Supervised and unsupervised methods for prostate cancer segmentation with multispectral MRI

Sedat Ozer[a)]
Department of Electrical and Computer Engineering, Medical Imaging Research Center (MIRC), Illinois Institute of Technology, Chicago, Illinois 60616

Deanna L. Langer
Joint Department of Medical Imaging, Princess Margaret Hospital, University Health Network and Mount Sinai Hospital, Toronto, Ontario M5G 2M9, Canada and Institute of Medical Science, University of Toronto, King's College, Toronto, Ontario M5S 1A8, Canada

Xin Liu
Department of Electrical and Computer Engineering, Medical Imaging Research Center (MIRC), Illinois Institute of Technology, Chicago, Illinois 60616

Masoom A. Haider
Joint Department of Medical Imaging, Princess Margaret Hospital, University Health Network and Mount Sinai Hospital, Toronto, Ontario M5G 2M9, Canada and Institute of Medical Science, University of Toronto, King's College, Toronto, Ontario M5S 1A8, Canada

Theodorus H. van der Kwast and Andrew J. Evans
Department of Pathology and Laboratory Medicine, Toronto General Hospital, Toronto, Ontario M5G 2C4, Canada

Yongyi Yang, Miles N. Wernick, and Imam S. Yetik
Department of Electrical and Computer Engineering, Medical Imaging Research Center (MIRC), Illinois Institute of Technology, Chicago, Illinois 60616

(Received 15 May 2009; revised 7 December 2009; accepted for publication 1 February 2010; published 30 March 2010)

Purpose: Magnetic resonance imaging (MRI) has been proposed as a promising alternative to transrectal ultrasound for the detection and localization of prostate cancer and fusing the information from multispectral MR images is currently an active research area. In this study, the goal is to develop automated methods that combine the pharmacokinetic parameters derived from dynamic contrast enhanced (DCE) MRI with quantitative *T*2 MRI and diffusion weighted imaging (DWI) in contrast to most of the studies which were performed with human readers. The main advantages of the automated methods are that the observer variability is removed and easily reproducible results can be efficiently obtained when the methods are applied to a test data. The goal is also to compare the performance of automated supervised and unsupervised methods for prostate cancer localization with multispectral MRI.

Methods: The authors use multispectral MRI data from 20 patients with biopsy-confirmed prostate cancer patients, and the image set consists of parameters derived from *T*2, DWI, and DCE-MRI. The authors utilize large margin classifiers for prostate cancer segmentation and compare them to an unsupervised method the authors have previously developed. The authors also develop thresholding schemes to tune support vector machines (SVMs) and their probabilistic counterparts, relevance vector machines (RVMs), for an improved performance with respect to a selected criterion. Moreover, the authors apply a thresholding method to make the unsupervised fuzzy Markov random fields method fully automatic.

Results: The authors have developed a supervised machine learning method that performs better than the previously developed unsupervised method and, additionally, have found that there is no significant difference between the SVM and RVM segmentation results. The results also show that the proposed methods for threshold selection can be used to tune the automated segmentation methods to optimize results for certain criteria such as accuracy or sensitivity. The test results of the automated algorithms indicate that using multispectral MRI improves prostate cancer segmentation performance when compared to single MR images, a result similar to the human reader studies that were performed before.

Conclusions: The automated methods presented here can help diagnose and detect prostate cancer, and improve segmentation results. For that purpose, multispectral MRI provides better information about cancer and normal regions in the prostate when compared to methods that use single MRI

techniques; thus, the different MRI measurements provide complementary information in the automated methods. Moreover, the use of supervised algorithms in such automated methods remain a good alternative to the use of unsupervised algorithms. © *2010 American Association of Physicists in Medicine*. [DOI: 10.1118/1.3359459]

Key words: relevance vector machine, support vector machine, Markov random fields, fuzzy segmentation, prostate cancer localization, threshold selection, multispectral MRI

I. INTRODUCTION

Prostate cancer is one of the most prevalent cancer types and one of the leading causes of cancer death for men in the United States.[1] The detection of prostate cancer at an early stage is crucial to increase the probability of successful treatment. Traditional prostate cancer detection uses digital rectal examinations and serum prostate specific antigen levels.[2] The final clinical diagnosis of prostate cancer is based on histological tissue analysis, which is currently performed using needle biopsy to obtain multiple tissue samples from the prostate. Imaging techniques to segment the most significant cancer can be used to target biopsy, reducing sampling errors and the number of tissue samples required at biopsy. Accurate segmentation of prostate cancer could also be used to guide radiotherapy, surgery, or emerging focal therapies such as cryoablation, focusing ultrasound, or laser ablation.[3] In this paper, our goal is to develop automated methods to segment the cancerous tissue using multispectral MRI.

