Learning image context for segmentation of the prostate in CT-guided radiotherapy

Learning image context for segmentation of the prostate in CT-guided radiotherapy

Wei Li[1,2]**, Shu Liao**[2]**, Qianjin Feng**[1]**, Wufan Chen**[1]
and Dinggang Shen[2,3]

[1] Biomedical Engineering College, Southern Medical University, Guangzhou,
People's Republic of China
[2] IDEA Lab, Department of Radiology and BRIC, University of North Carolina at Chapel Hill,
130 Mason Farm Road, Chapel Hill, NC 27599-7513, USA

E-mail: shibazislw@gmail.com, liaoshu.cse@gmail.com, qianjinfeng08@gmail.com,
wufanchen@gmail.com and dgshen@med.unc.edu

Abstract

Accurate segmentation of the prostate is the key to the success of external
beam radiotherapy of prostate cancer. However, accurate segmentation of the
prostate in computer tomography (CT) images remains challenging mainly due
to three factors: (1) low image contrast between the prostate and its surrounding
tissues, (2) unpredictable prostate motion across different treatment days and
(3) large variations of intensities and shapes of the bladder and rectum around
the prostate. In this paper, an online-learning and patient-specific classification
method based on the location-adaptive image context is presented to deal with
all these challenging issues and achieve the precise segmentation of the prostate
in CT images. Specifically, two sets of location-adaptive classifiers are placed,
respectively, along the two coordinate directions of the planning image space
of a patient, and further trained with the planning image and also the previous-
segmented treatment images of the same patient to jointly perform prostate
segmentation for a new treatment image (of the same patient). In particular,
each location-adaptive classifier, which itself consists of a set of sequential sub-
classifiers, is recursively trained with both the static image appearance features
and the iteratively updated image context features (extracted at different scales
and orientations) for better identification of each prostate region. The proposed
learning-based prostate segmentation method has been extensively evaluated
on 161 images of 11 patients, each with more than nine daily treatment three-
dimensional CT images. Our method achieves the mean Dice value 0.908
and the mean ± SD of average surface distance value 1.40 ± 0.57 mm.
Its performance is also compared with three prostate segmentation methods,

[3] Author to whom any correspondence should be addressed.

indicating the best segmentation accuracy by the proposed method among all methods under comparison.

(Some figures may appear in colour only in the online journal)

1. Introduction

The prostate is a walnut-sized gland that forms part of the male reproductive system. Prostate cancer is one of the most common cancers in males and is a leading cause of male cancer death in the USA (American Cancer Society 2006). External beam radiation treatment can provide a non-invasive and effective therapy for prostate cancer, which mainly consists of two subsequent stages. First, in the planning stage, a planning CT image is acquired from a patient for a clinical expert to delineate the prostate and nearby critical structures manually. In this way, a patient-specific treatment plan, including the intensities and shapes of the radiation beams to be delivered to the prostate, can be designed based on the planning image and its segmented prostate. Second, in the treatment stage, dose is delivered to the prostate in daily fractions during a period of eight to nine weeks, guided by the location of the prostate identified from each treatment image. Specifically, by localizing the prostate in the treatment image, the treatment plan designed in the planning image space can be transformed and applied to the patient on the treatment day with the goal of maximizing the dose delivered to the prostate and minimizing the dose delivered to the surrounding areas such as bladder and rectum. Therefore, the success of image-guided radiotherapy highly depends on the accurate localization of the prostate in the treatment CT images. Moreover, accurate prostate segmentation also has important applications such as dose tracking and plan correction.

However, accurate segmentation of the prostate in CT images is challenging mainly due to four factors. First, the image contrast between the prostate and its surrounding tissues is generally very low in CT images. This can be visually inspected from figure 1(a), where the red contours denote the manually delineated prostate boundaries. Second, the shape and size of the prostate among different patients of different ages vary significantly (Lee *et al* 2001). Third, the prostate motion/deformation (mainly caused by both water in the bladder and fillings/air in the rectum) is complex among different treatment days (Langen and Jones 2001). This factor can also be illustrated in figure 1(b) which shows the prostate regions of two treatment images (in white and gray, respectively) of a same patient after rigid alignment based on their pelvic bone structures. It can be observed that, even after rigid alignment based on bone structures, the relative prostate motion/deformation is still large. Fourth, the uncertain presence of rectum gas makes image appearance change dramatically across different treatment days. This factor can be again visualized by figure 1(a), which shows large variations of the shape and appearance of the rectum gas (black) in different treatment images.

So far, several categories of methods have been proposed for prostate segmentation from CT images. The first and the most popular category of prostate segmentation methods is the deformable-model-based segmentation methods (Broadhurst *et al* 2005, Freedman *et al* 2005, Rousson *et al* 2005, Feng *et al* 2010, Chen *et al* 2011). For instance, Broadhurst *et al* (2005) proposed a deformable model based on histogram-statistics appearance model at multiple scales for prostate segmentation. Freedman *et al* (2005) proposed a segmentation algorithm based on matching the probabilistic distributions of photometric variables that incorporates the learned shape and appearance models for the objects of interest. Also, Feng *et al* (2010) presented a deformable model based on the statistical shape and local image

(a)

(b)
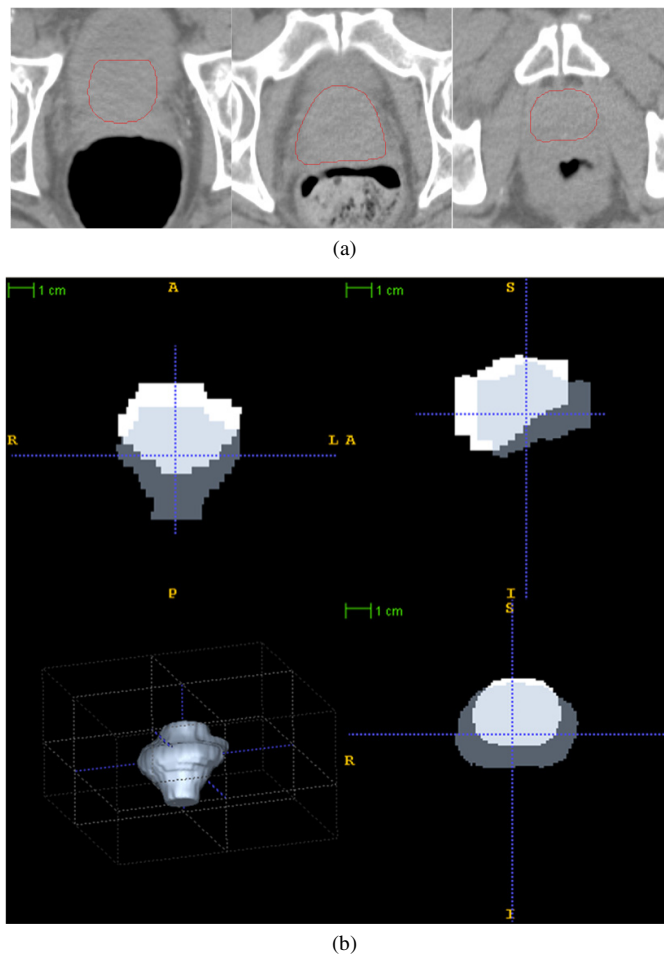
Figure 1. (a) Slices from the three treatment images showing the low contrast between the prostate and surrounding organs (i.e. bladder, rectum and muscles), as well as large variation of the shape and appearance of the rectum gas (black). Here, the three red curves denote the prostate boundaries manually delineated by a clinical expert. (b) Overlapping display of the two prostate shapes (white and gray) from two treatment images of the same patient, showing large motion/deformation even after bone-based rigid alignment.

appearance information which leverages both population and patient-specific statistics to segment the prostate in the CT images. Chen *et al* (2011) presented a prostate segmentation algorithm in which both the anatomical constraints from bones and a learned appearance model are incorporated under the Bayesian framework. Rousson *et al* (2005) proposed a Bayesian framework to impose a shape constraint on the prostate and couple its extraction with that of the bladder, where the prostate and bladder are represented by two level-set surfaces with non-overlapping constraint. Overall, these deformable-model-based segmentation methods, which consider the statistical information extracted from both organ shape and image appearance, have been proven effective in segmenting the prostate from CT images. However, this category of methods normally requires a good initialization and a specific optimization strategy in order to obtain accurate and robust segmentation results.

The second category of prostate segmentation methods is registration-based segmentation (Davis *et al* 2005, Wang *et al* 2005) methods. For instance, Davis *et al* (2005) developed a registration-based prostate localization approach for CT images by combining a fluid-based image registration algorithm and a bowel gas deflation algorithm. Wang *et al* (2005) implemented a three-dimensional (3D) deformable registration algorithm to aid dose tracking for the targeted prostate radiotherapy. However, the uncertain presence of water in the bladder and fillings/air in the rectum makes the accurate registration of two treatment images difficult.

Besides, other methods have also been developed based on mathematical morphology (Mazonakis *et al* 2001), combination of deformable model and registration (Chen *et al* 2009) and different optimization strategies (Ghosh and Mitchell 2006, Song *et al* 2009). Note that, besides segmenting the prostate from CT images, there are also other methods segmenting the prostate from other imaging modalities such as MR (Langerak *et al* 2010, Gao *et al* 2010) and US (Shen *et al* 2003, Zhan and Shen 2006, Zhan *et al* 2007). In this paper, we focus on developing a learning-based method for segmenting the prostate from 3D CT images.

The context and high-level information plays an important role in object recognition and scene understanding, especially for medical image analysis (Oliva and Torralba 2007, Belongie *et al* 2002). Several models in statistical classification field, such as Markov random field (Geman and Geman 1984) and conditional random field (Lafferty *et al* 2001), have been proposed to capture the context information using a neighborhood (clique) structure with a limited and fixed topology. However, this topology limits the modeling capability. Therefore, other alternative methods were proposed to extract context information more effectively. For instance, Jin and Geman (2006) proposed a framework called composition machine to construct the hierarchical probabilistic image models in order to accommodate arbitrary contextual relationships. Loog and Ginneken (2006) proposed an iterative contextual pixel classification method to segment the posterior ribs in chest radiographs, in which the classification of each pixel was iteratively updated based on the context information. Hupse and Karssemeijer (2009) developed a set of context features that represent suspiciousness of tissue for detection of malignant masses in mammograms and achieved performance improvement when multiple views were used for computing context features. Recently, similar to Loog and Ginneken (2006), a supervised machine-learning method with context information has also been proposed in Tu and Bai (2010). In this work, the classification probability maps created by the learned classifiers were used as context information in a recursive manner to improve classification result. However, due to the use of a single global classifier, the boundary of object in their classification result, where the image appearance and context have large variation, is still blur.

