich image information in a low dimensional feature space. We assume that the samples from different classes lie on different nonlinear submanifolds and design a new segmentation criterion called local independent projection (LIP). In our method, a dictionary containing training samples is constructed. To utilize the latest image information, we use an online updated strategy to construct this dictionary. In the proposed LIP, locality is emphasized rather than sparsity; local anchor embedding is performed to determine the dictionary coefficients. Several morphological operations are performed to improve the achieved results. The proposed method has been evaluated based on 330 3-D images of 24 patients. Results show that the proposed method is robust and effective in segmenting prostate in CT images.

Index Terms—Local anchor embedding, local independent projection (LIP), prostate segmentation, variant scale patch.

I. INTRODUCTION

IMAGE-GUIDED radiotherapy (IGRT) has been extensively used for prostate cancer treatment. IGRT is performed mainly in two stages: the planning stage and the treatment stage. The planning stage involves the acquisition of a computed tomography (CT) image prior to treatment. The prostate image obtained in the planning stage (i.e., planning image) should be accurately segmented to guide radiotherapy. This step is usually conducted by radiologists by manual
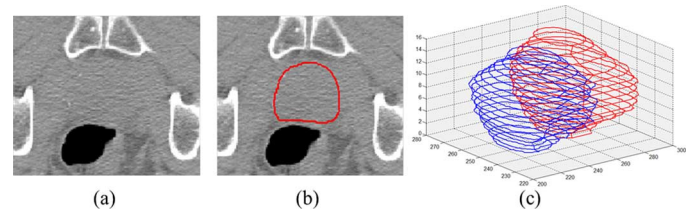
Fig. 1. (a) Image slice of a 3-D prostate CT image. (b) Manual segmentation of the prostate in this slice. (c) Manual segmentations of prostates from two treatment images of the same patient.

contouring. Furthermore, a dose plan, which assists the transmission of radiation beams during treatment, is designed based on the planning image. In the treatment stage, a CT image (i.e., treatment image) is acquired at each treatment day. The prostate in the treatment image should also be segmented. Thus, the dose plan designed in the planning image can be used for the treatment image by applying registration techniques. During treatment, the prostate in the treatment image is radiated by high-energy X-ray beams. This therapy is administered in daily fractions for several weeks [1]. The entire process requires accurate segmentation of the prostate. If the segmentation is not accurate, the dose plan which assists the transmission of radiation beams during treatment will not be aligned with the treatment image. In this case, the radiation beams will not be accurately located at the cancer tissue during treatment. Generally, if the segmentation is larger than the prostate, the radiation beams will affect some nearby normal tissues of the body. If the segmentation is smaller than the prostate, the cancer tissue will not be fully radiated and thus spread to the normal tissues. Specially, the prostate is near the rectum and the bladder, thus erroneous segmentation of the prostate may cause radiation proctitis and radiocystitis. Therefore, the precise segmentation of the prostate in CT images is important for prostate cancer radiotherapy. In traditional treatments, the prostate is manually segmented by radiologists; however, manual segmentation is time consuming and results obtained by different radiologists often differ. Thus, effective segmentation methods for prostate in CT images are valuable in clinical applications.

Distinguishing the prostate from its surrounding tissues is difficult because of the low contrast in CT images [Fig. 1(a) and (b)]. In addition, the position and appearance of the prostate in the CT image of the same patient greatly varies on different treatment days [Fig. 1(c)]. Therefore, accurate segmentation of the prostate in CT images is difficult. Since then, this topic has been extensively explored and it is still under investigation in methodological aspects.

Recently, many novel methods have been proposed on this topic. Deformable model based method is commonly used to perform prostate segmentation in CT images [1]–[6]. Freedman *et al.* [4] used the probabilistic distributions of photometric variables to facilitate prostate radiotherapy. Feng *et al.* [1] used population- and patient-specific statistics to guide the deformation of a prostate shape model; image gradient and probability distribution function have been combined to improve segmentation accuracy. Chen *et al.* [5] combined the anatomical constraints and appearance model into a Bayesian framework to achieve segmentation results. However, the performance of these methods highly depends on an accurate initialization of the deformable model, which is difficult to obtain because of variations in prostate appearance. Registration based method is another common category in this field [7]–[12]. Davis *et al.* [7] combined a nonrigid registration algorithm and a bowel gas deflation algorithm to align the planning image with the treatment image. Shi *et al.* [8] used the correspondence established between the prostate boundaries in the treatment and planning images; a fast registration approach has also been proposed to align each treatment image with the planning image. However, accurate registration results in prostate CT images are difficult to obtain because of the uncertain presence of bladder water and rectum gas [13]. More recently, classification based method has been introduced to prostate segmentation [13]–[15]. Liao *et al.* [15] used patch-based representation in a discriminative feature space and proposed a multi-atlas label-fusion method to locate the prostate in CT images. Li *et al.* [13] proposed an online-learning and patient-specific classification method based on location-adaptive image context. These studies have suggested that the image context information can improve the classification accuracy of prostate CT images. In addition, numerous prostate segmentation methods using other imaging modalities, such as magnetic resonance imaging (MRI) [16]–[20] and ultrasound (US) [21]–[25], have also been proposed.

This study aimed to propose a new classification-based method to segment the prostate in 3-D CT images automatically. The performance of this method mainly depends on the discriminative power of the image feature and efficiency of the classifier. The main contributions in our study are described in the following.

First, a novel image feature extraction algorithm, variant scale patch (VSP), is proposed in this study. For segmentation or classification, the discriminative features that can easily identify objects from their backgrounds are highly preferred. The traditional feature extraction strategy obtains the optimal features by evaluating several candidate features using automatic [26]–[28] or manual methods. These approaches depend on the performance of the feature selection method or subjectivity of users. SIFT [29] is a well-known descriptor which is based on the oriented gradient histograms. It possesses high descriptive property but is time consuming. DAISY [30] is computationally efficient, which depends on gradient histograms like SIFT and uses the Gaussian kernel by combining with a circular grid to compute the histograms. DAISY captures the image gradient information and is suitable for stereo reconstruction, object recognition and matching. BRISK [31] is another efficient descriptor, which calculates local intensity gradient and uses

a circular sampling neighborhood by intensity comparison to form the binary string descriptor. BRISK is often used in keypoint detection, description and matching. For prostate segmentation in CT images, the CT intensity information is more desired than the gradient information; thus, these methods (SIFT, DAISY, and BRISK) are not suitable in our study. The raw image patch has been shown to exhibit good performance in many cases [32]–[35]. The traditional patch-based feature of one point used in [32]–[35] is obtained by vectorizing the intensities of the image patch centered at this point. The context information is helpful for prostate segmentation [13]. Thus, if the context should be used sufficiently and the patch size is enlarged, the high dimension features will remarkably increase the computing time and consume more memory space. To overcome the limitation of the traditional patch-based feature, we apply the proposed VSP. Similar to DAISY and BRISK, VSP uses circular neighborhood to extract image information. Considering our special task (i.e., prostate segmentation in CT images), we focus on capturing rich context information in a low dimensional feature space, where we distort the traditional regular sampling patch and extract context information from variant scale circular neighbors centered at the transformed sampling points. As a result, the central image information can be emphasized and the computational burden can be reduced simultaneously. Experimental results have verified the effectiveness of VSP for prostate segmentation in CT images.

Second, a new classification approach called local independent projection (LIP) is proposed in our method. To date, sparse representation has been extensively used in many research fields, such as compressed sensing [36], [37], image denoising [38], [39], and classification [40], [41]. Similar to sparse representation-based classification (SRC) [40], the proposed LIP is based on the concept that a sample can be sparsely represented in a specific dictionary. In contrast to SRC, the proposed method uses training samples from the object and background that are presumably located on two submanifolds. These training samples are then divided into two groups: object and background dictionaries. The testing sample is independently projected on these two submanifolds by using a locally linear representation [42], in which reconstruction errors associated with both submanifolds are used as classification metrics.

Third, a novel dictionary construction approach is proposed in the current study. A series of daily scans should be obtained during radiotherapy because the appearance of the prostate in CT images greatly varies on different treatment days. However, we assume that the prostate images scanned on adjacent treatment days are highly correlated, and the information in the latest treatment images is potentially important in facilitating segmentation in the new treatment image. Considering this concept, we propose an online dictionary construction strategy. The training samples from the latest several treatment images rather than all of the treatment images are used to establish the dictionary for the new treatment image.

The rest of the current study is organized as follows. Section II introduces the methodology of the proposed algorithm.Section III presents the experimental analysis and results. Section IV provides some discussions. Section V shows the conclusion of the proposed method.
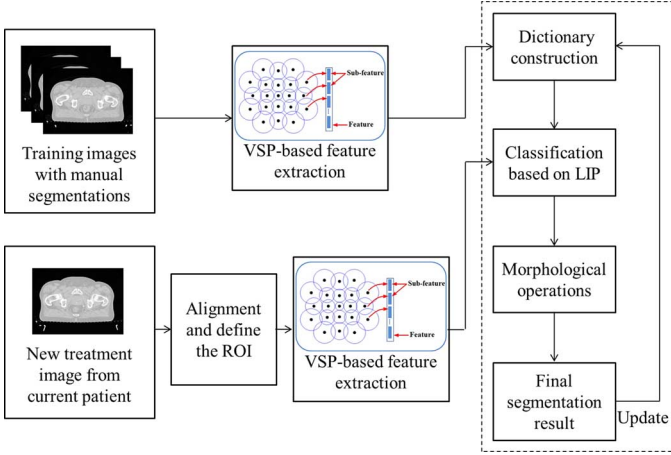
Fig. 2. Flowchart of the proposed method for prostate segmentation in 3-D CT images.