To reduce the number of biopsy samples taken from patients, the use of transrectal ultrasound (TRUS) to localize the disease has been proposed[4] and is currently the most widely used clinical imaging technique. However, since the accuracy of TRUS is limited, magnetic resonance imaging (MRI) has been proposed as an alternative to TRUS, due to its superior soft-tissue imaging capabilities.[5,6] Several studies have demonstrated that MRI can provide better resolution to assist in detecting smaller volumes of prostate cancer with a higher accuracy than TRUS,[5,7] and is a promising method for prostate cancer localization.

Cancer regions exhibit low signal values on $T2$, when compared to surrounding normal peripheral zone (PZ) tissue. However, common benign processes such as prostatic inflammation, postbiopsy hemorrhage, and fibrosis also exhibit low $T2$ signal. Therefore, although single-type MR images, such as $T2$-weighted MRI, can be used to localize prostate cancer, the accuracy of such methods is not sufficient to consistently target regions of tumor. Thus, methods such as diffusion weighted imaging (DWI), quantitative $T2$ MRI, and dynamic contrast enhanced (DCE) MRI have been suggested, often in combination with anatomic $T2$-weighted, to improve localization accuracy.[8]

Parameters derived using all of these techniques have been demonstrated to be altered in prostate cancer compared to normal peripheral zone tissue.[1,6–9] Therefore, recent studies have proposed using a multispectral approach for localizing prostate cancer[10–13] since each MRI technique adds complementary information about tumor and normal tissue physiology. However, most of these studies have relied on human observers to detect and segment the cancer.[10–12] An automated method to localize and segment prostate tumors is very useful because it decreases subjective evaluation, hence interobserver variability. It is also more efficient since individual reading of each parameter or data set is no longer necessary.

Although several image segmentation algorithms have been used to localize cancer regions in different sites in humans,[14,15] studies for automated segmentation of human prostate cancer with MRI have been very limited.[13] In Ref. 13, Chan *et al.* used texture features derived from $T2$-weighted MRI, quantitative $T2$ maps, and maps of apparent diffusion coefficient (ADC), and used generic support vector machines (SVMs) and Fisher linear discriminate to compare the performance of both methods. Our study differs from Ref. 13 in several aspects: First, we use DCE–MRI-derived parametric maps in addition to ADC and quantitative $T2$ maps for automated segmentation. Second, as opposed to the use of generic SVM methods, we use competing methods such as relevance vector machines (RVMs) and an unsupervised method based on fuzzy Markov random fields (MRF). A comparison of supervised and unsupervised methods has not been performed before and is an important information for the researchers working in the area. Finally, we propose to use thresholding schemes that tune the segmentation methods to provide an increased performance with respect to a certain criterion (accuracy, sensitivity, etc.) depending on the desired application instead of applying generic thresholding as in Ref. 13.

II. METHODS

II.A. Supervised learning methods

Supervised learning methods require labels along with the training data. The SVM and RVM methods are such popular methods that we use in this paper.[16]

II.A.1. Support vector machine

SVM is widely used in many applications because of its superior generalization capability.[18] To separate the given two classes for segmentation, SVM finds a hyperplane in the higher dimensional feature space based on the maximum margin idea. The SVM algorithm yields the optimal coefficients for the linear expansion of the decision function in terms of the training data. Besides its accuracy, SVM also provides sparseness reducing the computation time on test data.

For a given test vector x, SVM uses the following formula to classify the data:[18,19,23]

$$f(x) = \sum_{i=1}^{k} \alpha_i y_i K(x,x_i) + b, \tag{1}$$

$$y = \mathrm{sgn}(f(x)), \tag{2}$$

where α_i is nonzero Lagrange multiplier for each support vector x_i, k is the support vector number, $y_i \epsilon \{-1,+1\}$ the class label, and b is the bias value for the hyperplane.

The inner product in the high dimensional feature space can be found by applying a kernel function directly onto the input vectors without the need of finding any transformation function $\phi(.)$ to make the data linearly separable in that higher dimensional space as

$$K(x,x_i) = \langle \phi(x) \,.\, \phi(x_i) \rangle, \tag{3}$$

where $K(.)$ is defined as the kernel function. Thus for any nonlinear data set, α_i values in Eq. (1) can be found by maximizing the dual optimization problem $Q(\alpha)$

$$Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i,x_j), \tag{4}$$

subject to $\Sigma_{i=1}^{n} \alpha_i y_i = 0$ and $C \geq \alpha_i \geq 0$.

The x_i input (training) vectors with the nonzero Lagrange multiplier α_i are called support vectors (SVs). Polynomial kernel function is an example to the kernel functions

$$K(x,x_i) = (\langle x,x_i \rangle + 1)^d, \tag{5}$$

where d is the polynomial degree, which we select as one resulting in a linear SVM in this study.