To deal with difficulty in segmentation of the prostate in CT images, we propose a patient-specific classification method to segment the prostate from 3D CT images by incorporating both appearance features (section 2.3.1) and context features (section 2.3.2) into an online-learning framework. Our method avoids the negative side effect of inaccurate initialization and improves the boundary segmentation accuracy. In particular, first, we learn a different classifier for each local region of the prostate, rather than a single global classifier for the whole prostate. These local classifiers are named as *location-adaptive* classifiers in this paper. Second, we employ patient-specific information and the online-learning strategy. In other words, we improve the segmentation accuracy by learning information from the previous-segmented treatment images of a patient, thus achieving accurate segmentation for the new treatment images of the same patient in the following days. Third, we extend the auto-context method of Tu and Bai (2010) by introducing context information from both original image and classification probability. Different from the traditional learning-based methods, each *location-adaptive* classifier itself consists of a series of *sub-classifiers*. This means that a series
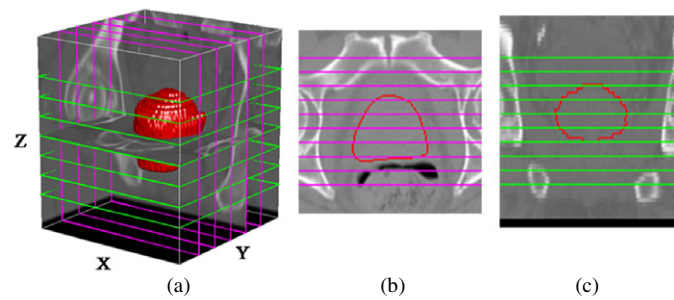
Figure 2. Illustration of location-adaptive classifiers placed along the two coordinate directions (*Y* and *Z*) around the prostate region in the extracted ROI. (a) A 3D ROI around a prostate (red object), where each magenta/green rectangle represents the place where we will place a location-adaptive classifier. (b) and (c) The sample slices displayed in the views of the axial (*Z* coordinate) and coronal (*Y* coordinate) directions, respectively. The red contours represent prostate boundaries. The magenta/green straight line represents where a location-adaptive classifier will be placed.

of learning processes are performed iteratively to produce a set of sequential *sub-classifiers* to compose a *location-adaptive* classifier. By using this learning process, the classifiers can effectively gather information locally and propagate it to other places. Last, context features are collected from a large number of locations far away from a pixel of interest. By using the context features, the partial shape information can also be learned by our classifiers.

Specifically, in our method, two sets of *location-adaptive* classifiers along the two coordinate directions are trained within a region of interest (ROI) around the prostate which is predefined, as shown in figure 2. The ROI is separated into parallel rectangular blocks in two coordinate directions. Each rectangular block has a corresponding *location-adaptive* classifier. Each *location-adaptive* classifier is trained only on two-dimensional (2D) slices within its corresponding rectangular block. The classification probability maps produced by the two sets of *location-adaptive* classifiers will be stacked along each direction first and then fused into a single 3D classification probability map for the daily image under segmentation. Next, this single 3D classification probability map is further refined by a 3D level-set algorithm (Chan and Vese 2001) for final segmentation. This two-coordinate-direction-based learning strategy encodes different local regions of the whole ROI in the two directions. It enables our *location-adaptive* classifiers to learn the complementary information about the prostate in 3D space in a more effective way. Note that the two coordinate directions (i.e. axial and coronal directions) are selected based on the anatomical symmetry of pelvis and the knowledge of observed prostate motion (i.e. along superior/inferior and anterior/posterior directions (Langen and Jones 2001)). Along these two directions, the symmetric shape information (e.g., bones and muscles) and motion information can be both learned by our classifiers based on the features designed in our method.

Notably, our proposed method for prostate segmentation is designed under the scenario of clinical daily radiotherapy. So, the whole method can be further optimized by an offline verification process, which allows the physicist to check the accuracy of automatic segmentation by our method. In this way, only those accurate segmented previous images will be used to retrain our *location-adaptive* classifiers to gradually improve the accuracy of automatic segmentation for the following daily images. Furthermore, our method requires only the first several treatment days to be segmented manually to initialize the whole training procedure in our method. Compared to the requirement of segmenting all treatment images (i.e. about 45 images) in the course of a patient's radiotherapy in the current clinical application,
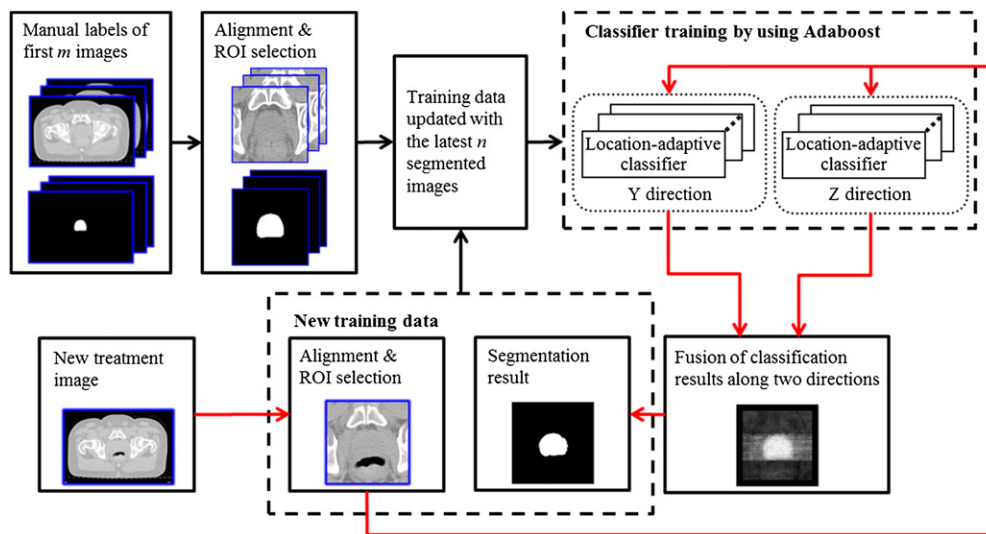
Figure 3. Flowchart of the proposed patient-specific prostate segmentation algorithm for 3D CT images. The black solid arrows indicate the training process, while the red solid arrows indicate the segmentation of a new treatment image acquired at the subsequent treatment day.

our method can significantly reduce manual efforts and thus make the whole process more efficient.

Our method has been extensively evaluated on 11 patients, each with more than nine 3D treatment CT images of the prostate. It is further compared with three state-of-the-art prostate segmentation algorithms for CT images. Experimental results show that our method can consistently achieve the highest segmentation accuracy among all the methods under comparison. We note that a preliminary version of our method was previously reported in a conference paper (Li *et al* 2011). The present paper provides a complete description of the method and further presents more experimental results to evaluate its performance.

The rest of this paper is organized as follows. Section 2 introduces the proposed algorithm. Section 3 provides and analyzes the experimental results. Section 4 provides some discussions on the proposed method. Section 5 concludes this paper.

2. Methodology

Our method mainly consists of an online training stage and a segmentation stage in the course of radiotherapy for a patient. The flowchart of the proposed method is shown in figure 3. Hereafter, we denote and index all images starting from number 1 for the sake of simplicity (e.g., the planning image is denoted by image 1 and the first treatment image is denoted by image 2).

Training stage. Two sets of location-adaptive classifiers are initially trained by the patient's first m daily scans (e.g., $m = 3$ in this paper) with manual segmentations. The classifiers will be further retrained and updated based on the n latest-acquired scans (e.g., $n = 4$) of the same patient and their corresponding segmentation results produced by the proposed method during the radiotherapy. That is, with more and more subsequent treatment images acquired

and automatically segmented, our classifiers can be updated with the automatic segmentation results of the n latest-acquired treatment images, along with the first m daily scans. In this way, our trained classifiers can track the up-to-date patient-specific image context and appearance information to improve the classification accuracy for the next daily treatment image. In this paper, we set $m = 3$, which means that we use only the first three images and their corresponding manual segmentations to initialize our training process in the whole course of radiotherapy for a specific patient. Then, the classifiers are updated by using the automatic segmentation results (with or without offline verification) of the subsequent daily treatment images produced by the pre-trained classifiers. Our method can work automatically for segmenting a treatment image after retraining classifier for the next day, which is important for a segmentation method to be practical for clinical application.

Specifically, the patient's poses in these images need to be aligned first to remove whole body rigid motion. Therefore, the pelvic bone structures of all training images are rigidly aligned to the pelvic bone structures in the planning image of the same patient. After alignment, the remaining movement and shape variation of the prostate in each training image are mainly caused by other facts such as water in the bladder and gas/fillings in the rectum. The goal of our classifiers is to capture these movements and shape variations in all aligned training images.

Next, a 3D ROI in the planning image is determined around the prostate (e.g., shown in figure 2) to reduce the computation burden in both training and segmentation stages. The position of ROI is determined based on the manual segmentation of the planning image, and the size of ROI (130 mm $\times$ 130 mm $\times$ 180 mm) is set large enough to include the whole prostate in all daily scans. With this defined ROI, we can use it to extract a same-sized subvolume from all rigidly aligned training images, which can be used to form the training data set for training our location-adaptive classifiers.

To train each location-adaptive classifier, the training pixels are first sampled from the corresponding slices for which this classifier is responsible. Then, we use features of the training pixels and the respective segmentation labels to train each classifier by using a machine-learning technique (Tu and Bai 2010). Each trained classifier is retrained when the training data set is updated by a new treatment image and its automatic segmentation result.