II. METHODS

The proposed method includes five steps: image preprocessing, VSP-based feature extraction, online dictionary construction, LIP-based classification, and postprocessing by morphological operations. The flowchart of the proposed method is shown in Fig. 2.

A. VSP-Based Feature Extraction

The proposed VSP is established by investigating the features of nonuniform visual acuity on human retina. Cones are a kind of photoreceptor cells in the retina. Cones are responsible for color vision, and they perceive detail and rapid changes in images. The distribution of cones in the fovea is denser than that in the periphery of the retina. Thus, the eye captures the scene in variant scales. In the central region, a high number of receptors are used to generate high-resolution images; by comparison coarse context information is sampled by fewer receptors in the peripheral region. The variant scale sampling manner of the eye contributes to the human classification decision. Using the nonuniform image captured by the eye, the human brain can reliably perform the classification task, which indicates that the nonuniform image carry enough information to satisfy the demand of the classification. This finding shows that the variant scale sampling manner of the eye is effective and can be introduced to machine leaning. The basic foundation of VSP is similar to that of the sampling manner of the eye. In contrast to uniform sampling in traditional patch-based approaches, the variant scale sampling method is introduced to obtain information from the surroundings of the candidate point to compose its feature. In this scheme, the sampling density is designed to decrease as location becomes farther from the candidate point, which ensures that the extracted feature contains detailed information of the central region and coarse context of the peripheral region. At the same time, the dimensionality of the feature space is maintained at a relatively low level. In the next paragraph, VSP is described in detail.

To compute the feature at a point c, an image patch P_c centered at c is considered. A scale parameter σ (σ is defined in the physical space in millimeters), which shows the sampling interval along the ray starting from point c, is initially introduced to compute VSP. In particular, a smaller σ corresponds to a denser sampling and vice versa. σ is variable over P_c, and σ_p (denotes the scale at point p) increases as p is located farther from c. Thus, we define σ_p as follows:

$$\sigma_p = g\left(D\left(c, p\right)\right) \tag{1}$$

where $D\left(c, p\right)$ is the Euclidean distance between points c and p. g is an increasing function. In the current study, g is defined as an exponential function

$$g\left(D\right) = \lambda e^{D/a}, \ a > 0 \tag{2}$$

where a is used to control the increase in speed of the scale along the distance D. λ determines the sampling interval in point c and is set to a minimum value of the voxel size as $\lambda = \min\{v_x, v_y, v_z\}$. In this situation, sampling is performed using the original image resolution at point c. Using this scale, we can transform each point in P_c into a new coordinate space and establish VSP. We then provide the details of coordinate transformation between VSP and original patch.

Considering point s' located at $x_{s'}$ ($x_{s'}$ is the relative coordinate to the central point c') in VSP, we can calculate the corresponding location x_s on the original image patch according to the following equation:

$$x_s^i = x_{s'}^i \left(\int_{c'}^{s'} \frac{\sigma_s ds}{D(s', c')} \right) \tag{3}$$

where x_s^i and $x_{s'}^i$ represent the ith coordinate element of x_s and $x_{s'}$, respectively. $\int_{c'}^{s'} \sigma_s ds$ denotes the line integral of scale σ from point c' to point s' in VSP. According to (1) and (2), (3) can be rewritten as follows:

$$x_s^i = x_{s'}^i \left(\int_{c'}^{s'} \frac{\lambda e^{D(s', c')/a} ds}{D(s', c')} \right)$$
$$= \frac{a\lambda x_{s'}^i (e^{D(s', c')} - 1)}{D(s', c')}. \tag{4}$$

To implement VSP conveniently, we define a sampling template $S' = \{s_i'\}_{i=1}^N$ consisting of N points on regular grids in VSP [Fig. 3(a)]. According to (4), the template S' can be transformed into an original image space and a warped template $S = \{s_i\}_{i=1}^N$ on the original image patch P_c [Fig. 3(b)]. The warped template is further truncated using a circular region around the central point c because the farthest sampling points are mostly irrelevant to the central point. The sampling points in this circular region remain and compose the final template $S_{\text{final}} = \{s_i\}_{i=1}^M$ [Fig. 3(c)]. Each point in VSP covers a region on the original patch because of the variant scale. For a sampling point s_i in S_{final}, the covered region on the original patch can be approximated by a circle neighborhood with a diameter of scale σ_{s_i} [Fig. 3(d)]. The image contents in this circular neighborhood can be extracted as the subfeature noted as f_{s_i} at point s_i, and the subfeatures of s_i in S_{final} correspond to the feature of central point c as $f = \{f_{s_1}, f_{s_2}, \ldots, f_{s_M}\}$ [Fig. 3(d)].

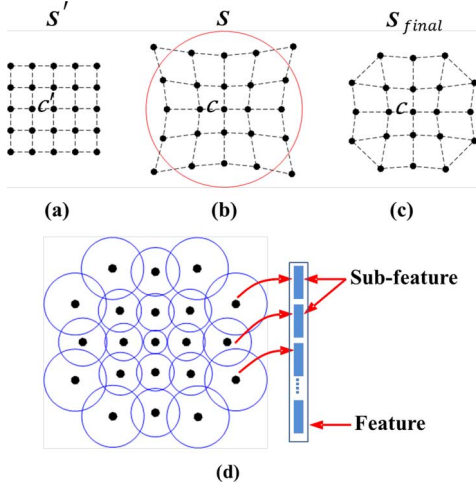Numerous approaches, such as mean and variance of image intensities and gray level histogram, can be exploited in each

Fig. 3. Schematic diagram of VSP construction.

Fig. 4. Variations in circle neighborhoods using different a values in VSP.

circle to extract the subfeature in VSP, in which users can choose an applicable method based on their special tasks. In our experiments, we considered the intensities of the pixels within each circle to calculate the subfeature in VSP. The mean of intensities was used to capture the image contents in each local region. The extracted feature can use a lower dimensional vector to cover a larger image region and thus acquire richer context information compared with the traditional patch-based feature.

Fig. 4 shows the variations in VSP at different a [(2)] in 2-D images. In Fig. 4, the size of the initial sampling template is 5×5, and 21 sampling points remain in the truncated template after the four farthest sampling points were removed. We also assumed that the pixel size is 1 mm $\times$ 1 mm and λ is set as 1. As a decreases, the covered region of VSP enlarges accordingly. At $a = 0.5$, the area covered by VSP is > 100 mm $\times$ 100 mm [Fig. 4(a)]. Thus, if we calculate and set the mean of intensities in each circle neighborhood as the subfeature, the final feature of the central point can be represented by a 21-dimensional vector. Such a low-dimensional vector not only contains detailed image information in the central region, but also reserves the coarse image context in the peripheral area. It should be noted that the VSP algorithm can be extended to the 3-D image space, and the features used in the current experiments were calculated in a 3-D environment.

B. LIP-Based Classification

1) Basic Idea of LIP: Before the proposed LIP is introduced, the following assumption is considered as the basis for LIP:

Assumption I: The samples from different classes are located on different nonlinear submanifolds, and a sample can be approximately represented as a linear combination of several nearest neighbors from its corresponding submanifold.

For N-class classification, this assumption indicates that the samples are found on a manifold $\{M^i\}_{i=1}^N$, which consists of N submanifolds, where M^i represents the submanifold associated with the ith class. Dictionary $D = \{D^i\}_{i=1}^N$ consists of N subdictionaries, where $D^i = \{d_1^i, d_2^i, \ldots, d_o^i\}$ consists of o typical samples from the ith submanifold.
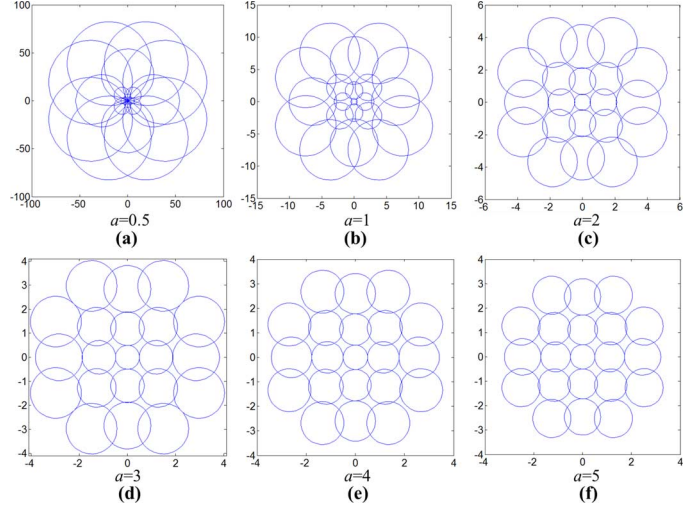
According to assumption I, a testing sample f can be projected onto each submanifold using the following linear representation:

$$f = D^i a^i + \varepsilon^i = \sum_{j=1}^o a_j^i d_j^i + \varepsilon^i \tag{5}$$

where $a^i = \{a_1^i, a_2^i, \ldots, a_o^i\}^T$ denotes the weight coefficient vector of the linear combination; ε^i is the reconstruction error for f associated with M^i. The L2-norm of ε^i noted as $\|\varepsilon^i\|_2$ can be used to measure the proximity between f and its projection on M^i, and smaller $\|\varepsilon^i\|_2$ is associated with increased likelihood that f belongs to the ith class. Thus, the classification rule can be expressed as follows:

$$l^* = \min_i \|\varepsilon^i\|_2 \tag{6}$$

where l^* is the label of the sample f. The proposed locally independent coding-based classification is different from SRC [40]. In SRC, the testing sample f is represented by the whole dictionary $D = \{D^i\}_{i=1}^N$ instead of each subdictionary in the proposed LIP [(5)]. The representation of f in SRC is defined as

$$f = Da \tag{7}$$

where $a = \{(a^1)^T, (a^2)^T, \ldots, (a^N)^T\}^T$ is the coefficient vector of $D = \{D^i\}_{i=1}^N$, which is obtained by solving a sparse coding problem with L1-norm [(9)]. In SRC, the reconstruction error of each class is obtained by the following criterion:

$$\varepsilon^i = f - D^i a^i, \quad i \in \{1, \ldots, N\}. \tag{8}$$

After obtaining the reconstruction error of each class, the class label of f is determined using the same criterion in (6).