II.A.2. Relevance vector machine

RVM is a probabilistic method that uses the same functional form of SVM[20,21] for decision and used in many applications including target tracking and classification.[22] In this sparse Bayesian probabilistic model, for a given test vector x, the class label y can be found as

$$f(x,w) = y = \sum_{i=1}^{n} w_i K(x,x_i) + w_0, \tag{6}$$

where the w_i is the ith weight for the ith training input, $w = [w_1, w_2, \ldots, w_n]^T$, n is the training sample number, and $K(x,x_i)$ is the kernel function.

In this Bayesian approach, each w has a Gaussian distribution parameterized by a hyperparameter α, whose most probable value is calculated iteratively through the RVM algorithm by using expectation maximization algorithm.[20] Notice that this hyperparameter α differs from the one defined for SVM.

$$p(w|\alpha_1, \ldots, \alpha_n) = (2\pi)^{-n/2} \prod_{i=1}^{n} \alpha_i^{1/2} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \alpha_i w_i^2 \right), \tag{7}$$

where α_i is the hyperparameter for the corresponding w_i value. Each hyperparameter can be considered as the controlling parameter for the inverse variance of each weight. The training vectors that have noninfinite α value are called relevance vectors (RVs). Therefore, in general, RVs do not lay on the margin as opposed to SVs for SVM.

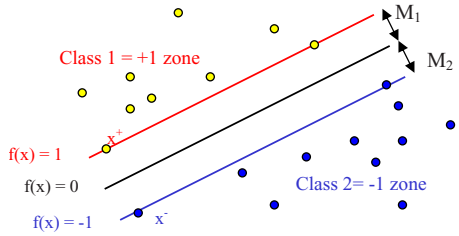

FIG. 1. Maximum margin width.

RVM assumes that the targets are noisy observations of the actual values as shown below

$$t = y + \varepsilon, \tag{8}$$

where t is the noisy observed output, ε is the noise value, and y is the actual target output. RVM can learn the noise variance along the training procedure as well. While SVM assumes -1 or $+1$ values for class labels, RVM classification assumes that the label values for t to be 0 or 1 for each class. Therefore the RVM output value is mapped to a range between 0 and 1 with the following equation:

$$y' = \frac{1}{1 + \exp(-y)}. \tag{9}$$

II.A.3. Threshold selection for SVM and RVM

In this section, we propose to select a threshold value for increased segmentation performance with respect to a certain criterion depending on the desired application using both SVM and RVM algorithms. Classical SVM selects a threshold such that the distances M_1 and M_2 are equal assuming test data are equally likely to come from two classes. This concept is illustrated in Fig. 1. However, this is not true in cancer segmentation and the optimum bias provided by classical methods, which is the midpoint, is no longer optimum. In this study, we show that selecting a different threshold value can improve the performance of SVM and RVM with respect to a desired performance criterion.

The goal of selecting a threshold is not to create a "better" classifier in the universal sense, since this is not possible considering the fact that there is no single goodness criterion. Our goal is to develop methods that can tune the threshold value depending on the application. Then, a threshold can be selected such that the accuracy is optimized, or another threshold can be selected such that a minimum required specificity is achieved while maximizing sensitivity. The user will be able to determine what is more important for a particular application (accuracy, sensitivity, specificity, etc.) and use the method that yields a tuned threshold targeted at that particular performance criterion. Next, we describe several

of these performance criteria that can be useful and explain how a threshold can be selected that maximizes each of these criterion.

II.A.3.a. Accuracy maximization. In this approach, we select a threshold that maximizes the accuracy, which is defined as

$$\text{accuracy} = \frac{\text{DTN} + \text{DTP}}{\text{TN} + \text{TP}}, \tag{10}$$

where DTN is the number of true detected negative pixels, DTP is the number of true detected positive pixels, TP is the number of true positive pixels, and TN is the number of true negative pixels.

II.A.3.b. Weighted accuracy maximization. Another approach is maximizing the weighted sum of the DTN and DTP value such that

$$\arg\max_T(\pi\text{DTN} + (1-\pi)\text{DTP}), \tag{11}$$

where π is the weighting factor. This approach can be useful where the number of true negatives is excessively higher than the number of true positives or vice versa in order to avoid bias toward one class. However for this approach, finding the optimal π value is an open problem. In this study we choose π such that $\pi=\text{TP}/(\text{TN}+\text{TP})$. Then the above maximization problem is also equivalent to maximizing

$$\arg\max_T(P(f=1|y=1)P(y=-1) + P(f=-1|y=-1)P(y=1)). \tag{12}$$

Proof: From Bayes theorem we have

$$P(f=1|y=1) = \frac{P(f=1,y=1)}{P(y=1)}, \tag{13}$$

which can be written as

$$\frac{P(f=1,y=1)}{P(y=1)} = \frac{\hat{P}(f=1,y=1)}{\hat{P}(y=1)}, \tag{14}$$

where $\hat{P}$ is the estimate of the actual probability value P. Then we can rewrite the following equation as:

$$\frac{\hat{P}(f=1,y=1)}{\hat{P}(y=1)} = \frac{\text{DTP}/(\text{TP}+\text{TN})}{\text{TP}/(\text{TP}+\text{TN})} = \frac{\text{DTP}}{\text{TP}}, \tag{15}$$