Segmentation stage. Each new treatment image is first rigidly aligned onto the planning image space based on the pelvic bone structures. Then, the ROI defined in the training stage will be used to extract the target region in the new aligned treatment image, which will be further classified/labeled by the two sets of our trained location-adaptive classifiers (along the two coordinate directions). The two prostate classification probability maps produced by the two sets of classifiers are stacked along their directions, respectively, to achieve two 3D classification results. Then, the two 3D classification results can be finally fused into a single 3D classification result and further segmented by a 3D level-set-based segmentation technique extended from Chan and Vese (2001) to produce the final binary segmentation result for the new treatment image.

The above training stage and the segmentation stage are recursively repeated in the whole course of radiotherapy of a patient with the aid of an automatic or manual result verification process. For example, if the period of radiotherapy for a patient consists of 45 days and the number of images required for training initialization is set to 3, the clinician needs to manually label the first three images to start our method. Then, the classifiers for segmenting the fourth daily treatment image are trained based on the segmentations of the first three images. After automated or manual verification, the segmentation result of the fourth daily image can be added into our training data set, and our classifiers for the fifth daily image are retrained based

on the updated training data set. This procedure is repeated in all following treatment days. Since the retraining process can be done overnight, the updated classifiers can be used for prostate segmentation for the next treatment day. Due to this repeated retaining procedure, our method is thus called an online-learning method.

2.1. Auto-context classification formulation

In this section, we present the formulation of auto-context classification and notations used in this paper. For more details refer to Tu and Bai (2010). Here, a 2D/3D image is represented as a function $X = (x_1, \ldots, x_N)$ mapping from the spatial domain $\Omega \subset \mathbb{R}^2/\mathbb{R}^3$ to an intensity value in $\mathbb{R}$, where N is the number of pixels in the image. For each image X, its corresponding ground-truth segmentation/label image $Y = (y_1, \ldots, y_N)$ is also available (i.e. manual segmentation result in the first several scans), with the label value y_j of each pixel x_j ($j = 1, \ldots, N$) belonging to $\{0, 1\}$.

We aim to train a classifier to find an optimal classification result through maximum *a posteriori* given a new image X, i.e. $\text{argmax}_Y(p(Y|X))$. Here, the marginal distribution is employed to calculate $p(Y|X)$. To better approximate the marginal distribution, an auto-context model is proposed, in which context features are extracted from the classification probability map F^t produced by a traditional classifier such as SVM or AdaBoost at each iteration t and further integrated into the training set recursively to provide context information. The iterative algorithm updates $p^{(t)}(y_j|B_j(X), F^{t-1})$ to approach $p(y_j|X)$ asymptotically without explicit integration, where $B_j(X)$ denotes the pixels in an image patch centered at the jth pixel. This iterative process can be expressed by

$$p^{(t)}(y_j|B_j(X), F^{t-1}) \rightarrow p(y_j|X) = \int p(y_j, Y_{-j}|X)\, \mathrm{d}Y_{-j}, \tag{1}$$

where Y_{-j} denotes the group of segmentation labels of all the pixels except y_j.

2.2. Location-adaptive classifiers

In this section, we describe how to train each location-adaptive classifier. First, the corresponding slices used to train a specific location-adaptive classifier are selected from all ROIs in the training data set and further denoted as X_i ($i = 1, \ldots, M$), where $M \geqslant 1$ is the number of the slices. For example, suppose that there are in total four training ROIs available and each location-adaptive classifier is responsible for five slices; then the number of slices M used to train each local-adaptive classifier is $M = 4 * 5 = 20$.

Details of the training instance sampling procedure are summarized as follows. First, different types of organs and tissues such as bone and air in each prostate training image can be identified based on an intensity thresholding process (i.e. more details of this step can be referred to section 3). Then, we can randomly draw samples from regions belonging to different types of organs and tissues with the following constraints enforced: (1) half of the training instances are constrained to be positive instances belonging to the prostate region; (2) 50% of the negative instances are constrained to lie within 20 mm around the prostate, 30% of the negative instances are constrained to lie in bones and air and 20% of the negative instances are constrained to lie in other regions. Figure 4 illustrates this sampling procedure.

For a specific location-adaptive classifier, assume that K_i training instances are obtained for each training slice X_i. All training instances (i.e. sampled training pixels) are denoted as $\{x_{ij}, i = 1, \ldots, M, j = 1, \ldots, K_i\}$. Then, the training set can be represented as $S = \{(y_{ij}, \boldsymbol{v}_{ij}), i = 1, \ldots, M, j = 1, \ldots, K_i\}$ for each location-adaptive classifier, where $\boldsymbol{v}_{ij}$ is the corresponding feature vector of x_{ij}. The image appearance features (refer to
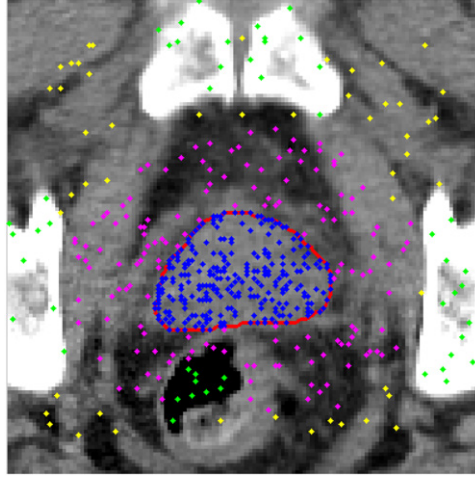
Figure 4. An illustration of the training instance sampling procedure. The red curve is the boundary of the prostate. Pixels of different colors represent the training instances drawn from different regions belonging to different types of organs and tissues (i.e. prostate (blue), region within 20 mm around the prostate (magenta), bone and air (green) and other regions (yellow)).


Given training set $S_0 = \{(y_{ij}, A_{ij}, C_{ij}^0)\}$.
For $t = 0 \cdots T$:
   1) Train a classifier L_t on S_t;
   2) Use the trained classifier L_t to compute the new classification map F_i^{t+1} for each image slice X_i;
   3) Compute context features C_{ij}^{t+1} from the new classification map F_i^{t+1} and X_i;
   4) Construct the new training set $S_{t+1} = \{(y_{ij}, A_{ij}, C_{ij}^{t+1})\}$;
End For

Figure 5. Recursive training process of a location-adaptive classifier.

section 2.3.1) of each training pixel x_{ij} are denoted as A_{ij} ($i = 1, \ldots, M, j = 1, \ldots, K_i$). Also, the corresponding context features (described in section 2.3.2) of x_{ij} at iteration t are denoted as C_{ij}^t ($i = 1, \ldots, M, j = 1, \ldots, K_i, t = 0, \ldots, T$), which are computed from the slice classification map F_i^t and X_i. Accordingly, the feature vector v_{ij} for the pixel x_{ij} at iteration t is now represented by both appearance features and context features, i.e. $v_{ij} = (A_{ij}, C_{ij}^t)$. Then, the training set becomes $S_t = \{(y_{ij}, A_{ij}, C_{ij}^t)\}$ at iteration t. Finally, we can iteratively train our location-adaptive classifier using the AdaBoost algorithm (Freund and Schapire 1995). Figure 5 outlines our training process for a location-adaptive classifier. At the end of the training procedure, we get a sequence of classifiers L_t ($t = 0, \ldots, T$), namely the *sub-classifiers*, for each location-adaptive classifier, where T denotes the last iteration and in total $T + 1$ iterations are used. Note that for the first *sub-classifier*, we use the *default* uniform classification probability map (i.e. F_i^0 is a uniform classification probability map) since the new classification probability map is not computed yet.
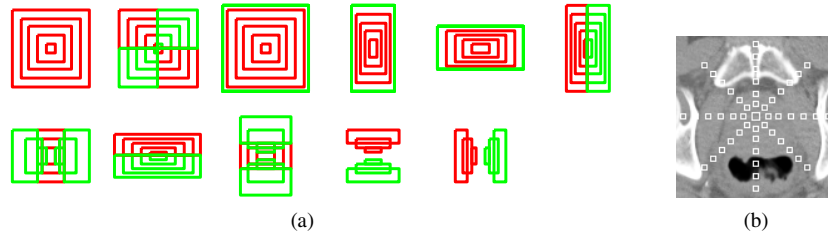
Figure 6. (a) Examples of some 2D Haar-like descriptors. (b) Context sample pattern.

2.3. Feature extraction

Two types of features are employed in our learning-based segmentation algorithm. The first type of feature is image appearance features, which are calculated from intensity images. The second type of feature is context features, which are computed both from intensity images and classification probability maps. These two types of features construct our feature pool.

2.3.1. Appearance features. The appearance features used in this paper consist of both Haar features (Viola and Jones 2004) and histogram of oriented gradient (HOG) features (Dalal and Triggs 2005). The two types of appearance features are employed to capture the local information around each voxel in a CT image. Haar features are widely used in object recognition, which can be computed at different scales of interest and with high speed by using integral images. A Haar-like feature considers adjacent rectangular regions at a specific location in a detection window, sums up the pixel intensities in these regions and calculates the difference between them. In this paper, we extract Haar features from different scales of interest to facilitate our prostate classification. Figure 6(a) shows some 2D Haar-like descriptors used in our method.

The other appearance features adopted in this paper are the HOG features. The HOG features can be extracted by dividing the image into small connected regions, called cells, and then compiling a histogram of gradient directions for the pixels within each cell. To improve the segmentation accuracy, the local histograms are contrast-normalized for correcting intensity inhomogeneity in the prostate CT images by calculating the mean of the intensity (i.e. average intensity) across a larger region of the image, called a block, and then, this value is used to normalize all cells within the block. Our HOG features consist of nine cells by nine histogram bins. The size of the block is set to be the size of image patch used for computing appearance features. In this paper, the size of the image patch is set empirically to be 21×21. Finally, the coordinates of each pixel are also adopted as the appearance features in our method to learn the movement information of the prostate within the patient's pelvis.