Although the reconstruction errors are both used as the classification metric in LIP and SRC, the testing sample is projected onto each submanifold independently in the proposed method instead of the whole manifold in SRC. For each submanifold, the calculation of the projection is independent; as such, the

proposed method is more applicable than SRC if assumption I holds.

2) Implementation of LIP-Based Classification: To implement LIP-based classification, we perform three steps: dictionary construction; locally linear representation; and classification.

Step 1) In general, dictionary construction is a task-driven step. Many approaches, such as method of optimal directions (MOD) [43] and K-SVD [44], can be used to learn the dictionary for the best representation for samples. The discriminability of the dictionary is a crucial factor for the classification, which is ignored in the learning procedures of MOD and K-SVD. k-means is a clustering method that groups the training samples according to their similarities. Using k-means to cluster samples can reduce the size of the dictionary and maintain the discriminability of the training samples. k-means has been effectively applied for the dictionary learning in the classification task [45]. Therefore, we use k-means to cluster the original samples and then select the cluster centers to construct the dictionary.

Step 2) To linearly represent a sample using the dictionary, i.e., solving (5), researchers proposed several methods, such as sparse coding (SC) [40], local anchor embedding (LAE) [42], and locality-constrained linear coding (LLC) [46].

SC emphasizes the sparsity of the representation and uses several training samples to reconstruct the testing sample with minimum error. In SC, the coefficient vector of dictionary is solved using the following criterion:

$$\min_a \|f - Da\|_2^2 + \beta \|a\|_1 \tag{9}$$

where $D = \{d_1, d_2, \ldots, d_L\}$ consists of L training samples, and $a = \{a_1, a_2, \ldots, a_L\}^T$ is the coefficient vector. The second term with L1-norm is used to control the sparsity of vector a. Many efficient methods are available to solve the SC problem, such as basis pursuit (BP) [47] and orthogonal matching pursuit (OMP) [48]. In our experiments, we use OMP, which is more efficient and effective than BP [49], to solve the cost function in SC.

LLC and LAE belong to the special cases of sparse representation. These two approaches focus on locality rather than sparsity by limiting linear coding within a local neighborhood. In LLC, the cost function is as follows:

$$\min_a \|f - Da\|_2^2 + \beta \|b \odot a\|_1$$
$$s.t. 1^T a = 1 \tag{10}$$

where $\odot$ denotes the element-wise multiplication, and b is the locality adaptor which is defined as

$$b = \exp\left(\frac{dist(f, D)}{\gamma}\right) \tag{11}$$

where
$dist(f, D) = [dist(f, d_1), \ldots, dist(f, d_L)]^T$, and $dist(f, d_L)$ is the Euclidean distance between f and d_L. γ is used to adjust the weight decay speed for the locality adaptor. In LLC, the coefficient vector is obtained using a gradient descent procedure [46].

Compared with LLC, LAE imposes a nonnegative constraint to the weight coefficients. In LAE, the reconstructed sample is located in a convex region on a hyperplane spanned by its closest neighbors. Here, a special dictionary D' is defined, which consists of a set of k nearest neighbors of f in dictionary D. LAE uses the following formula to obtain the coefficient vector:

$$\min_{a'} \|f - D'a'\|_2^2$$
$$s.t. 1^T a' = 1, \ a' \geq 0 \tag{12}$$

where $a' = \{a_1, a_2, \ldots, a_k\}^T$ denotes the coefficient weights of the k nearest samples, which is obtained using the projected gradient method [42].

According to assumption I, the locality of the representation is highly preferred for LIP. Thus, LAE is applicable in solving linear representation in the current study. Furthermore, constraints $1^T a' = 1$ and $a' \geq 0$ ensure that the reconstructed sample is a convex combination of its closest neighbors. Using LAE, only one parameter needs to be set, that is, the number of k nearest neighbors. The selection of k is discussed in Section III-B.

Step 3) To classify a testing sample f, we initially project f to each submanifold according to (5) and then use LAE to obtain the coefficient vector. Calculating the reconstruction error on each submanifold, we finally determine the class of f based on (6).

3) Summary of LIP: For clarity, we summarize the proposed LIP in Algorithm 1.

Algorithm 1: LIP-based classification

Input: testing sample f and training samples from each class.

Output: label of f.

Step 1) Dictionary construction

k-means clustering method is used to learn a compact subdictionary for each class, denoted as D^i for class i ($i = 1 \ldots N$).

Step 2) Locally linear representation

D^i is used to represent f according to (5). LAE is adopted to solve the coefficient vector a^i.

Step 3) Classification

Given D^i and a^i, the reconstruction error ε^i on each subdictionary is calculated based on (5). f is labeled to the class with the minimum ε^i [(6)].

C. Online Updated Dictionary

An online updated strategy, which utilizes the latest image information of the current patient, is proposed to construct the dictionary for each new treatment image. Radiologists should manually segment the first three daily scans of the same patient. The corresponding results are then used to construct the original dictionary for the fourth daily scan. In the following days of radiotherapy, the dictionary is automatically updated using the segmentation results of the new treatment images. All training samples are collected in a region of interest (ROI), which is defined in Section II-D. In our experiments, only the latest five segmented scans are used to construct the dictionary for the new treatment image. For the fourth and fifth daily scans, only three and four segmented images are obtained, respectively. Nevertheless, the scans are sufficient to construct an over-complete dictionary. Thus, the dictionary is online updated during the whole radiotherapy procedure; this dictionary can provide the latest image information and improve the segmentation accuracy of the current treatment image.

D. Pre- and Post-Processing

In the preprocessing step, all images are rigidly aligned to a common space based on the pelvic bone structure. According to the manual segmentation results of the planning images, the centroid of the prostate is computed and a ROI (121 mm × 121 mm × 90 mm) around the prostate is then defined based on the centroid. This defined ROI is used not only in dictionary construction but also in the classification step; as such, computational burden is reduced.

Postprocessing is finally performed to improve the results. This step is implemented by several morphological operations on the achieved segmentations. The details are as follows: 1) The current segmentations (binary images) are eroded and dilated using a line structuring element with a length of six pixels at 45° (measured in a counterclockwise direction from the horizontal axis); 2) The results of Step 1 are further eroded and dilated using a line structuring element with a length of six pixels at 135°. This step is automatically performed slice-by-slice using the MATLAB tool box. These operations can eliminate the isolated points and thus generate smooth results.

III. EXPERIMENTAL RESULTS

In this section, a set of experiments is presented to evaluate the performance of the proposed method. The following criteria were considered: 1) overall performance of the proposed method; 2) effectiveness of VSP-based feature extraction; 3) effectiveness of the proposed LIP-based classification; 4) performance of the online updated dictionary; and 5) comparison to state-of-the-art methods.

To compare the proposed method with relevant prostate segmentation methods, we used five common evaluation measures: dice similarity coefficient; average surface distance (ASD); centroid distance (CD); probability of detection (PD); and probability of false alarm (FA). These measures are commonly used to evaluate the performance of the segmentation methods [11], [13], [15].

Dice similarity coefficient is used to measure the similarity between ground truth and segmentation results

TABLE I
COMPUTATIONAL COMPLEXITY ANALYSIS OF THE PROPOSED METHOD

Dictionary learning (*k*-means)	VSP feature extraction	LIP classification
$O(Q^{Cd+1} \log Q)$	$O(dne^{\frac{3P}{2a}})$	$O(Cnk)$

Q is the number of training samples to be clustered, C is the number of clustering centers, d is the feature dimension, n is the number of testing samples, P is the patch size of the VSP template, a is a parameter in VSP feature extraction [Eq. (2)], and k is the number of nearest neighbors in LAE.

$$\text{Dice} = \frac{2 * |A_A \cap A_G|}{|A_A| + |A_G|} \tag{13}$$

where A_A represents the automatic segmented prostate, and A_G denotes the ground truth. A higher Dice ratio corresponds to a better segmentation result.

CD measures the distance between the centroid of the automatic segmented prostate and that of the ground truth. The centroid of the prostate, which is defined as the mean position of all prostate voxels, is very important in radiotherapy because it is the isocenter of the treatment plan [7]. Therefore, CDs in lateral (x), anterior–posterior (y), and superior–inferior (z) directions are commonly used to evaluate the performance of the segmentation methods.

ASD is the average distance between the surfaces of the segmented prostate and the ground truth along 360 × 180 rays evenly directed to the surface from the ground truth centroid.

PD (also called sensitivity or true positive rate) is the percentage of the ground truth that overlaps the segmented prostate.

FA (also called false positive rate) is the percentage of the segmented prostate that lies outside the ground truth.

A. Data Acquisition and Computing Time

The dataset used in the current study consists of 24 patients, each with more than nine daily CT scans and with a total of 330 3-D CT images. Each image has an XY size of 512 × 512 with more than 44 slices. The voxel resolution of each image is 1 mm × 1 mm × 3 mm. All images were resampled to 1 mm × 1 mm × 1 mm. Manual segmentations of the prostates in all of the images were provided to construct the training dataset and evaluate segmentation results.