which is also known as the *Sens* sensitivity value. Similarly using

$$P(f=-1|y=-1) = \frac{\hat{P}(f=-1,y=-1)}{\hat{P}(y=-1)}, \tag{16}$$

we have

$$\frac{\hat{P}(f=-1,y=-1)}{\hat{P}(y=-1)} = \frac{\text{DTN}/(\text{TP}+\text{TN})}{\text{TN}/(\text{TP}+\text{TN})} = \frac{\text{DTN}}{\text{TN}}, \tag{17}$$

which is also known as the *Spec* specificity value. Thus we have

$$\arg\max_T\left(\frac{\text{DTN}}{\text{TN}} + \frac{\text{DTP}}{\text{TP}}\right) = \arg\max_T(\text{Spec}+\text{Sens}). \tag{18}$$

$$\equiv \arg\max_T(\hat{P}(f=1|y=1)\hat{P}(y=-1) + \hat{P}(f=-1|y=-1)\hat{P}(y=1)). \tag{19}$$

II.A.3.c. Maximizing sensitivity subject to a required minimum specificity. A third approach for thresholding is fixing the specificity at a predefined value and then finding the threshold value which yields the predefined specificity value.

In this study, we have used first order polynomial kernel for both vector machines resulting in a linear kernel based on preliminary studies. In our preliminary tests, we have applied alternative kernels as well, and found that the linear kernel performs similar to that of other kernel results; these results are not shown here because of the limited space. Moreover, using a linear kernel simplifies the formulation of the process as well. Once the process can be understood, the same framework can be applied on any other kernel machine as well. For a linear kernel, Eqs. (1) and (6) can be rearranged as

$$f(x,w) = \sum_{i=1}^{n} w_i \sum_{j=1}^{m} (x^j x_i^j) + \sum_{i=0}^{n} w_i, \tag{20}$$

where x^j is the jth element of given test vector x, and m is the feature number of each sample. Here, for the SVM case, $w_0=b$ and $w_i=\alpha_i y_i$ for $I=1,2,\ldots,n$.

Equation (20) can then be written in the linear form such that for a given test vector x

$$f(x,w) = \sum_{j=1}^{m} h_j x^j + h_0. \tag{21}$$

Then, we can find the linear coefficients such that

$$h_0 = \sum_{i=0}^{n} w_i, \tag{22}$$

$$h_j = \sum_{i=1}^{n} w_i x_i^j \quad j=1,2,\ldots,m. \tag{23}$$

The specificity Spec for the training data of given M number of patients can be written as follows:

$$\text{Spec} = \frac{\sum_{k=1}^{M}\left[\sum_{l=1,x\in S}\left[\left(\left(\sum_{j=1}^{m} h_j x_l^j + h_0\right) < T\right) \cap (\text{TN})_{l,k}\right]\right]}{\sum_{k=1}^{M}\sum_{l=1,x\in S}(\text{TN})_{l,k}}, \tag{24}$$

where the true negative $(\text{TN})_{l,k}$ is the ground truth mask for the lth pixel of kth patient provided by the experts, S is the region of interest, i.e., the peripheral zone, and T is the threshold value.

Consequently, a threshold value that results in a predefined specificity value can be selected once the optimal

coefficients are learnt by RVM and SVM, respectively. We choose to select a threshold value resulting in a 0.8 Spec value in this study. This value of specificity is justified by previous clinical experience and experiments, and can be altered when there is a need depending on the clinical application.

For all these three methods, the threshold selection is purely based on the training data, and once it is fixed by maximizing the accuracy of the training data or by fixing the specificity to a certain value, it is not varied for individual test patients.

The proposed algorithms can be summarized as follows:

(1) Calculate the mean value of TN and TP regions from each patient.
(2) Train the vector machine to find the optimal w and b values for i.i.d. case.
(3) Use all training pixels for cross validation and then adjust the threshold T value according to one of the proposed criteria.
(4) In order to use sign function for decision, calculate the new (final) bias value b' to set the threshold value to 0, where $b'=b-T$.
(5) Apply the final w and b' values to the test subject to obtain segmentation results.

II.B. Unsupervised segmentation methods

Unsupervised methods play an important role in many applications where the class labels are unavailable. In this study we choose fuzzy Markov random fields to represent unsupervised algorithms since the fuzzy Markov random fields method was compared to several other unsupervised methods in Ref. 17 and as it was found to yield higher localization performance than alternative unsupervised methods for prostate cancer segmentation with multispectral MRI.