2.3.2. Context features. Context features are used to capture global scene information around each pixel in a CT slice. A context sample pattern is employed to extract information at different pixels around each pixel as its context features. Different from the local appearance features, the context features of a pixel are extracted from a large region centered at the pixel rather than a small neighborhood region. Specifically, for a pixel, eight rays at $45°$ intervals are extended out from the current pixel, and then, we sparsely sample the context locations on these rays at different radii as illustrated in figure 6(b). In this paper, a radius sequence $R = \{R_u, u = 1, \ldots, 15\} = \{4, 5, 6, 8, 10, 12, 14, 16, 20, 25, 30, 40, 50, 60, 80\}$ is used for choosing context locations on each ray. Furthermore, the location of the current pixel and

pixels in its 3×3 neighborhood are also included in the whole context sample pattern. For prostate segmentation, the relative coordinates $(\alpha_{uw}, \beta_{uw})$, i.e. $(0, 0)$ is the center pixel, of context (except 3×3 neighborhood of the center pixel) are constructed with the following formulation:

$$\begin{cases} \alpha_{u,w} = \left\lfloor R_u \cos\left(\frac{\pi}{4} * w\right) + \frac{1}{2} \right\rfloor \\ \beta_{u,w} = \left\lfloor R_u \sin\left(\frac{\pi}{4} * w\right) + \frac{1}{2} \right\rfloor \end{cases}, \qquad u = 1, \ldots, 15, \quad w = 0, \ldots, 7 \quad (2)$$

where $\lfloor \cdot \rfloor$ gives the floor of real number.

For each of the above context locations in the context sample pattern, its classification probability and the mean probability within a 3×3 window are adopted as the context feature as in Tu and Bai's (2010) work. However, different from their work, we also extract context information from the intensity image at these selected context locations rather than from the classification probability only. In other words, we consider that there is important context information included in the intensity variation of image at these selected context locations. Therefore, the absolute intensity variations between the context locations and the current pixel of interest are also included as context features in our algorithm. The construction and update of the context features are shown in figure 7. Thus, the *location-adaptive* classifiers can learn the location information from the air or bone regions (which often have large intensity variation from the current pixel of interest) through our designed context features. These context features implicitly represent the shape and configuration information of the prostate. Due to the nature of iteratively learning context information, our location-adaptive classifiers can achieve better performance. This can be seen from figure 7, where the prostate region in the sequential classification maps becomes more and more prominent and the boundary between the prostate and background becomes sharper and sharper. Most importantly, our proposed algorithm is general which can incorporate other types of features for possible improvement of prostate segmentation.

3. Experimental results

In this section, several experiments were conducted to evaluate the proposed method. Since our classifiers were designed to learn the patient-specific information, all training and segmentation processes were done on the images from the same patient. In the following experiments, each treatment image was first rigidly aligned with the planning image based on pelvic bone structures. The alignment process was performed using the FLIRT toolkit in Fischer and Modersitzki (2003). The bone and air regions in each image were detected according to their values in Hounsfield units (HU) by using the predefined thresholds: $+400$ and -150 (i.e. regions with intensity of more than $+400$ were determined as bones, while regions with intensity of less than -150 were determined as air). Given the segmentation ground truth of the planning image, we defined an ROI with a size of 130 mm $\times$ 130 mm $\times$ 180 mm around the prostate, which was used in the training and segmentation stages to trim out ROIs from the aligned treatment images. Note that the segmentation ground truth of each image is provided by one clinical expert in this study, and different clinical experts could provide different segmentation ground truths. However, the corresponding location-adaptive classifiers will be trained consistently with the segmentation ground truths provided by the specific clinical expert.

According to the statistics of position variation of prostates in different daily images based on our data set, we found that the prostate mainly moves in the Y (coronal) and Z
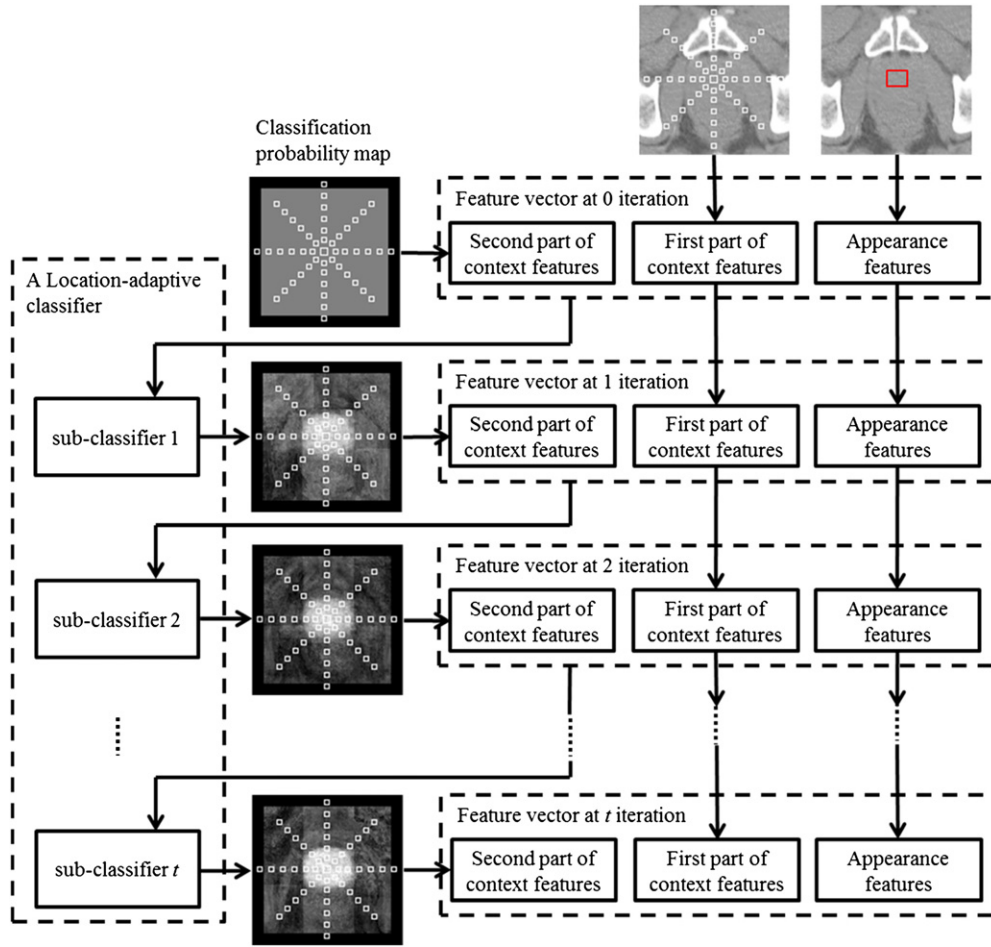
Figure 7. Illustration of construction and update of the features used in our method. The red rectangle in the top-right image represents the patch used for calculating appearance features. The small rectangles located along rays (i.e. context sample pattern) are the positions where context features are extracted. In the column of the classification probability map, the classification map is iteratively updated from a uniform probability map (top) to a better and better classification result (bottom). Note that the classification probability map in the 0 iteration is a default uniform classification probability map.

(axial) directions (Langen and Jones 2001). So the *location-adaptive* classifiers can learn the most important movements of prostates when they are trained on slices along the Y and Z directions. Adding another set of *location-adaptive* classifiers, i.e. trained along the X (sagittal) direction, may slightly increase the performance of our method. However, to save computation time, we do not include classifiers along the X direction. The numbers of location-adaptive classifiers along the Y and Z directions were determined based on the resolutions of patient image along the two directions and also the size of ROI determined in the planning image. For our data set, 12 and 20 classifiers were placed along the Y and Z directions, respectively. Each location-adaptive classifier includes a series of three *sub-classifiers*, each of which is an AdaBoost-based classifier, consisting of 100 three-level decision trees as weak classifiers.

The number of the features used in the experiments was 568 in total. The image appearance features consisted of 118 Haar features and 81 HOG features, which were all calculated from the 21×21 image patch of each pixel. The number of context features was 369. The parameters of the 3D level-set-based segmentation technique used in this paper were set as follows: $\mu = 0.02 * 255^2$, $\nu = 0$, $\lambda_1 = \lambda_2 = 1$. The first two parameters (i.e. μ, ν) denote the weightings given to surface area and inside volume of the level-set surface for regularization. The parameters λ_1 and λ_2 are the weightings given to the two fitting terms in the energy function (Chan and Vese 2001).

3.1. Data and evaluation measures

Our data consist of 11 patients, each with more than nine daily prostate CT scans, with a total image count of 161. Each image has *XY* size 512×512 and 34 to 97 slices with a voxel resolution of 1 mm $\times$ 1 mm $\times$ 3 mm. For each patient, the first three images and their manual segmentations were used to initialize the patient-specific learning process. To validate the proposed algorithm quantitatively, the expert manual segmentation results for each image were used as ground truth. Three quantitative measures were used to evaluate the performance of our method as listed below.

(1) The Dice similarity coefficient (Dice) (Dice 1945). The Dice is defined as $(2 * V_{a \cap g})/(V_a + V_g)$, where $V_{a \cap g}$ denotes the number of voxels classified as the prostate by both the proposed algorithm (*a*) and the ground truth (*g*), and V_a and V_g denote the number of voxels classified as the prostate by the proposed algorithm and the ground truth, respectively.
(2) The average surface distance (ASD) between the automated and manual segmentation results. The ASD is calculated as the average distance between the surfaces of the ground truth and the automatically segmented prostates along 300 rays evenly directed to the spherical surface from the ground-truth centroid.
(3) The centroid distance (CD) is calculated to measure the segmentation accuracy with respect to the centroid measurement since the centroid is of particular interest for a radiation oncologist when designing a treatment plan. We compare the distance between the centroid of the ground truth with the centroid of the automatic segmentation results. The differences in the lateral (*X*), anterior–posterior (*Y*) and superior–inferior (*Z*) directions are measured separately.