A PC with Intel P4 3.0-GHz processor and 16 GB RAM was used as the workstation in this research. In the training step, we used k-means clustering to learn a compact dictionary, which took 1.1 h on average. During radiotherapy, the training step is an offline procedure. The segmentation of a new treatment image required 1.8 min on average, including VSP feature extraction (0.5 min), LIP classification (1.3 min), and morphological operations (0.002 min). In practice, a radiologist needs 15–20 min for the segmentation of a treatment image [5]. The computational complexity of the proposed method is listed in Table I.

B. Parameter Choices

1) Parameter Setting in VSP: In Section II-A, two parameters were assigned in VSP feature extraction: a in (2) and the size of the initial sampling template. These two parameters should be correlated during feature extraction. Different a values may

TABLE II
MEAN DICE RATIOS OF 24 PATIENTS USING DIFFERENT a
VALUES AND PATCH SIZES IN VSP

a\patch size(voxels)	5×5×5	7×7×7	9×9×9	11×11×11	13×13×13
1	**91.4**	\	\	\	\
2	86.1	88.9	\	\	\
3	84.7	86.6	\	\	\
4	82.9	85.2	86.7	85.9	\
5	80.5	83.4	85.1	85.2	85.8
6	78.3	82.3	83.9	84.3	84.9

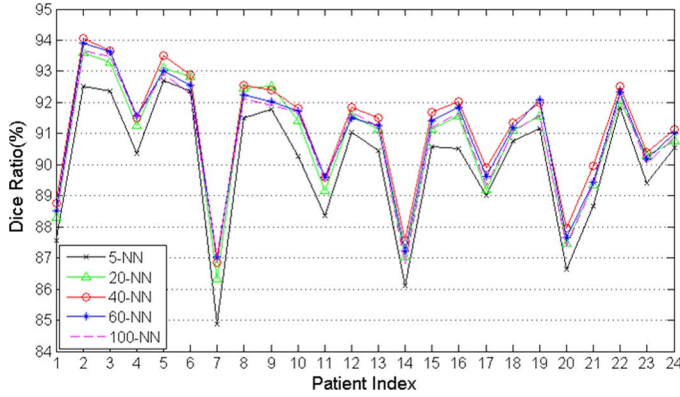Fig. 5. Mean Dice ratios of all 24 patients obtained using different numbers of the nearest neighbors in LAE.
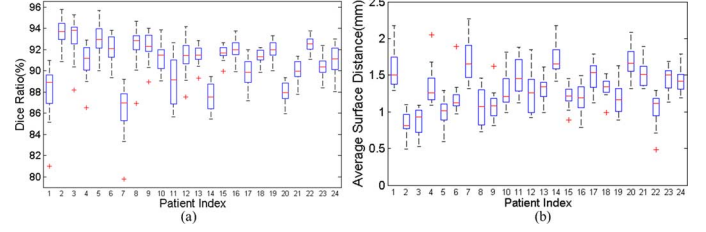

Fig. 6. Whisker plots of the (a) Dice and (b) ASD measures for 24 patients. Each box denotes the distribution of the results of the daily scans for one patient. Red line in each box corresponds to the median of this group; bottom and top lines of the box indicate the 25th and 75th percentiles, respectively. The whiskers of each box are extended to extreme data points.
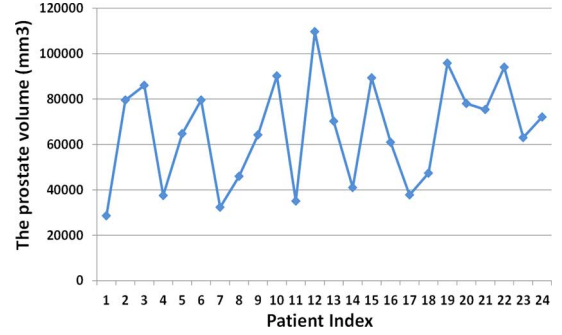

Fig. 7. Mean prostate volumes in all treatment images of 24 patients.

match different patch sizes. Therefore, we varied these values simultaneously in the current experiment to combine these two parameters accurately. We varied parameter a from 1 to 6, and varied the template size from 5 to 13 with an interval of 2. All 330 images of 24 patients were used to find the proper settings of these two parameters. Here, we used the manual segmentation results of the first three daily scans of each patient to construct an original dictionary, and sequentially tested on the following daily scans. The new segmentation results were added to update the dictionary. Segmentation results of all treatment images (except for the first three daily scans) of the 24 patients under different values of parameter a and patch sizes are shown in Table II.

In the current experiment, not all the combinations in Table II were tested because if we set parameter a as a small value, a large image area can be covered even with a small patch size (Fig. 4). Thus, as a decreases, patch size simultaneously decreases. Table II demonstrates that $a = 1$ and patch size = $5 \times 5 \times 5$ are applicable in our dataset. It should be noted that different image data may need different a values and patch sizes, which should be tested during the course of the experiments.

2) Choice of k in LAE: Parameter k is the number of nearest training samples in LAE algorithm. To set k properly, we varied it from 1 to 100. In this experiment, all images in our dataset were used. We also used the manual segmentation results of the first three daily scans of each patient to construct an original dictionary and added the new segmentation results to update the dictionary. Fig. 5 shows the segmentation results using 5, 20, 40, 60, and 100 nearest neighbors. Here, we used the mean Dice ratio of all treatment images (except for the first three daily scans) of each patient to show the results obtained using various values of k. As shown in Fig. 5, the LAE algorithm generated

the best performance among most of the patients when $k = 40$. Performance began to decline when k deviated from $k = 40$. Therefore, we set k to 40 in the LAE coding step.

3) Size of the Dictionary: In LIP-based classification, we used the k-means cluster centers to construct the dictionary. All 330 images of 24 patients were used in this experiment. To set the size of the dictionary properly, we varied it from 1000 to 10 000 with an interval of 1000. During the experiments, segmentation accuracy improved as the size of the dictionary increased. Compared with using 1000 cluster centers, the mean Dice ratio of 24 patients using 10 000 cluster centers improved from $89.3\% \pm 2.4\%$ to $91.6\% \pm 2.0\%$, with an increase of 2.3% (paired t-test; $p < 0.001$). However, it took more computing time as the number of cluster centers increased. The increase rate of Dice ratio decelerated when the number of cluster centers was larger than 5000. To balance both efficiency and effectiveness, we finally set the size of the dictionary to 5000.

C. Overall Performance of the Proposed Segmentation Method

Experiments were performed to evaluate the performance of the proposed method. Fig. 6 shows the whisker plots of the Dice and ASD measures of the proposed method for 24 patients. It can be observed that the proposed method can generate high Dice ratios, and the mean Dice for all 24 patients using the proposed method is 91.4%. Since the Dice ratio cannot provide insight on the spatial distribution of the segmentation errors, the distance between the manual and automatic segmentation boundaries is usually computed to judge the segmentation error [50]. A segmentation error of a few millimeters at the boundaries is clinically acceptable. Generally, the ASD between the manual and automatic segmentation boundaries below 3–4 mm
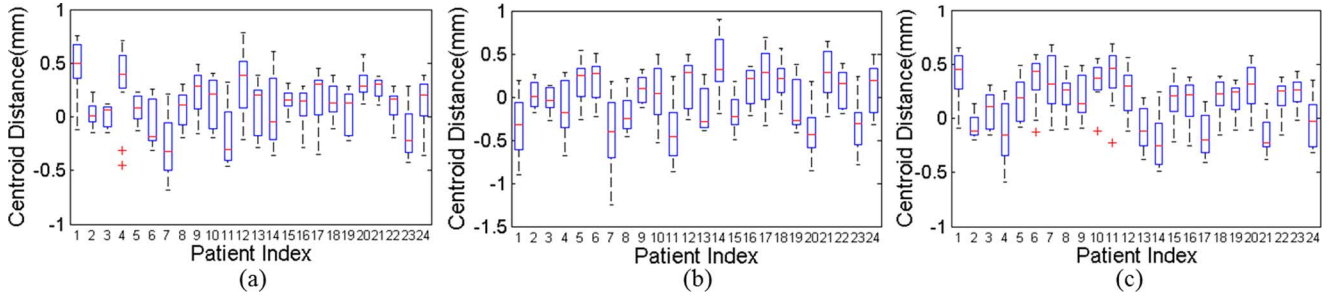
Fig. 8. CD measures for 24 patients in lateral (a), anterior–posterior (b), and superior–inferior (c) directions.

is clinically acceptable. As shown in Fig. 6(b), the ASD values of all patients obtained using our method are below 2.5 mm and within tolerance level for clinical practice. Considering that the Dice ratio is sensitive to the size of the object measured [51], we also show the prostate volumes of all patients in Fig. 7. The sizes of the prostates illustrated in Fig. 7 considerably vary across different patients. Generally, a higher Dice ratio is easier to be obtained when the prostate is large.

CD measures in lateral (x), anterior–posterior (y), and superior–inferior (z) directions are shown in Fig. 8. Many CDs along these three directions are < 1 mm. A lower CD value indicates that the position of the segmented prostate is closer to the ground truth. All of these results indicate that the proposed method is robust in segmenting prostate in the CT images.

D. Effectiveness of VSP-Based Feature Extraction

In this section, we further evaluated the effectiveness of the proposed VSP. We compared VSP-based feature with the traditional patch-based feature used in [32]–[35]. The traditional patch-based feature of one point was obtained by vectorizing the intensities of an image patch centered at this point. In our experiment, an initial $5 \times 5 \times 5$ patch was chosen in VSP, parameter a was set to 1, and the intensity mean value of each circular neighborhood was used to denote the subfeature. The image region covered by a $5 \times 5 \times 5$ VSP template with $a = 1$ is approximately 25 mm $\times$ 25 mm $\times$ 25 mm [Fig. 4(b)]. Given that the size of VSP template was $5 \times 5 \times 5$, we compared it with the traditional patch-based feature with a $5 \times 5 \times 5$ patch size. Additionally, we compared the approximately 25 mm $\times$ 25 mm $\times$ 25 mm image area covered by VSP with the traditional patch-based feature with a $25 \times 25 \times 25$ patch size. All these features were compared using the proposed classification framework with different feature extraction procedures. The results of all 24 patients obtained using different features are shown in Fig. 9.