The Markov random field models the objects in a probabilistic way. In MRF model, we consider two random fields, X and Y for the region of interest S, where the measured data to be segmented is an instance of X denoted by x and the corresponding class label y is an instance of a hidden field Y. For fuzzy MRF, each pixel in S is said to belong simultaneously to more than one of K classes with membership y_k, yielding a vector of memberships $(y_1,y_2,\ldots,y_K)\in[0,1]$, with $\Sigma_{k=1}^{K}y_k=1$. Thus, with $y_k=1$, the pixel purely belongs to the class k.[24]

The problem of estimating the class label y can be accomplished by choosing a maximum *a posteriori* (MAP) estimation of y. For two classes classification problem, means and variances are denoted by μ_1, μ_2, σ_1^2, and σ_2^2, respectively, y_1,y_2 denote the degree of membership, β is the weighting parameter, $t\epsilon\{0,1\}$ is the class label, and P is the clique. The likelihood function to be maximized can be expressed as

$$L(y_1,y_2)=\sum_i\left\{\log\left[\frac{y_{1i}\exp\frac{(x_i-\mu_1)^2}{2\sigma_1^2}}{\sqrt{2\pi\sigma_1^2}}+\frac{y_{2i}\exp\frac{(x_i-\mu_2)^2}{2\sigma_2^2}}{\sqrt{2\pi\sigma_2^2}}\right]-\beta\sum_{j\in P(i)}|y_{1i}-y_{1j}|^2+|y_{2i}-y_{2j}|^2\right\} \quad (25)$$

for a Gaussian distribution and Gibbs prior.[25]

From Eq. (25), we can see that parameters μ_1, μ_2, σ_1^2, and σ_2^2 play a crucial role in the estimation of the class label. The performance of the classification algorithm would suffer significantly when the MRF parameters cannot be accurately estimated. The main idea of our unsupervised method proposed in Ref. 17 is that instead of estimating the parameters used in the models, including the means and variances of the Gaussian distributions in $p(x)$ as a preprocessing step, we estimate these parameters simultaneously while performing classification. In this method, expressions that are functions of degree of class membership y_k replace the means and variances

$$\hat{\mu}_k=\frac{1}{\sum_{i=1}^{N}y_{ki}}\sum_{i=1}^{N}y_{ki}x_i,\quad \hat{\sigma}_k^2=\frac{1}{\sum_{i=1}^{N}y_{ki}}\sum_{i=1}^{N}y_{ki}(x_i-\mu_k)^2, \quad (26)$$

where $k=1,2$, $y_{1i},y_{2i}\in[0,1]$, $y_{1i}+y_{2i}=1$, N is the total number of pixels in the region of interest, and x_i is the intensity of the ith pixel.

Since, y_{1i} and y_{2i} are dependent, we substitute $y_{2i}=1-y_{1i}$ into the likelihood function. Then, the segmentation result consisting of the class parameters for each pixel can be performed by finding the y_{1i}'s that maximize the likelihood function

$$\hat{y}_1=\arg\max_{y_1}\sum_i\left\{\log\left[y_{1i}\exp\frac{(x_i-\mu_1)^2}{2\sigma_1^2}+(1-y_{1i})\exp\frac{(x_i-\mu_2)^2}{2\sigma_2^2}\right]-\beta\sum_{j\in P(i)}|y_{1i}-y_{1j}|^2\right\}, \quad (27)$$

where μ_1, μ_2, σ_1^2, and σ_2^2 are functions of the y_1's defined by Eq. (26).

This method can easily be extended to multidimensional feature vectors, such as multispectral MR data in our problem. The only required modification is to extend Eq. (27) to its vector and matrix counterparts. Then, these expressions will be used to obtain the final cost function which will be a function of y_1's only.

Since the labels are fuzzy, we need to assign exact discrete labels with a hard decision which can be done by using the thresholding method proposed by Otsu *et al.* in Ref. 26. Automatic selection of the threshold allows our unsupervised method to complete the segmentation task with no human intervention. The Otsu threshold is obtained by minimizing the intraclass variance, which is defined as a weighted sum of variances of the two hard classes.

$$\sigma^2_{\text{intra}}(T) = q_1(T)\sigma_1^2(T) + q_2(T)\sigma_2^2(T), \quad (28)$$

where $q_1(T)$ and $q_2(T)$ are the probabilities of the two classes defined by threshold T and σ_1^2 and σ_2^2 are variances of these two classes. We can calculate $\sigma^2_{\text{intra}}(T)$ iteratively as we test all possible threshold and choose one that minimizes the intraclass variance.

II.C. Multispectral MRI data: Description and protocol

The multispectral prostate MRI data we use in this study consist of three different types of images and is a combination of the morphological information and functional parameters derived from the different types of MR images: *T*2-weighted MRI, quantitative *T*2, DWI, and DCE-MRI obtained with a *T*1-weighted series. The data set is obtained from 20 patients with biopsy-confirmed prostate cancers. The study was approved by the patients' institutional research ethics board and all patients gave informed consent. The prostate has several anatomical zones such as the peripheral and transition zones and as the majority of prostate cancers occur in the peripheral zone;[27] therefore we consider only the peripheral zone in this study.