3.2. Location-adaptive classifiers versus global classifier

To justify that our location-adaptive classification scheme is better than using a single global classifier, we designed two experiments. In the first experiment, we compared our method with a 3D population-based global classifier trained on the population images (i.e. different patients' images were used for training). For training the 3D global classifier, similar features in 3D image space (such as Haar, HOG and context features) were employed. Then, 80 images out of our data set were randomly selected to train the 3D global classifier. Some typical results produced by our method and the 3D global classifier are shown in figure 8 for visual comparison. As indicated in figure 8, the results by our method are significantly better than those by the 3D global classifier. This indicates that the prostate cannot be segmented accurately if using only the population-based global classifier. It also indicates that the information learned from the population images cannot improve the classification performance due to large variations of image appearance across different patients.
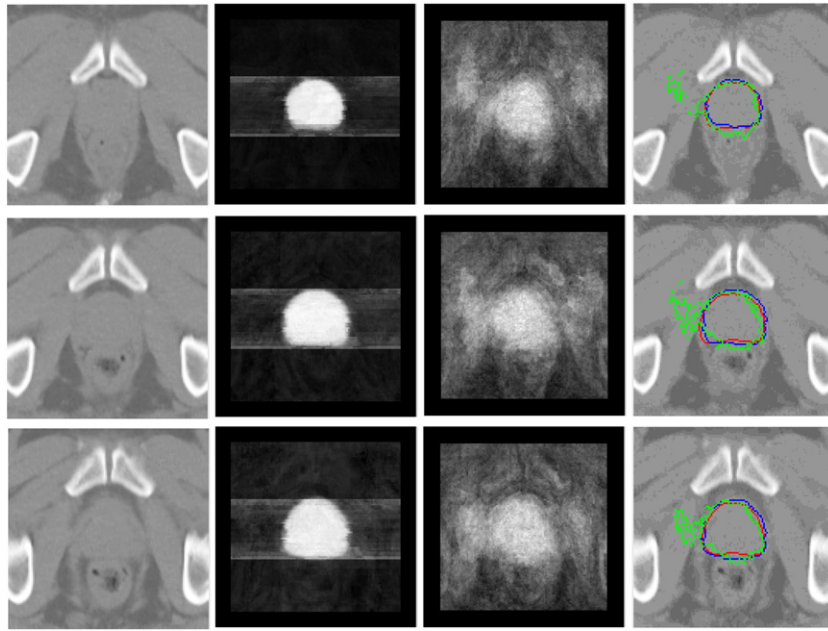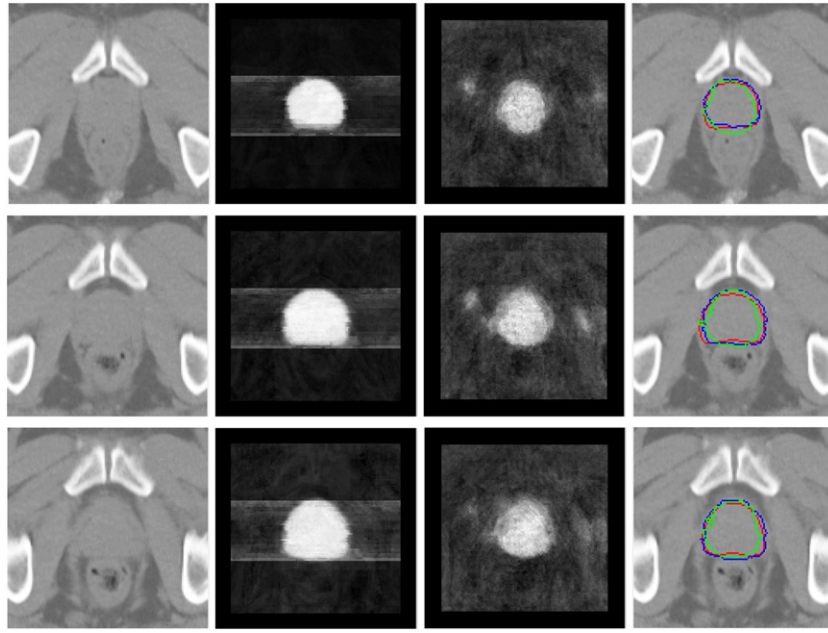
Figure 8. Classification results produced by our proposed method and a single 3D global classifier trained on the population images. From left to right columns, intensity slices, classification results of our classifiers, classification results of the 3D global classifier, and intensity slices overlapped by ground truth (red curves), segmentation results of our method (blue curves) and segmentation results of the 3D global classifier (green curves) are displayed.

In the second experiment, we compared our method with a 3D patient-specific global classifier, trained on the first five images from the same patient, which has in total 16 images. The similar features as the 3D population-based global classifier were employed again. Then, 11 images from the same patient (except the first five images used for training) were used as testing images. Some typical results are shown in figure 8. As indicated in figure 9, our method also performed much better than the 3D patient-specific global classifier, especially around the boundary of the prostate. This again approves the difficulty of training only a global classifier for accurate prostate segmentation. On the other hand, we can observe that the 3D patient-specific global classifier is better than the 3D population global classifier, which means that the performance of classifiers can be improved by using the information learned from images of the same patient. In figure 10, the Dice values of 11 segmentation results by our method and by the 3D patient-specific global classifier are displayed. The Dice values of our method are significantly ($p < 0.001$) greater than those of the 3D patient-specific global classifier.
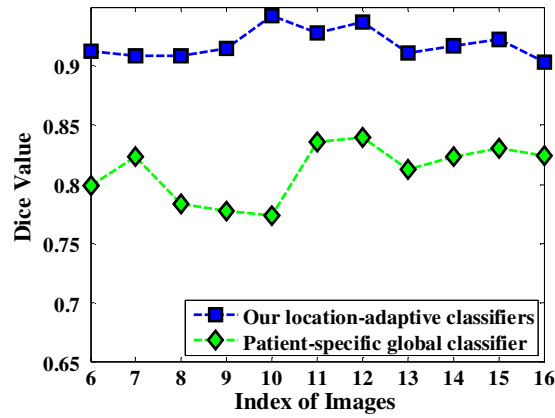
3.3. Role of context features

By using image context features, our classification results of image slices have more uniform high values in the region of the prostate and low values in the background. In training the location-adaptive classifiers, the context features play an important role in training the three sequential *sub-classifiers* used in this paper. In our method, the context features of each pixel consist of two parts at each context location (see section 2.3.2). Figure 11 shows an example

Figure 9. Classification results produced by our proposed method and a 3D patient-specific global classifier trained on images from the same patient. From left to right columns, intensity slices, classification results of our proposed method, classification results of the 3D patient-specific global classifier, and intensity slices overlapped by ground truths (red curves), segmentation results of our method (blue curves) and segmentation results of the 3D patient-specific global classifier (green curves) are displayed.


Figure 10. Dice values of segmentation results of 11 images from the same patient by our location-adaptive classifiers and the 3D patient-specific global classifier.

about the numbers of the two parts of context features used by the three sequential *sub-classifiers* in a learned location-adaptive classifier. From figure 11, it can be observed that different sets of context features are selected by each sub-classifier. It can also be observed that features from the two parts of context are used in *sub-classifiers* jointly. However, the

(a)   (b)   (c)
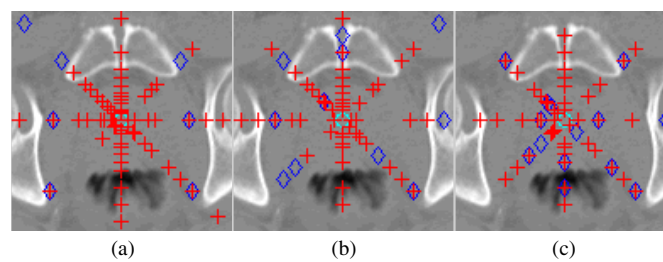
Figure 11. An example illustrating the locations of two parts of the context features learned by the three sub-classifiers of a location-adaptive classifier at three different iterations. (a)–(c) Illustration of the context features learned by the first, second and third sub-classifiers. The green circle, blue diamond and red '+' represent the center of context features, context features calculated from the classification maps and the absolute intensity difference between the current pixel of interest and the context location, respectively.

(a)   (b)   (c)   (d)   (e)
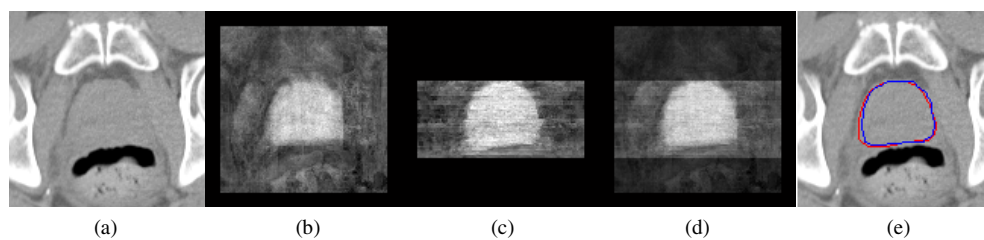
Figure 12. Demonstration of the superior classification results obtained after fusing the classification results produced by the two sets of classifiers placed along the two coordinate directions. (a) Original image. (b) Classification result by the classifiers placed along the *z* coordinate direction. (c) Classification result by the classifiers placed along the *y* coordinate direction. (d) Fusion of (b) and (c). (e) Ground truth of the prostate contour (red) and the estimated prostate contour (blue) by our proposed segmentation algorithm.

number of context features from the second part is larger than that from the first part, indicating that the absolute intensity variations in context features are more important than other context features in our method.

3.4. Fusion of results of location-adaptive classifiers

To illustrate the effectiveness of the fusion of the classification probability maps produced by the two sets of classifiers along the two coordinate directions, a visual comparison between the results of fusion, *Y* direction and *Z* direction is given in figure 12. Also, a Dice ratio comparison between their results for all images of 11 patients is shown in figure 13. It can be observed from both figures 12 and 13 that the segmentation results after classification fusion is better than the segmentation results directly based on the classification along each direction.