In this experiment, the dimension of VSP-based feature was 81 with a $5 \times 5 \times 5$ initial template after some peripheral circles were removed [Fig. 3(c)]. The dimensions of the traditional patch-based feature with $5 \times 5 \times 5$ and $25 \times 25 \times 25$ patch sizes were 125 and 15 625, respectively. Compared with the traditional patch-based feature with a $5 \times 5 \times 5$ patch size, the Dice ratio of VSP-based feature was enhanced from $77.5\% \pm 3.9\%$ to $91.4\% \pm 1.9\%$, with an increase of 13.9% (paired *t*-test; $p < 0.001$). Using a larger patch size to extract the traditional patch-based feature significantly improved the segmentation results (Fig. 9). The mean Dice ratio of all patients using the
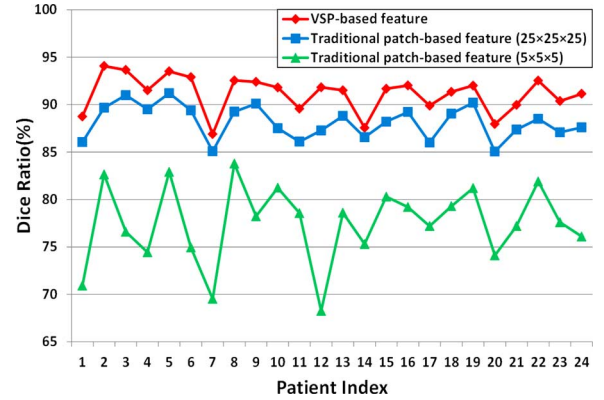

Fig. 9. Mean Dice ratio of each patient obtained using 1) VSP-based image feature, 2) traditional patch-based feature with a $25 \times 25 \times 25$ patch size, and 3) traditional patch-based feature with a $5 \times 5 \times 5$ patch size.

traditional patch-based feature with a $25 \times 25 \times 25$ patch size was 88.2%, with 10.7% improvement compared with using a $5 \times 5 \times 5$ patch size. However, the computing time also considerably increased using the traditional patch-based feature with a $25 \times 25 \times 25$ patch size, which took approximately 163 min on average for the segmentation of one treatment image. The traditional patch-based feature with a $5 \times 5 \times 5$ patch size only took approximately 1.5 min on average. VSP uses a lower dimensional feature vector to represent more image information compared with the traditional patch-based feature extraction. Moreover, the variant scales of the circles in VSP guarantee that the extracted feature of the central point contains detailed information of the central region and coarse context of the peripheral region. Thus, VSP-based feature reduces the computing time and simultaneously improves the segmentation results compared with the traditional patch-based feature.

Fig. 10 shows some segmentation results using VSP- and traditional patch-based features. It can be observed that the proposed VSP algorithm can produce more accurate segmentation results.

E. Effectiveness of the Proposed LIP

1) Performance of the k-Means Clustering Method: In LIP, we first used k-means to learn a compact dictionary. To evaluate the performance of k-means, we compared it with MOD and K-SVD, which are extensively used dictionary learning methods [43], [44]. In this experiment, all 330 images of the 24 patients were used. The three methods were compared using the same segmentation framework, except the dictionary learning
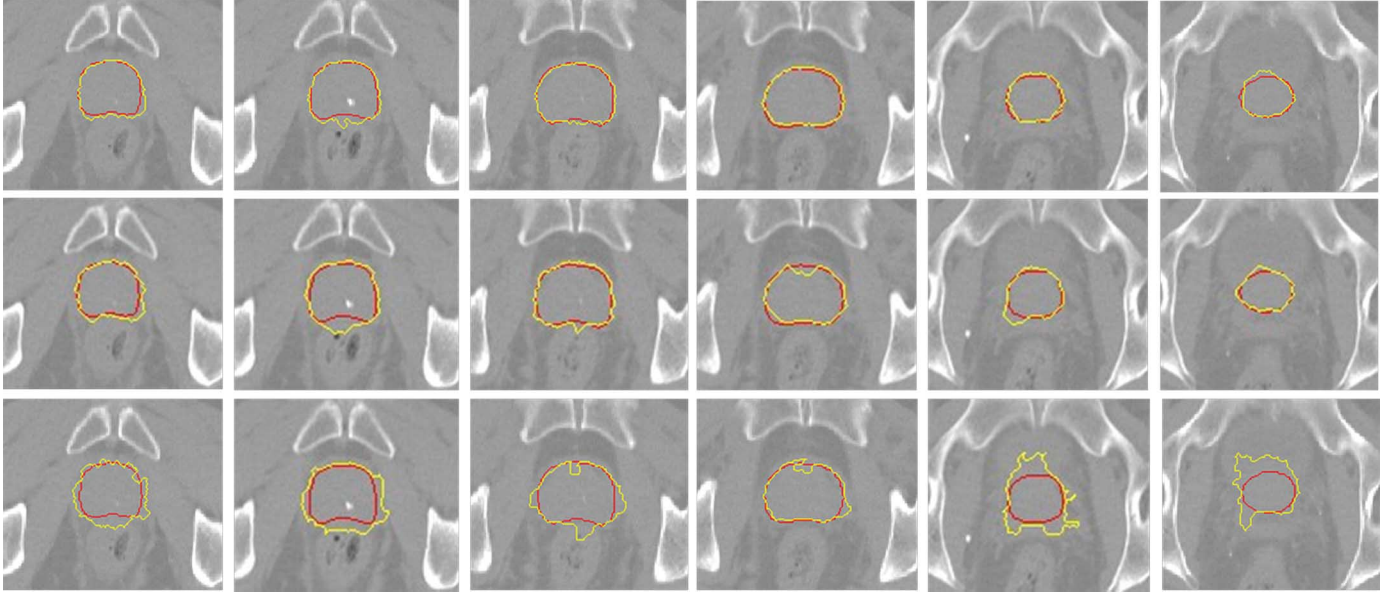
Fig. 10. Slices of the seventh treatment image of the ninth patient using VSP-based feature (first row), traditional patch-based feature with a $25 \times 25 \times 25$ patch size (second row), and traditional patch-based feature with a $5 \times 5 \times 5$ patch size (third row). Red curves show the ground truth and yellow curves show the automatic segmentation results obtained using different features. The corresponding Dice ratios are 91.57%, 88.45%, and 79.23%, respectively.
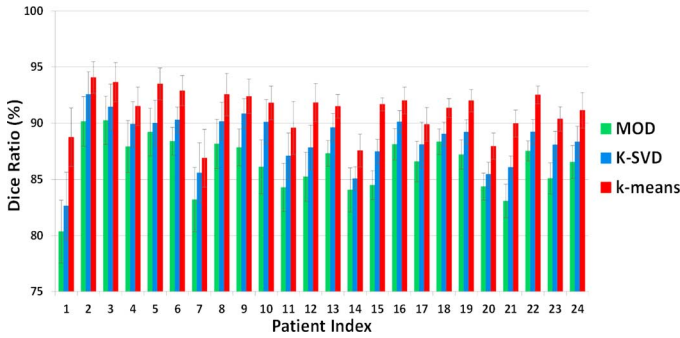

Fig. 11. Mean and standard deviation of Dice ratios obtained using MOD, K-SVD, and k-means to perform the dictionary learning step.
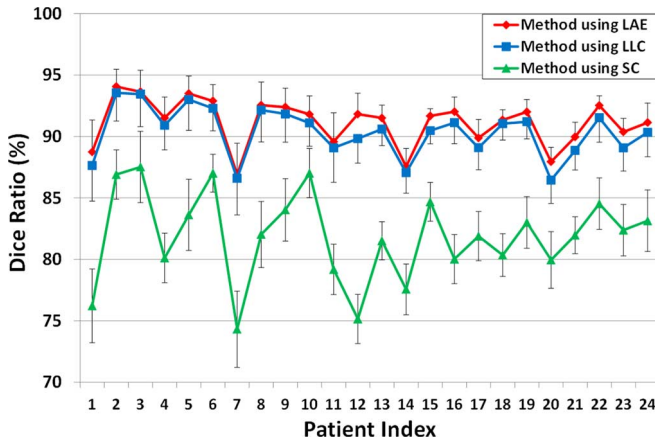

Fig. 12. Mean and standard deviation of Dice ratios of all 24 patients obtained using SC, LLC, and LAE to perform the coding step. For clarity, only one side of the standard deviations are shown on the upper two lines (red and blue).
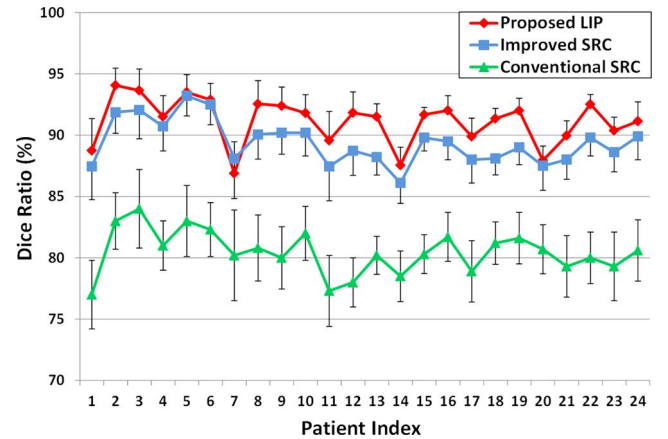

Fig. 13. Mean and standard deviation Dice ratios of all 24 patients obtained using the 1) conventional SRC [40] with SC in the coding step, 2) improved SRC with LAE in the coding step, and 3) proposed LIP. For clarity, only one side of the standard deviations are shown on the upper two lines (red and blue).