Quantitative *T*2 maps are calculated using an exponential decay fit to the multiecho FSE data sets (a series of echo time measurements). This fitting process removes the variations in signal intensity as a function of proximity to the endorectal coil seen in *T*2*w*. ADC maps are calculated from the DWI acquisition, and can assist in differentiating between *T*2 shine through effects or artifacts and real ischemic lesions. The DCE-MRI data are obtained by injecting contrast agent into the patient and acquiring a series of *T*1-weighted images. One purpose of DCE-MRI is to characterize tissue regions based on certain processes such as blood flow, vascular characteristics, or tissue integrity, as these values may differ in malignant tissue when compared to normal tissue.

From DCE-MRI, time concentration curves can be calculated using a *T*1-weighted series, and pharmacokinetic parameters can be obtained based on compartmental models. Maps of k_{ep} (contrast agent wash-in rate), k_{el} (contrast agent wash-out rate), and A (a fitting parameter incorporating contrast agent dose and other tissue properties) were generated based on Brix model to fit the DCE-MRI data.[9] The signal model with kinetic parameters A, k_{ep}, and k_{el} is $s(t)/s(0) = 1 + Ak_{\text{ep}}(e^{-k_{\text{ep}}t} - e^{-k_{\text{el}}t}/k_{\text{el}} - k_{\text{ep}})$, where $s(t)$ denotes the DCE-MRI time series and t denotes the time.[8] There are alternative models to quantitatively analyze the DCE-MRI data. The "optimal" model for prostate cancer localization is a subject of debate, and a comparison of alternative compartmental models is outside the scope of this paper; and certainly, alternatives that use arterial input function and *T*1 correction has the potential to improve localization performance. Besides, the methods developed in this paper are easily applicable to other models as well. We have preferred the Brix model since it does not require the arterial input function, which can be problematic to measure. A full MRI and parametric map data set is shown for one patient in Figs. 2(a)–2(f). Figures 2(h) and 2(i) show the ground truth for tumor and normal regions on the *T*2*w* image.

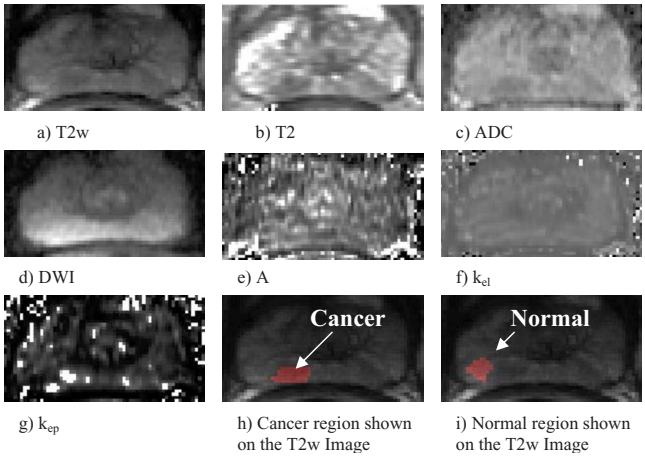

FIG. 2. Multispectral MRI data set for a patient. (a)–(g) are the feature images, (h) shows the cancer regions on *T*2*w* image, and (i) shows the normal region. Parts (h) and (i) are obtained based on pathology.

Prior to prostatectomy, each patient underwent endorectal MRI on a 1.5 T MRI system (Echospeed or Excite HD; GE Healthcare, Milwaukee, WI) using a four-channel phased-array surface coil coupled to an endorectal coil (MEDRAD, Warrendale, PA). In addition to anatomic *T*2-weighted MRI, DWI, quantitative *T*2, and DCE-MRI data sets were acquired. All data sets were obtained at an axial-oblique orientation, perpendicular to the rectal wall/prostate interface. Slices were 3 mm thick, were at the same location for all data sets, and had no intersection gap.

*Fast spin-echo T*2-weighted MR images obtained with no phase wrap. *DWI* images obtained when $b=0$, 600 s/mm^2. Q*uantitative T*2 mapping—Multiecho FSE values are obtained by using ten echo times (9.0–90.0 ms, in 9 ms increments; phase encode=left-right). *DCE-MRI*—Multislice fast spoiled gradient echo values are obtained with the temporal resolution=10 s, 50 phases. For DCE-MRI, 20 ml contrast agent was administered [gadopentate dimeglumine (Magnevist; Bayer Schering Pharma, Berlin, Germany)] at a rate of 4 ml/s, followed by a 20 ml saline flush using a power injector (MEDRAD Spectris MR injection system). Two phases were acquired prior to injection and used for baseline. Other acquisition parameters are listed on Table I for each imaging technique.