3.5. Segmentation accuracy of the proposed algorithm against different number of training images

The number of available training images is also a critical factor affecting the segmentation accuracy of the classifiers. Intuitively, by increasing the number of available training images, the segmentation accuracy of the classifiers can be improved as more information can be
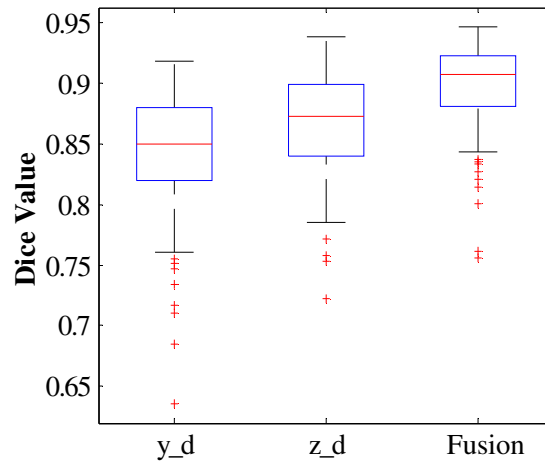
Figure 13. (a) Dice measure of segmentation results by using only *y* direction classifiers (y_d), using only *z* direction classifiers (z_d), and using fusion of *y* and *z* directions classifiers together (Fusion). The horizontal lines in the box plots represent the lower quartile, median and upper quartile values, respectively. The whiskers show the extent of the rest of the data. Outliers, which fall outside 1.5 times the interquartile range, are denoted with the '+' symbol.

learnt from the training images. However, in real-world clinical application, the number of available training images is limited. Therefore, the robustness of the segmentation algorithm against the number of available training images is also interesting to explore. In this paper, we tested the segmentation accuracy of the proposed algorithm using different number of training images. In particular, we selected three patients (patient IDs 2, 3 and 8) with large variations of segmentation accuracy across different treatment images (i.e. these three patients have large variations of Dice measures compared to other patients (see figure 17(a))). Next, we selected one particular treatment image from each of the three selected patients (i.e. image 11 of patient 2, image 12 of patient 3 and image 14 of patient 8). Then, we performed the prostate segmentation on each of these three images by using our location-adaptive classifiers trained by the latest two to six treatment images of the same patient. The corresponding Dice measures are shown in figure 14. It can be observed that the Dice values become steady when the number of training images is 4 or larger. Therefore, the number of the training images was set to 3–7 for the tradeoff between training time and segmentation accuracy in all our experiments, as well as considering the increasing available treatment images for training and clinical feasibility.
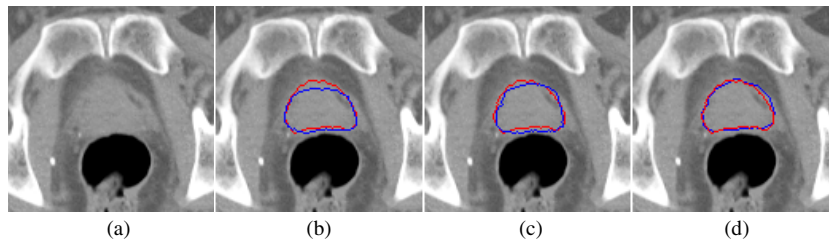
The improvement of the segmentation accuracy by increasing the number of training images is visually demonstrated in figure 15. From figure 14, it can be observed that the segmentation result is improved as the number of training images is increased, i.e. the prostate boundaries of the automatic segmentation results gradually approach to the prostate boundary of the ground truth.

3.6. Evaluation of the online-learning mechanism

The online-learning mechanism can avoid the trained classifiers to be possibly overfitted on some training data by introducing more new training data. As a result, the classifiers trained by online-learning mechanism can produce good results robustly and consistently as demonstrated
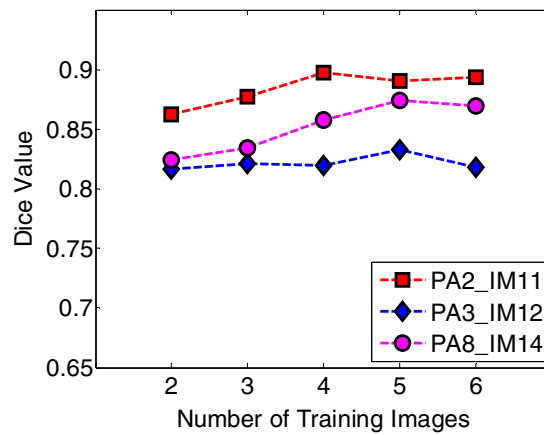
Figure 14. Segmentation accuracy of our proposed algorithm with respect to different numbers of available training images. Here, 'PA2_IM11' means image 11 of patient 2.


Figure 15. An example illustrating the improvement of the segmentation accuracy by increasing the number of training images (image 14 of patient 8). (a) Original image slice. (b)–(d) Segmentation results obtained by the classifiers trained with two, four and six training images, respectively. The red curves are the ground-truth prostate contours. The blue curves are the prostate contours of the automatic segmentation results.

above. In this experiment, we used the first three images with manual segmentation and the four latest treatment images with automatic segmentation results (produced by the proposed algorithm) to train our classifiers. When the number of available training images is less than 7, all the available images were used for training in order to avoid the lack of sample problem. The number of images with manual ground-truth segmentation (i.e. 3 in this experiment) used for initializing the training process is considerably less than the total number (about 45) of the treatment images in the whole radiotherapy period of one patient. With the online-learning mechanism, our location-adaptive classifiers can learn the most up-to-date patient-specific information and also reserve the information from the manual segmentation results from the first three images. Figure 16 shows an example of the segmentation results on a patient using our location-adaptive classifiers trained along the two coordinate directions. We applied our method to the serial images of 11 patients, respectively. The prostate segmentation for each patient started from the fourth image. The online-learning-based segmentation results on the images of the 11 patients are summarized by the whisker plot in figure 17(a) for Dice measures and figure 17(b) for ASD measures, respectively. Also, the centroid differences in lateral (x), anterior–posterior (y) and superior–inferior (z) directions are shown in figure 18. Our method achieved the mean Dice value 0.908 and the ASD (mean $\pm$ SD) of 1.40 $\pm$ 0.57 mm.
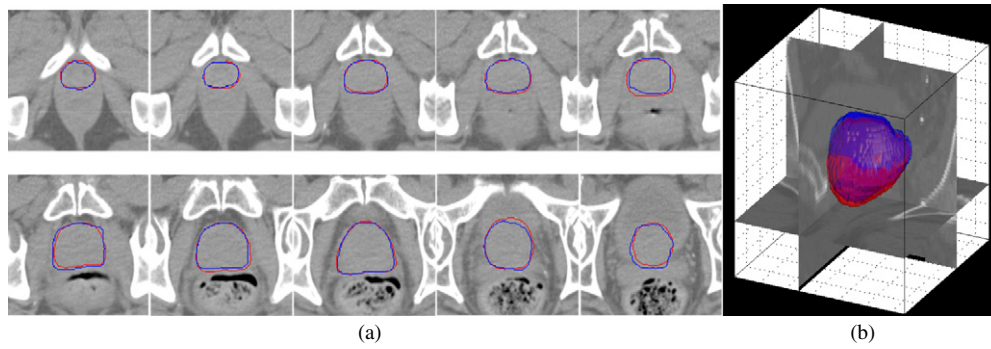
Figure 16. (a) Segmentation results on evenly spaced slices of image 11 of patient 5 (Dice 0.906). The red contours show the manually delineated boundaries of the prostate by a radiation oncologist, while the blue contours show the prostate boundaries determined by the proposed algorithm. (b) 3D comparison of two segmentation results.
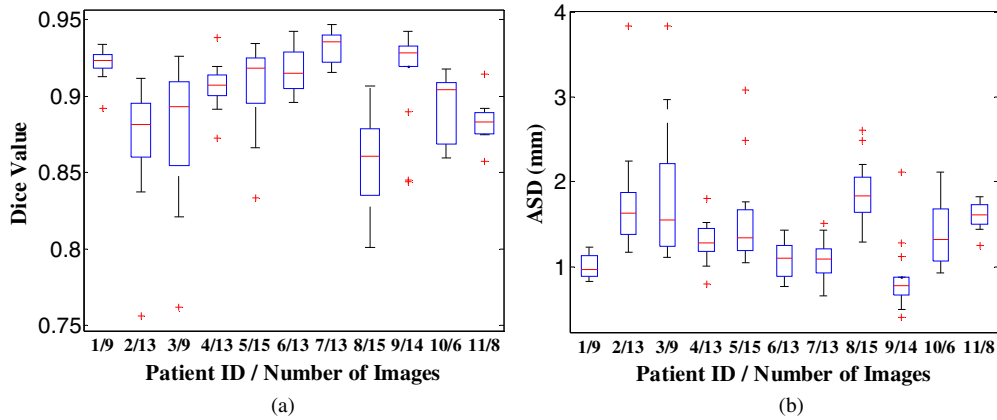

Figure 17. (a) Dice measure of segmentation results. (b) ASD of segmentation results.
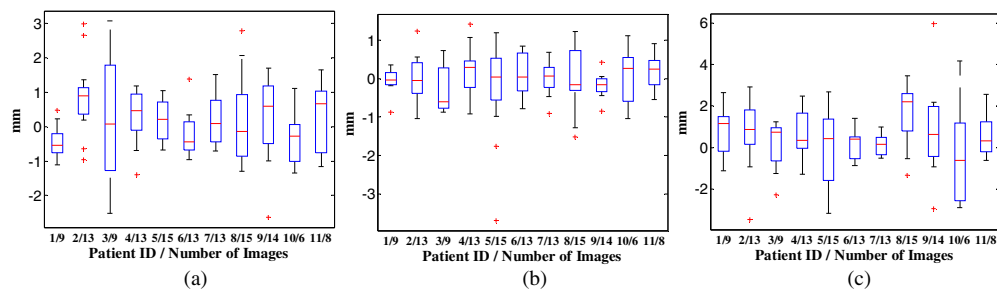
Figure 18. (a)–(c) The centroid differences in lateral (x), anterior–posterior (y) and superior–inferior (z) directions, respectively. All symbols are the same as those in figure 16.