[52] and is a conventional initialization strategy in k-means clustering, was used to initialize the cluster centers. The Dice ratios of all 24 patients obtained using MOD, K-SVD, and k-means are shown in Fig. 11. k-means shows the best performance among the methods because MOD and K-SVD learn the dictionary based on representation and cannot maintain the discriminability of the training samples. When k-means was used, the mean Dice ratio of all 24 patients significantly improved by 4.7% against MOD and 2.6% against K-SVD (paired t-test; $p < 0.001$).

2) Comparison Between LAE and SC, LLC: In this section, we compared three coding methods (i.e., LAE, SC, and LLC) to choose a proper one for LIP. Fig. 12 shows the mean and standard deviation of Dice ratios obtained using different coding methods. Considering that the standard deviations on the upper

step. In these three methods, the number of dictionary elements to train was set to 5000, and the number of iterations was empirically set to 300. Random selection, which can avoid bias
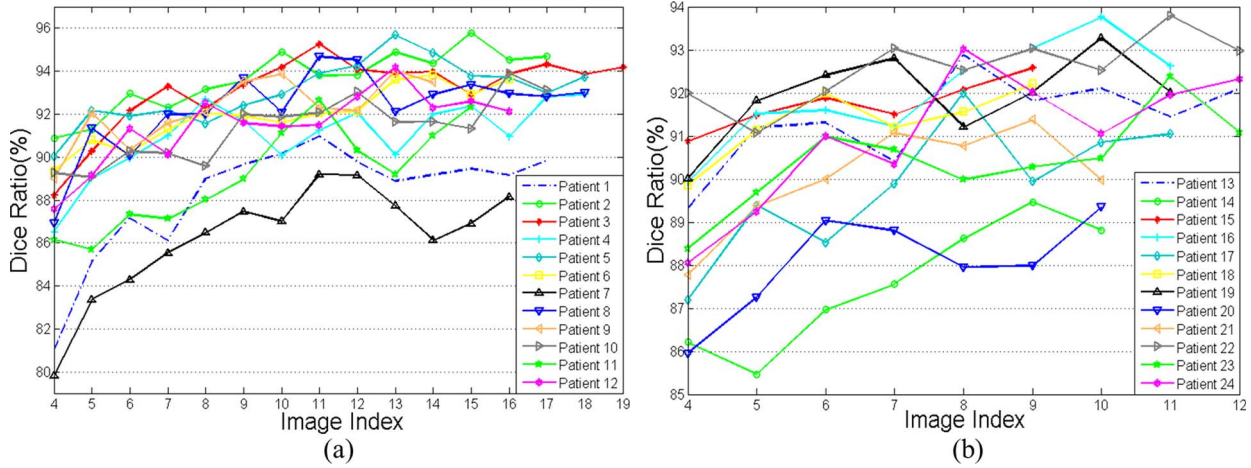
Fig. 14. Dice ratios of different treatment images of 24 patients obtained using the proposed method with the online updated dictionary. The number of treatment images varies in different patients.

two lines (red and blue) will have some overlaps, we only show one side of the standard deviations on these two lines for clarity.

Fig. 12 shows that LAE exhibits better robustness compared with SC and LLC. Compared with SC, the Dice ratio using LAE was improved from $81.8\% \pm 3.6\%$ to $91.4\% \pm 1.9\%$ with 9.6% increase (paired t-test; $p < 0.001$). LLC exhibits locality constraint and can produce satisfactory results, indicating that locality is more important than sparsity. As shown in Fig. 12, addition of a nonnegative property in LAE is beneficial in segmentation. The mean Dice ratio of LAE was 0.8% higher than that of LLC (paired t-test; $p < 0.001$). These results show that LAE is more applicable in addressing difficulties in prostate segmentation compared with SC and LLC.

3) Comparison Between LIP and SRC: To evaluate the effectiveness of LIP, we compared it with the conventional SRC [40] and the improved SRC. The conventional SRC focuses on the sparsity property of the representation and uses SC to perform the coding step. In LIP, the locality property is emphasized, and LAE is used in the coding step. To further measure the effectiveness of the classification criterion in LIP, we also compared it with the improved SRC, which uses LAE in the coding step but with different classification criteria. Fig. 13 shows the results of all 24 patients obtained using different classification methods. Compared with the conventional SRC, the Dice ratio obtained by LIP improved from $80.2\% \pm 1.8\%$ to $91.4\% \pm 1.9\%$, with 11.2% increase. The mean Dice ratio of LIP was also 1.8% higher than that of the improved SRC. Both improvements are statistically significant, with the p value less than 0.001, which indicates that the assumption and classification criterion proposed in LIP are useful and can be effectively used in the segmentation task.

F. Performance of the Online Updated Dictionary

In this section, we evaluated the performance of the online updated dictionary. Fig. 14(a) and (b) show the segmentation results using the online updated dictionary for 24 patients. In our dataset, the patients have different numbers of treatment
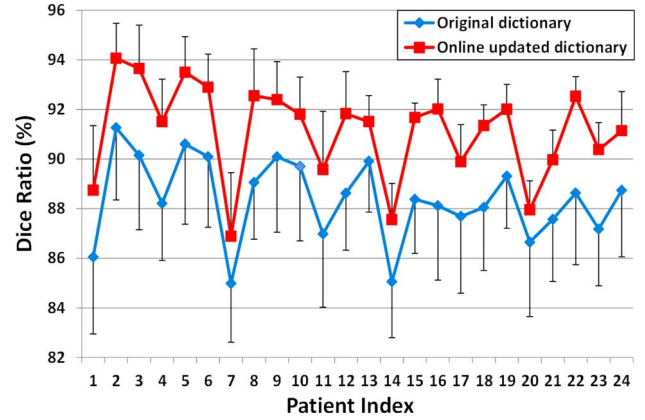

Fig. 15. Mean and standard deviation of Dice ratios obtained using the online updated dictionary and the original dictionary (the first three daily scans of each subject without update). For clarity, only one side of the standard deviations are shown on the two lines (red and blue).

images. The segmentation results of most treatment images improved when new segmentation results were added to the original dictionary [Fig. 14], which demonstrates that the online updated strategy is helpful for the segmentation of the new treatment images. To further evaluate the performance of the online updated dictionary, we also compared it with the method that only utilizes the original dictionary without online update for the segmentation of the new treatment images. Fig. 15 shows the mean and standard deviation of Dice ratios for 24 patients with and without the online updated dictionary. The mean Dice ratio for all patients using the online updated dictionary was 2.8% higher than that using the original dictionary (paired t-test; $p < 0.001$). Hence, the online updated dictionary can provide the latest image information and improve the segmentation accuracy of the new treatment images.

G. Comparison With Relevant Methods

To evaluate the performance of the proposed method, we compared it with the performance of six relevant methods for prostate segmentation in CT images. Feng *et al.* [1] and Chen

TABLE III
COMPARISONS WITH THE RELEVANT METHODS USING FIVE EVALUATION MEASURES. N/A MEANS THIS MEASUREMENT
WAS NOT REPORTED IN THE CORRESPONDING PAPER

Methods	Mean Dice (%)	Mean ASD (mm)	Mean CD (x\y\z) (mm)	Median PD (%)	Median FA (%)
Davis *et al.* [7]	82.0±6.0	N/A	−0.26±0.6\0.35±1.4\0.22±2.4	N/A	N/A
Feng *et al.* [1]	89.3±5.0	2.08±0.79	N/A	N/A	N/A
Chen *et al.* [5]	N/A	1.10±N/A	N/A	84	13
Liao *et al.* [11]	89.9±N/A	1.08±N/A	0.21±N/A\−0.12±N/A\0.29±N/A	89	11
Li *et al.* [13]	90.8±N/A	1.40±N/A	0.18±N/A**−0.02**±N/A\0.57±N/A	90	10
Liao *et al.* [15]	90.9±N/A	**0.97**±N/A	0.17±N/A\0.09±N/A\0.18±N/A	90	**8**
Our method	**91.4**±1.9	1.30±0.49	**0.11**±0.32\-0.03±0.47**0.14**±0.39	**93**	10

et al. [5] are based on deformable models. Methods proposed by Davis *et al.* [7] and Liao *et al.* [11] are based on the registration technique. The other two methods (i.e., Li *et al.* [13] and Liao *et al.* [15]) belong to the classification category. Note that, most of the six compared methods adopted the same dataset. Feng *et al.* [1], Liao *et al.* [11], and Liao *et al.* [15] used the same dataset consisting of the same 24 patients with a total of 330 prostate CT images. The methods proposed by Davis *et al.* [7] and Li *et al.* [13] were evaluated on a subset of the 24 patients used in [1], [11], [15]. Davis *et al.* [7] used three patients with 40 images, and Li *et al.* [13] used 11 patients with 161 images to evaluate their methods. The method in Chen *et al.* [5] was evaluated on 13 patients with 185 images. Our dataset is the same as the dataset in [1], [11], [15]. Therefore, the effectiveness of our method can be evaluated upon comparison with these methods. To fairly compare with these relevant methods, we used the same metrics (mean Dice, ASD, CD and median PD, FA) as reported in most of these papers. Table III shows the results of these methods under five evaluation measures. The best results are shown in bold in Table III.