After prostatectomy, each patient's prostate was fixed in formalin for 24 h, and hematoxylin and eosin stained whole mount sections generated in 3 mm increments. Sections were assessed by a pathologist and regions of tumor were outlined on the sections as pathologic ground truth. These regions were then transferred to MRI by an experienced radiologist. The radiologist viewed the histological slide, the *in vivo* MR image, as well as the *ex vivo* MR image of the pathological slide. In this way (with the addition of *ex vivo* MRI of the histological slide), the regions from histological slide can more accurately be transferred into the *in vivo* MR image, when compared to viewing only the *in vivo* MR and the

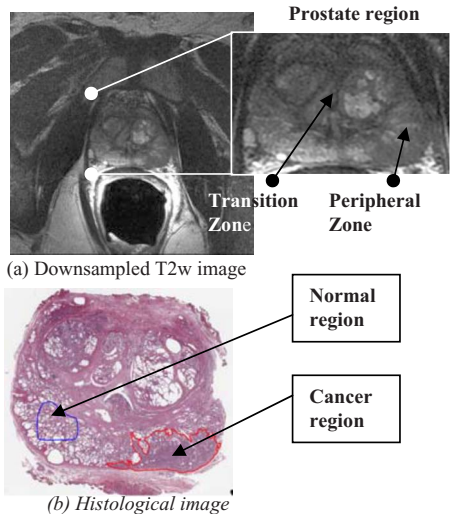

FIG. 3. (a) *T2w* image; (b) corresponding whole mount section.

histological slide. Figure 3 shows a sample digitized histologic section and the corresponding *T2w* MRI.

The MRI image size is 256×256. However, as the region of interest is only the peripheral zone for this study, we used the pixel values within this smaller peripheral zone region to form feature vectors for the training of SVM and RVM algorithms and for applying our unsupervised segmentation method.

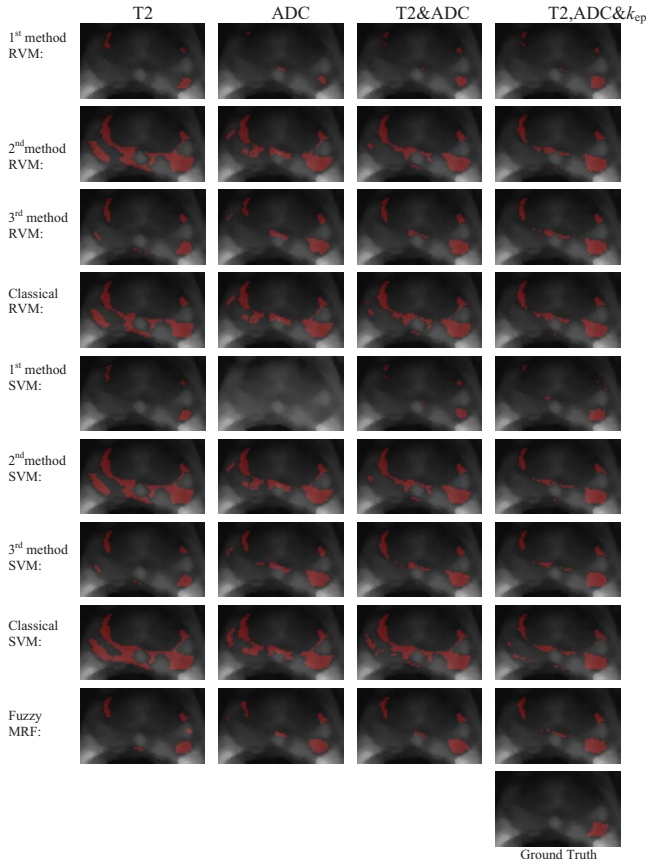

FIG. 4. Segmentation results for a sample patient, overlaid on the corresponding down-sampled *T2w* image.

III. EXPERIMENTS

III.A. Experimental setup

In this section, we describe the experimental setup and explain how we have applied the developed methods to multispectral MRI and performed the evaluation. In our experiments, we use one image representing each MRI technique. This image set was selected based on preliminary testing, where the parameter resulting in the highest accuracy was chosen for the final analysis. Thus, for *T2w* MRI, we use quantitative maps of *T2*; for DWI, we use ADC values; and for DCE-MRI, we use k_{ep}. As the images come from different imaging scans, a preprocessing step including image registration is necessary. The image registration for different modalities is done manually. Expert human readers have also extracted the peripheral zone from the images. Our data set consists of full 3D MRI including parametric images. However, only a subset of 2D slices (one slice per patient), which have tumors, are used.

We also need to normalize the intensity values since the intensity values for each image vary considerably across different images. Therefore, we transform the image values such that they have zero mean and unit variance with the following formula:

$$\text{Datanormalized} = \frac{\text{Dataraw} - \mu_{\text{PZ}}}{\sigma_{\text{PZ}}}, \tag{29}$$

where the Dataraw is the image to be normalized, and μ_{PZ} and σ_{PZ} are the mean value and the standard deviation of the data within the peripheral zone, respectively. Median filtering is used to reduce the noise before the normalization step.

TABLE I. Various acquisition parameters. ETL: Echo-train length, NEX: Number of excitations, and FOV: Field of view.