Furthermore, we demonstrated the advantage of our online-retraining method by running a comparison experiment, in which the classifiers were trained by only the first three images, without further including the four latest treatment images (with automatic segmentations). In particular, 4 patients with 16–18 treatment images were selected. Figure 19 shows that the
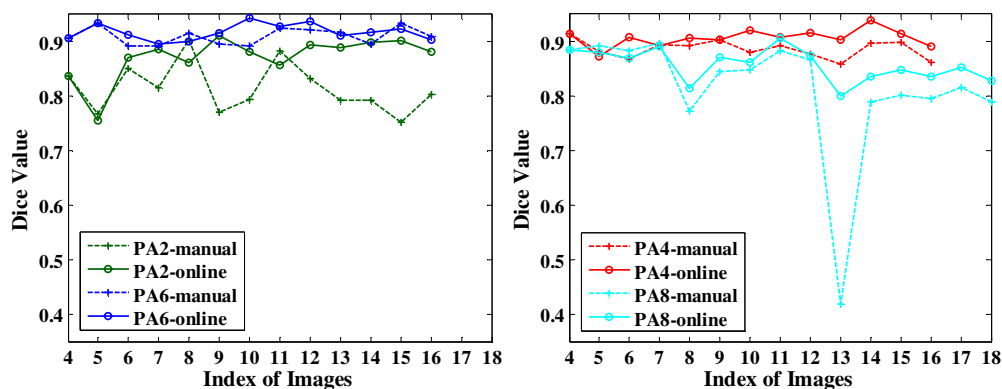
Figure 19. Demonstration of our online-learning mechanism. 'PA2-manual' means that the images of patient 2 are segmented by the classifiers trained with only the first three images (with manual segmentations). 'PA2-online' means that the images of patient 2 are segmented by our proposed online-learning method.

Dice ratio of each image obtained by the classifier trained with only the first three images and our proposed online-learning method, respectively. It can be observed that our proposed online-learning method can achieve much higher Dice ratios. For a typical example, i.e. the 13th image of patient 8, the low Dice ratio was produced by the classifiers trained with only the first three images, indicating that this image is significantly different from the first three images of the same patient. However, it can be satisfactorily segmented by our method which incorporates the latest information from the latest segmented treatment images.

3.7. Comparison with other methods

Three state-of-the-art prostate segmentation algorithms (Feng *et al* 2010, Chen *et al* 2011, Davis *et al* 2005) were compared with our algorithm in this section. The best segmentation performance of the methods reported in Feng *et al* (2010), Chen *et al* (2011) and Davis *et al* (2005) was used to compare with our algorithm. These three existing methods are deformable model based or registration based, while our algorithm is classification based. The method in Feng *et al* (2010) employs the patient-specific and online-learning strategy as our method. The method in Chen *et al* (2011) exploits population-based deformable model with anatomical constraints from pelvic bones. The method in Davis *et al* (2005) uses patient-specific registration. The comparisons of the above three methods with our method are given in table 1. Specifically, values of the five segmentation performance measures are provided in table 1, in which entries marked with N/A mean that the corresponding measures were not reported in the respective papers. Our mean Dice measure and mean ASD on all patients are 0.908 and 1.40 mm, respectively, which are better than the mean Dice measure of 0.893 and the mean ASD of 2.08 mm reported in Feng *et al* (2010). The paired *t*-test on the Dice measures shows that our method is significantly ($p < 0.002$) better than the method in Feng *et al* (2010). Our mean centroid differences are 0.18, -0.02 and 0.57 mm along the *x*, *y* and *z* directions, which are comparable to the best result of -0.26, 0.35 and 0.22 mm reported in Davis *et al* (2005), while our mean Dice measure 0.908 is significantly higher than the value 0.82 in Freedman *et al* (2005). Our median probability of detection (PD) and false alarm (FA) are 0.90 and 0.10, respectively, which are much better than 0.84 and 0.13 reported in Chen

Table 1. The comparison of different methods with respect to five performance measures.

Method	Method discription	Patients	Images	Mean Dice	Mean ASD (mm)	Mean CD $(x/y/z)$ (mm)	Median PD	Median FA
Feng *et al* (2010)	Online, patient specific, deformable model	24	306	0.893	2.08	N/A	N/A	N/A
Davis *et al* (2005)	Patient specific, registration-based method	3	40	0.820	N/A	−0.26/0.35/ 0.22	N/A	N/A
Chen *et al* (2011)	Anatomical constraints, deformable model	13	185	N/A	1.10	N/A	0.84	0.13
Our method	Online, learning-based classification	11	161	0.908	1.40	0.18/−0.02/ 0.57	0.90	0.10

N/A: the value was not reported in this paper.

et al (2011). However, the mean ASD 1.1 mm reported in Chen *et al* (2011) is better than the value 1.4 mm of our method. Furthermore, our method only uses three to seven images of one patient for training, while other methods use many more images from same patients or other patients for training through a leave-one-out strategy.

3.8. Online update of training data

One potential problem in our online-learning method is that if a low-quality automatic segmentation result is used to update the training data, the following retrained classifiers will be not reliable. The low-quality automatic segmentations are mainly owing to a significant variation of image appearance around the prostate in daily treatment images of some patients (Engels *et al* 2009). To illustrate this case, images of the first four days of patient 3 are shown in figure 20. It can be observed that the new treatment image (the fourth image) is significantly different from the first three images used for the training images; thus, the previous trained classifiers cannot effectively identify the prostate in the new treatment image (the fourth image).

To address this problem, our method is designed in a more flexible manner which ensures that only well-segmented images are used to update the training data, through an image dissimilarity indicator (IDI). Also, our method allows users to further check the automatic segmentation results offline, which is clinically feasible. Specifically, we define an IDI to determine whether the automatic segmentation can be used to update training data, as well as tell users whether the checking of the automatic segmentation result is needed. Here, the role of IDI is to estimate the quality of segmentation and assign a confidence degree to the automatic segmentation result of the current treatment image. When IDI has a high value, the confidence degree is low and in this case, manual checking is needed since a less accurate segmentation result is probably obtained. In contrast, if the IDI value is low, the confidence degree is high and in this case, manual checking is not needed. By using this IDI measure, we can reduce the time to manually check the automatic segmentation results and meanwhile eliminate the negative impact of including low accuracy automatic segmentation results to retrain the classifiers.

IDI is defined as a ratio of the mean dissimilarity between the new treatment image and all training images to the mean dissimilarity between each pair of training images, where dissimilarity is measured based on the features learned by our location-adaptive classifiers. When a new image comes in, it is first rigidly aligned with the planning image and further trimmed out with a 3D ROI (whose size and position are defined in the training process). Then,

Day 1   **Day 2**   **Day 3**   **Day 4**
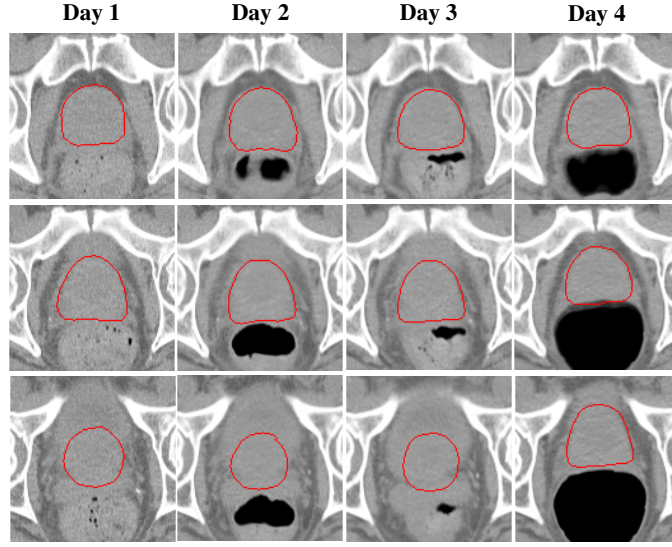
Figure 20. Illustration of large variation of position and image appearance of prostates in the first four images of patient 3 during the radiotherapy course. Each column represents slices of a 3D CT image scanned at a certain day. Each row represents the same slices of rigidly aligned images of four different treatment days. The red curves are ground truth.

the dissimilarity between image I_a and image I_b is computed within the ROI by the following steps.

(1) Select some evenly spaced slices at the same positions within the ROIs of both I_a and I_b, and further select the evenly spaced in-plane pixels from those selected slices.

(2) For each selected slice, its corresponding location-adaptive classifier is known. Next, features and weights used by the first sub-classifier of the corresponding location-adaptive classifier are obtained. Then, values of the above features, including image appearance features and image context features, are calculated for each sampled pixel in the current slice, respectively. (Note that, for the first sub-classifier, we can use only the *default* uniform classification probability map since the new classification probability map is not computed yet.)

(3) Finally, the square root of the weighted square sum of feature values at all selected pixels of selected slices is calculated as the dissimilarity between I_a and I_b.

Assume that the number of training images is P. The dissimilarities between the new image and each of the training images, $d_p^{\mathrm{new}}(p = 1, \ldots, P)$, as well as each pair of training images, $d_q^{\mathrm{train}}(q = 1, \ldots, Q, Q = C_P^2 = P \times (P - 1)/2)$, is calculated through steps (1)–(3), respectively. Finally, we can get the IDI through the following formulation:

$$\mathrm{IDI} = \frac{\mathrm{mean}\left(d_1^{\mathrm{new}}, \ldots, d_P^{\mathrm{new}}\right)}{\mathrm{mean}\left(d_1^{\mathrm{train}}, \ldots, d_Q^{\mathrm{train}}\right)},$$

where mean denotes the average operator.