It can be observed that the proposed method yields the highest mean Dice and median PD among the methods described in this study. Although the mean CD along y direction of the proposed method is larger than that in Li *et al.* [13], the mean CDs along the x and z directions are better than those of other methods. The mean ASD and the median FA of the proposed method are also acceptable. Five of the six compared methods reported their computing time in the corresponding paper. In Feng *et al.* [1], the computing time was 95.8 s. In Chen *et al.* [5], the average time was less than 1 min. In Liao *et al.* [11], the training step took approximately 6.7 h on average, and the segmentation of a new image took 3.8 min on average. In Li *et al.* [13], the training step took approximately 2 h, and the average time for segmentation of a new image was approximately 3 min. In Liao *et al.* [15], approximately 28.6 min on average was necessary in the training step, and 2.6 min on average for the segmentation of a new treatment image. In the current study, the training step took approximately 1.1 h on average. The segmentation of a new treatment image required approximately 1.8 min on average. Although the computing time of our method for the segmentation of a new image is slightly longer than those of Feng *et al.* [1] and Chen *et al.* [5], the results obtained are better than these two methods. For example, the mean Dice ratio of the proposed method is 2.1% higher than that of Feng *et al.* [1], and the

median PD is 9.0% higher than that of Chen *et al.* [5]. The computing time of our method is less than the other three methods in [11], [13], [15] as well as that of the manual segmentation by a radiologist (15–20 min). Considering both efficiency and effectiveness, the proposed method is suitable for the segmentation of prostate in CT images.

IV. DISCUSSIONS

The main contributions in the current paper are 1) VSP-based feature extraction, 2) LIP-based classification, and 3) the online dictionary updated strategy.

In VSP-based feature extraction, we choose a $5 \times 5 \times 5$ template size and set $a = 1$ after trying different parameter settings [Table II]. However, the overlaps of the circles lead to redundancy along with the decrease in parameter a [Fig. 4]. Information from the overlap is shared and contributes to more than one subfeature, which is similar to the soft assignment approaches widely used in bag-of-words (BoW) [53]. This strategy makes the feature more robust against the location variation of the distinguishable structure that appears in the peripheral region. The optimal selection of parameter a considers the trade-off between the covered region of VSP and the redundancy of the feature. In our experiments, $a = 1$ is an optimal setting upon the use of a $5 \times 5 \times 5$ VSP template.

The traditional patch-based feature with a larger patch size ($25 \times 25 \times 25$) obtains better results compared with using a $5 \times 5 \times 5$ patch size [Fig. 9], which indicates that a larger image region with more context information can improve the segmentation results. However, high-dimensional features also markedly increase the computing time. The proposed VSP can use a low dimensional feature vector to represent much image information, thus reducing the computational burden. Moreover, the central image information is emphasized using the proposed VSP, which can be explained as follows: 1) The circles, which are used to extract the subfeature in VSP, have variant scales. The variant circles in VSP guarantee that the extracted feature of the central point contains detailed information of the central region and coarse context of the peripheral region; 2) The location of each sampling point [black dot in Fig. 3(d)] is defined using (4). Fig. 3(d) illustrates that the sampling points in the central region are denser than those in the peripheral region, which is consistent with the nonuniform sampling manner of the eye. The proposed VSP gives detailed mathematical steps on how each sampling point contributes to the central point. Furthermore, Buades *et al.* [54] introduce

a nonlocal patch-based method, which has been widely used in image denoising [55] and segmentation [56]. Buades *et al.* [54] use a Gaussian kernel to provide higher weights to the pixels at the center of the patch, which indicates that the image information of the central region is more important than that of the peripheral region. This property is well maintained in the proposed VSP-based feature. Essentially, the traditional patch-based feature used in [32]–[35] can be viewed as a special case of VSP-based feature. If we use a sufficiently large parameter a, each circle in VSP only contains one pixel. In this case, VSP-based feature is the same as the traditional patch-based feature.

In the proposed LIP, based on the updated dictionary, we first use k-means method to cluster the training samples of each class. All the clusters of one class are assumed to be located on one submanifold. Based on assumption I, we use the cluster centers on each submanifold to construct a subdictionary for each class and use these two subdictionaries (prostate or non-prostate) to represent the testing sample, respectively. Then, we use LAE to solve the dictionary coefficients for each class. Given that both classes have reconstruction errors, we finally label the testing sample to the class with a smaller reconstruction error.

In our experiments, LIP is compared with SRC [40], a popular classification method and has been widely used in many fields [57], [58]. LIP and SRC differ as follows. 1) Their classification criteria are different. In LIP, we assume that the samples from different classes are located on different submanifolds. For the segmentation of a testing sample, we use the subdictionary, which contains training samples from each submanifold, to represent the testing sample and calculate the representation error associated with each subdictionary to classify the testing sample. The reconstruction error in LIP can be interpreted as the distance between the testing sample and each submanifold. We choose a minimum distance to determine the label of the testing sample. However, SRC uses a whole dictionary containing training samples from all classes to represent the testing sample and calculates the representation error associated with each class in the dictionary to determine the label of the testing sample. This representation error in SRC cannot be well explained. 2) Their coding steps are different. The proposed LIP uses LAE, which emphasizes the locality of the representation, to solve the dictionary coefficients. Different from LIP, SRC emphasizes the sparsity property and uses SC with L1-norm in the coding step. Actually, because locality must lead to sparsity, but not necessarily vice versa [59], LIP still maintains the sparsity property of the representation. Compared with the conventional SRC, the results can be markedly improved using the proposed LIP. To further evaluate the effectiveness of the classification criterion in LIP, we also compare it with the improved SRC employing the same coding method (LAE). The proposed LIP still obtains a significant improvement.

The online updated dictionary contains the latest image information of the daily scans. Compared with using the original dictionary without update, the online updated dictionary can improve the segmentation results, which indicates that the appearance of the prostate changes during the radiotherapy. In addition, the image information of the latest daily scans aids the segmentation of the new treatment images.

V. CONCLUSION AND FUTURE WORK

In the current study, a novel automatic algorithm is proposed for prostate segmentation in CT images. A new VSP-based feature extraction method is designed to address the disadvantages of the traditional patch-based feature extraction. The proposed VSP can use a lower dimensional feature space to represent more image information, and the central image information can be emphasized by using the variant scale circles, thereby reducing computational burden and achieving more accurate results compared with the traditional patch-based feature extraction. Considering the special research tasks, we can extract various features using the VSP framework. We also propose a new classification method, namely, LIP, which has been proven effective in handling the segmentation task. In contrast to sparse representation-based methods, LAE algorithm emphasizes local and nonnegative properties; this algorithm is used to perform the coding step in LIP. In addition, we use an online updated strategy to obtain an informative dictionary. Some morphological operations are also performed to improve the achieved results. Experimental results demonstrate that the proposed method is effective in segmenting prostate in CT images.

Since LIP classifies each voxel individually, there is no constraint to guarantee the smoothness of the segmentation results. In this paper, we use morphological operations to smooth the results. Future work includes combining the proposed method with the optimization algorithm (i.e., graph cuts [60]), which contains the boundary constraint and thus can generate more smooth results. Moreover, our segmentation framework needs three manual segmentation results of the same patient to construct an initial training dataset. Therefore, the radiologists need to manually segment the prostates in the first three daily scans. Although the workload using the proposed method has been significantly reduced compared with the traditional fully manual contouring, we still try to incorporate the proposed method in the population segmentation. Future study also contains the assessment of population information to initialize the training dataset and gradual addition of the segmented patient-specific data to the dictionary. Thus, manual segmentations in the first three daily scans are not necessary; this step is advantageous to radiologists.

REFERENCES

[1] Q. Feng, M. Foskey, W. Chen, and D. Shen, "Segmenting CT prostate images using population and patient-specific statistics for radiotherapy," *Med. Phys.*, vol. 37, pp. 4121–4132, Aug. 2010.

[2] J. V. Stough, R. E. Broadhurst, S. M. Pizer, and E. L. Chaney, "Regional appearance in deformable model segmentation," *Inf. Process. Med. Imag.*, vol. 20, pp. 532–543, 2007.

[3] S. M. Pizer, P. T. Fletcher, S. Joshi, A. G. Gash, J. Stough, A. Thall, G. Tracton, and E. L. Chaney, "A method and software for segmentation of anatomic object ensembles by deformable m-reps," *Med. Phys.*, vol. 32, pp. 1335–1345, May 2005.

[4] D. Freedman, R. J. Radke, T. Zhang, Y. Jeong, D. M. Lovelock, and G. T. Chen, "Model-based segmentation of medical imagery by matching distributions," *IEEE Trans. Med. Imag.*, vol. 24, no. 3, pp. 281–292, Mar. 2005.

[5] S. Chen, D. M. Lovelock, and R. J. Radke, "Segmenting the prostate and rectum in CT imagery using anatomical constraints," *Med. Image Anal.*, vol. 15, pp. 1–11, Feb. 2011.

[6] M. J. Costa, H. Delingette, S. Novellas, and N. Ayache, "Automatic segmentation of bladder and prostate using coupled 3-D deformable models," *Med. Image Comput. Comput. Assist. Interv.*, vol. 10, pp. 252–260, 2007.

[7] B. C. Davis, M. Foskey, J. Rosenman, L. Goyal, S. Chang, and S. Joshi, "Automatic segmentation of intra-treatment CT images for adaptive radiation therapy of the prostate," *Med. Image Comput. Comput. Assist. Interv.*, vol. 8, pp. 442–450, 2005.

[8] Y. Shi, S. Liao, and D. Shen, "Learning statistical correlation for fast prostate registration in image-guided radiotherapy," *Med. Phys.*, vol. 38, pp. 5980–5991, Nov. 2011.

[9] H. Wang, L. Dong, M. F. Lii, A. L. Lee, R. de Crevoisier, R. Mohan, J. D. Cox, D. A. Kuban, and R. Cheung, "Implementation and validation of a three-dimensional deformable registration algorithm for targeted prostate cancer radiotherapy," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 61, pp. 725–735, Mar. 2005.