In figure 21, the IDI value with respect to the Dice values of automatic segmentation results of the images for 11 patients is shown. From figure 21, it can be observed that all IDIs of low Dice value images (e.g. Dice value less than 0.80) have a larger value (e.g., larger than 1.5). It can also be observed that some images with a high Dice value also have a large IDI
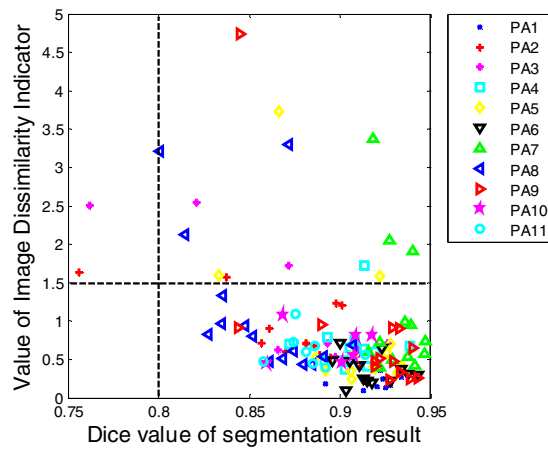
Figure 21. The relationship between the IDIs and the Dice values of automatic segmentation results for all images of 11 patients.

value (e.g., larger than the real value 1.5). In other words, our location-adaptive classifiers (each consisting of sequential *sub-classifiers*) can still produce high accuracy segmentation results even if there exists large dissimilarity between the new image and training images in some cases. In general, the most important point is that low accuracy automatic segmentation results will not be missed, which really need manual checking by using the proposed IDI measure, in order to keep high reliability of our online leaning classifiers. In our application on 11 patients, the total number of the IDI with value larger than the predefined threshold (e.g., 1.5) is actually small (i.e. 12.5% of the total number of the images used in this paper), indicating that little manual checking time is required in the whole radiotherapy based on the proposed IDI.

4. Discussion

In this section, important properties and possible alternative schemes of the proposed method are discussed. Our proposed algorithm is flexible which can incorporate other types of classifiers and features for possible improvement of prostate segmentation. First, since each location-adaptive classifier consists of a sequence of *sub-classifiers*, the choice of the *sub-classifiers* also influences the overall segmentation performance of the prostate. Although other learning algorithms, such as SVM and random forest, can also be used as sub-classifiers, boosting has many appealing properties (Friedman *et al* 2000): (1) it is a natural feature selection and fusion process; (2) it can deal with a large number of features from a considerable number of training data; (3) features do not need to be normalized; (4) it is less susceptible to the overfitting problem than most other learning algorithms and (5) it can be efficiently trained and used. Therefore, we use AdaBoost as the sub-classifier in this paper. Second, different types of features can be used in our algorithm. One reason to use the Haar and HOG features as the image appearance features in this paper is due to their computational efficiency. Another reason is that the HOG features provide complementary information to the Haar features (Dalal and Triggs 2005). On the other hand, the prostate region in the CT image has almost homogeneous intensity values except for some intermittent weak boundary. Therefore, texture descriptors seem not be effective in this application. As for the context

feature, different context sample pattern can be designed for integrating context information more effectively, as well as the features at each context sample point.

Currently, our method is implemented with Matlab2010 on an Intel Xeon 2.26 GHz CPU, and the typical computation time for the training stage, i.e. for training in total 32 location-adaptive classifiers (with 12 and 20 classifiers along two directions, respectively), was about 2 h. After training, the average time for segmenting a new image was about 3 min. However, our algorithm can be optimized and implemented using C/C++ in the future for significant improvement of segmentation speed. Thanks to the strategy of the location-adaptive classifiers in two coordinate directions, the program can be further improved by using parallel programming techniques to reduce the time of training and segmentation.

5. Conclusion

In this paper, we have presented a patient-specific, online-learning and location-adaptive classification method, based on image context information, to segment the prostate in the 3D CT scans during radiotherapy. Our method uses only the previous treatment images of the same patient as training samples to specifically learn the patient-specific information to enhance the segmentation accuracy. This is different from the conventional deformable-model- and image-registration-based methods that generally use the population information to guide the segmentation. Our trained classifiers can be updated online by the latest-acquired treatment images, thus better adapting to the prostate change of the current patient during the treatment. So, the location-adaptive classifiers make prediction more accurate. Experimental results on our data set show that our proposed method can consistently outperform the three state-of-the-art segmentation algorithms under comparison, using five different performance measures.

In summary, we have presented a patient-specific, online-learning and location-adaptive classification method, based on image context information, to segment the prostate in the 3D CT scans during radiotherapy. We applied the method to serial images of 11 patients. To evaluate the performance of the proposed method, we evaluated it using manually labeled prostates as the ground truth. The results show that the algorithm is able to segment the prostate efficiently and accurately, achieving mean Dice value 0.908 and mean ASD value 1.40.

Future works are summarized in the following aspects. First, shape priors or more effective constraints of the prostate shape can be integrated into our framework to get more accurate and applausive results, although the context features already contain some partial shape information implicitly. Second, the bladder and rectum can be segmented simultaneously with the prostate using a multi-class learning framework to provide more information of their correlation. Third, clinical feedbacks will be acquired from the clinicians to further improve the proposed method.

References

American Cancer Society 2006 *Cancer Facts and Figures 2006* (Atlanta, GA: American Cancer Society)

Belongie S, Malik J and Puzicha J 2002 Shape matching and object recognition using shape contexts *IEEE Trans. Pattern Anal. Mach.Intell.* **24** 509–22

Broadhurst R *et al* 2005 Histogram statistics of local model-relative image regions *Int. Workshop Deep Structure, Singularities, and Comput. Vision (DSSCV)* pp 72–83

Chan T F and Vese L A 2001 Active contours without edges *IEEE Trans. Image Process.* **10** 266–77

Chen T *et al* 2009 3D meshless prostate segmentation and registration in image guided radiotherapy *Int. Conf. Medical Image Computing and Computer-Assisted Intervention* pp 43–50

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

Dalal N and Triggs B 2005 Histograms of oriented gradients for human detection *Proc. 2005 IEEE Computer Society Conf. Comput. Vision and Pattern Recognition (CVPR '05)* vol 1-01 pp 886–93

Davis B C *et al* 2005 Automatic segmentation of intra-treatment CT images for adaptive radiation therapy of the prostate *Med. Image Comput. Comput. Assist. Interv.* **8** 442–50

Dice L R 1945 Measures of the amount of ecologic association between species *Ecology* **26** 297–302

Engels B *et al* 2009 Assessment of rectal distention in radiotherapy of prostate cancer using daily megavoltage CT image guidance *Radiother. Oncol.* **90** 377–81

Feng Q *et al* 2010 Segmenting CT prostate images using population and patient-specific statistics for radiotherapy *Med. Phys.* **37** 4121–32

Fischer B and Modersitzki J 2003 FLIRT: a flexible image registration toolbox *Biomedical Image Registration (Lecture Notes in Computer Science)* (Berlin: Springer) pp 261–70

Freedman D *et al* 2005 Model-based segmentation of medical imagery by matching distributions *IEEE Trans. Med. Imaging* **24** 281–92

Freund Y and Schapire R E 1995 A decision-theoretic generalization of on-line learning and an application to boosting *Proc. 2nd European Conf. Comput. Learning Theory* pp 23–37

Friedman J, Hastie T and Tibshirani R 2000 Additive logistic regression: a statistical view of boosting *Ann. Stat.* **28** 337–74

Gao Y *et al* 2010 A coupled global registration and segmentation framework with application to magnetic resonance prostate imagery *IEEE Trans. Med. Imaging* **29** 1781–94

Geman S and Geman D 1984 Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images *IEEE Trans. Pattern Anal. Mach.Intell.* **6** 721–41

Ghosh P and Mitchell M 2006 Segmentation of medical images using a genetic algorithm *Proc. 8th Annu. Conf. on Genetic and Evolutionary Comput. (Seattle, WA, USA)* pp 1171–8

Hupse R and Karssemeijer N 2009 Use of normal tissue context in computer-aided detection of masses in mammograms *IEEE Trans. Med. Imaging* **28** 2033–41

Jin Y and Geman S 2006 Context and hierarchy in a probabilistic image model *Proc. 2006 IEEE Computer Society Conf. Comput. Vision and Pattern Recognition* vol 2 pp 2145–52

Lafferty J D, McCallum A and Pereira F C N 2001 Conditional random fields: probabilistic models for segmenting and labeling sequence data *Proc. 18th Int. Conf. Machine Learning* pp 282–9

Langen K M and Jones D T 2001 Organ motion and its management *Int. J. Radiat. Oncol. Biol. Phys.* **50** 265–78

Langerak T R *et al* 2010 Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE) *IEEE Trans. Med. Imaging* **29** 2000–8

Lee T *et al* 2001 Prostate shape and symptom score in benign prostatic hyperplasia *Yonsei Med. J.* **42** 532–8

Li W *et al* 2011 Learning image context for segmentation of prostate in CT-guided radiotherapy *Proc. 14th Int. Conf. Medical Image Comput. and Computer-Assisted Intervention—Volume Part III (Toronto, Canada)* pp 570–8

Loog M and Ginneken B 2006 Segmentation of the posterior ribs in chest radiographs using iterated contextual pixel classification *IEEE Trans. Med. Imaging* **25** 602–11

Mazonakis M *et al* 2001 Image segmentation in treatment planning for prostate cancer using the region growing technique *Brit. J. Radiol.* **74** 243–9

Oliva A and Torralba A 2007 The role of context in object recognition *Trends Cogn. Sci.* **11** 520–7

Rousson M *et al* 2005 Constrained surface evolutions for prostate and bladder segmentation in CT images *Comput. Vision Biomed. Image Appl.* **3765** 251–60

Shen D, Zhan y and Davatzikos C 2003 Segmentation of prostate boundaries from ultrasound images using statistical shape model *IEEE Trans. Med. Imaging* **22** 539–51

Song Q *et al* 2009 Optimal graph search segmentation using arc-weighted graph for simultaneous surface detection of bladder and prostate *Med. Image Comput. Comput. Assist. Interv.* **12** 827–35

Tu Z and Bai X 2010 Auto-context and its application to high-level vision tasks and 3D brain image segmentation *IEEE Trans. Pattern Anal. Mach. Intell.* **32** 1744–57

Viola P and Jones M J 2004 Robust real-time face detection *Int. J. Comput. Vision* **57** 137–54

Wang H *et al* 2005 Implementation and validation of a three-dimensional deformable registration algorithm for targeted prostate cancer radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **61** 725–35

Zhan Y and Shen D 2006 Deformable segmentation of 3-D ultrasound prostate images using statistical texture matching method *IEEE Trans. Med. Imag.* **25** 256–72

Zhan Y *et al* 2007 Targeted prostate biopsy using statistical image analysis *IEEE Trans. Med. Imaging* **26** 799–88