[10] U. Malsch, C. Thieke, P. E. Huber, and R. Bendl, "An enhanced block matching algorithm for fast elastic registration in adaptive radiotherapy," *Phys. Med. Biol.*, vol. 51, pp. 4789–4806, Oct. 2006.

[11] S. Liao and D. Shen, "A feature based learning framework for accurate prostate localization in CT images," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3546–3559, Apr. 2012.

[12] J. Zhou, S. Kim, S. Jabbour, S. Goyal, B. Haffty, T. Chen, L. Levinson, D. Metaxas, and N. J. Yue, "A 3-D global-to-local deformable mesh model based registration and anatomy-constrained segmentation method for image guided prostate radiotherapy," *Med. Phys.*, vol. 37, pp. 1298–1308, Mar. 2010.

[13] W. Li, S. Liao, Q. Feng, W. Chen, and D. Shen, "Learning image context for segmentation of the prostate in CT-guided radiotherapy," *Phys. Med. Biol.*, vol. 57, pp. 1283–1308, Mar. 2012.

[14] B. Haas, T. Coradi, M. Scholz, P. Kunz, M. Huber, U. Oppitz, L. Andre, V. Lengkeek, D. Huyskens, A. van Esch, and R. Reddick, "Automatic segmentation of thoracic and pelvic CT images for radiotherapy planning using implicit anatomic knowledge and organ-specific segmentation strategies," *Phys. Med. Biol.*, vol. 53, pp. 1751–1771, Mar. 2008.

[15] S. Liao, Y. Gao, J. Lian, and D. Shen, "Sparse patch-based label propagation for accurate prostate localization in CT images," *IEEE Trans. Med. Imag.*, vol. 32, no. 2, pp. 419–434, Feb. 2013.

[16] N. Makni, P. Puech, R. Lopes, A. S. Dewalle, O. Colot, and N. Betrouni, "Combining a deformable model and a probabilistic framework for an automatic 3-D segmentation of prostate on MRI," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 4, pp. 181–188, Mar. 2009.

[17] D. Pasquier, T. Lacornerie, M. Vermandel, J. Rousseau, E. Lartigau, and N. Betrouni, "Automatic segmentation of pelvic structures from magnetic resonance images for prostate cancer radiotherapy," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 68, pp. 592–600, Jun. 2007.

[18] S. S. Chandra, J. A. Dowling, K. K. Shen, P. Raniga, J. P. W. Pluim, P. B. Greer, O. Salvado, and J. Fripp, "Patient specific prostate segmentation in 3-D magnetic resonance images," *IEEE Trans. Med. Imag.*, vol. 31, no. 10, pp. 1955–1964, Oct. 2012.

[19] N. Betrouni, P. Puech, A. S. Dewalle, R. Lopes, P. Dubois, and M. Vermandel, "3-D automatic segmentation and reconstruction of prostate on MR images," in *Proc. IEEE Eng. Med. Biol. Soc. Conf.*, 2007, pp. 5259–5262.

[20] R. Toth and A. Madabhushi, "Multifeature landmark-free active appearance models: Application to prostate MRI segmentation," *IEEE Trans. Med. Imag.*, vol. 31, no. 8, pp. 1638–1650, Aug. 2012.

[21] A. Ghanei, H. Soltanian-Zadeh, A. Ratkewicz, and F. F. Yin, "A three-dimensional deformable model for segmentation of human prostate from ultrasound images," *Med. Phys.*, vol. 28, pp. 2147–2153, Oct. 2001.

[22] W. D. Richard and C. G. Keen, "Automated texture-based segmentation of ultrasound images of the prostate," *Comput. Med. Imag. Graph.*, vol. 20, pp. 131–140, May-Jun. 1996.

[23] D. Shen, Y. Zhan, and C. Davatzikos, "Segmentation of prostate boundaries from ultrasound images using statistical shape model," *IEEE Trans. Med. Imag.*, vol. 22, no. 4, pp. 539–551, Apr. 2003.

[24] Y. Zhan and D. Shen, "Deformable segmentation of 3-D ultrasound prostate images using statistical texture matching method," *IEEE Trans. Med. Imag.*, vol. 25, no. 3, pp. 256–272, Mar. 2006.

[25] S. Mahdavi, M. Moradi, W. Morris, S. Goldenberg, and S. Salcudean, "Fusion of ultrasound B-mode and vibro-elastography images for automatic 3-D segmentation of the prostate," *IEEE Trans. Med. Imag.*, vol. 31, no. 11, pp. 2073–2082, Nov. 2012.

[26] F. Liu, C. Y. Wee, H. Chen, and D. Shen, "Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's Disease and mild cognitive impairment identification," *NeuroImage*, vol. 84C, pp. 466–475, Sep. 2013.

[27] S. Maggio, A. Palladini, L. D. Marchi, M. Alessandrini, N. Speciale, and G. Masetti, "Predictive deconvolution and hybrid feature selection for computer-aided detection of prostate cancer," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 455–464, Feb. 2010.

[28] F. Liu, H.-I. Suk, C.-Y. Wee, H. Chen, and D. Shen, "High-order graph matching based feature selection for Alzheimer's disease identification," in *Proc. MICCAI*, 2013, pp. 311–318.

[29] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, pp. 91–110, 2004.

[30] E. Tola, V. Lepetit, and P. Fua, "DAISY: An efficient dense descriptor applied to wide-baseline stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 815–830, May 2010.

[31] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 2548–2555.

[32] S. F. Eskildsen, P. Coupe, V. Fonov, J. V. Manjon, K. K. Leung, N. Guizard, S. N. Wassef, L. R. Ostergaard, and D. L. Collins, "BEaST: Brain extraction based on nonlocal segmentation technique," *NeuroImage*, vol. 59, pp. 2362–2373, Feb. 2012.

[33] F. Rousseau, P. A. Habas, and C. Studholme, "A supervised patch-based approach for human brain labeling," *IEEE Trans. Med. Imag.*, vol. 30, no. 10, pp. 1852–1862, Oct. 2011.

[34] M. Kim, G. Wu, and D. Shen, "Sparse patch-guided deformation estimation for improved image registration," in *Machine Learning in Medical Imaging*. New York: Springer, 2012, pp. 54–62.

[35] D. Zhang, Q. Guo, G. Wu, and D. Shen, "Sparse patch-based label fusion for multi-atlas segmentation," in *Multimodal Brain Image Analysis*. New York: Springer, 2012, pp. 94–102.

[36] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, pp. 1182–1195, Dec. 2007.

[37] H. Lee, D. S. Lee, H. Kang, B. N. Kim, and M. K. Chung, "Sparse brain network recovery under compressed sensing," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1154–1165, May 2011.

[38] L. J. Bao, Y. M. Zhu, W. Y. Liu, P. Croisille, Z. B. Pu, M. Robini, and I. E. Magnin, "Denoising human cardiac diffusion tensor magnetic resonance images using sparse representation combined with segmentation," *Phys. Med. Biol.*, vol. 54, pp. 1435–1456, Mar. 2009.

[39] S. Li, H. Yin, and L. Fang, "Group-sparse representation with dictionary learning for medical image denoising and fusion," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 12, pp. 3450–3459, Dec. 2012.

[40] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[41] S. Zhang, Y. Zhan, and D. N. Metaxas, "Deformable segmentation via sparse representation and dictionary learning," *Med. Image Anal.*, vol. 16, pp. 1385–1396, Oct. 2012.

[42] W. Liu, J. He, and S.-F. Chang, "Large graph construction for scalable semi-supervised learning," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 679–686.

[43] K. Engan, S. O. Aase, and J. H. Husoy, "Method of optimal directions for frame design," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1999, pp. 2443–2446.

[44] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[45] Y. Gao, S. Liao, and D. Shen, "Prostate segmentation by sparse representation based classification," *Med. Phys.*, vol. 39, pp. 6372–6387, Oct. 2012.

[46] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 3360–3367.

[47] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, pp. 33–61, 1998.

[48] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[49] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[50] S. Klein, U. A. van der Heide, I. M. Lips, M. van Vulpen, M. Staring, and J. P. Pluim, "Automatic segmentation of the prostate in 3-D MR images by atlas matching using localized mutual information," *Med. Phys.*, vol. 35, p. 1407, 2008.

[51] T. Rohlfing, R. Brandt, R. Menzel, and C. R. Maurer Jr., "Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains," *NeuroImage*, vol. 21, pp. 1428–1442, Apr. 2004.

[52] W. J. Sutherland, D. Spiegelhalter, and M. A. Burgman, "Policy: Twenty tips for interpreting scientific claims," *Nature*, vol. 503, pp. 335–337, Nov. 2013.

[53] F. Perronnin, "Universal and adapted vocabularies for generic visual categorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 7, pp. 1243–1256, Jul. 2008.

[54] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 60–65.

[55] J. Boulanger, C. Kervrann, P. Bouthemy, P. Elbau, J. B. Sibarita, and J. Salamero, "Patch-based nonlocal functional for denoising fluorescence microscopy image sequences," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 442–454, Feb. 2010.

[56] S. Hu, P. Coupe, J. C. Pruessner, and D. L. Collins, "Nonlocal regularization for active appearance model: Application to medial temporal lobe segmentation," *Human Brain Mapp.*, Sep. 2012.

[57] J. X. Mi and J. X. Liu, "Face recognition using sparse representation-based classification on k-nearest subspace," *PloS One*, vol. 8, p. e59430, 2013.

[58] Y. Shin, S. Lee, J. Lee, and H. N. Lee, "Sparse representation-based classification scheme for motor imagery-based brain-computer interface systems," *J. Neural Eng.*, vol. 9, p. 056002, Oct. 2012.

[59] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," *Adv. Neural Inf. Process. Syst.*, pp. 2223–2231, 2009.

[60] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.