	TR/TE	Matrix size	ETL	Bandwidth (kHz)	NEX	FOV (cm)	Phase encode
T2-weighted MRI:	6550/101.5 ms	320×256	16	20.83		14	Left-right
DWI	4000/77 ms	128×256	144	166.7	10	14	Left-right
Quantitative *T2*	TR=2000 ms	256×128	1	31.25	1	20	Left-right
DCE-MRI	4.3/1.9 ms	256×128		62.5	0.5	20	Anterior-posterior

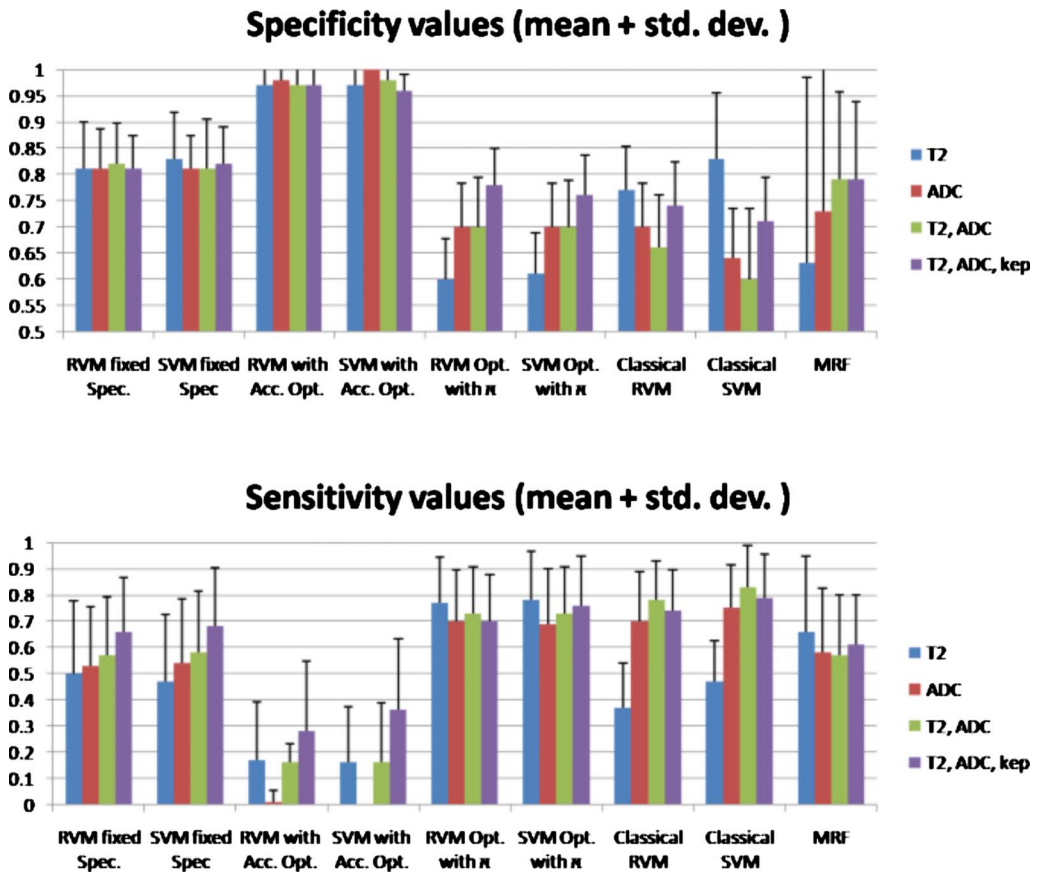

FIG. 5. Average specificity and sensitivity values with their corresponding variances for different algorithms and image sets.

The effect of different filter sizes was investigated as described later in this section.

After these preprocessing steps, we use SVM, RVM, and MRF algorithms separately to segment the prostate cancer on each patient. We use one-against-all approach for training, that is, for each test patient, we used the remaining 19 patients' data as training data, and use a polynomial kernel with polynomial order 1 for both SVM and RVM algorithms. A variant of LIBSVM code available in Ref. 28 for SVM algorithm, and we use the code available in Ref. 20 for RVM algorithm are used. Although SVM defines the threshold value as 0, we develop methods of varying this threshold targeted at improving a certain performance criterion as explained earlier. The selection of the threshold is purely based on the training data and is not affected by the test subject.

We calculate specificity, sensitivity, accuracy, and AUC values as measures of quantitative performance. The area under the curve values are calculated from receiver operational characteristics curve plots. First, the output of SVM and RVM are mapped into the region between 0 and 1 to give posterior probabilities as in Ref. 29. The AUC provides a performance criterion that is independent of the threshold value, and the other three criteria (specificity, sensitivity, and accuracy) depend on the threshold value. We have included both groups, since we both want to evaluate how well the images are fused together with appropriate weights (with AUC, independent of the threshold), and how well the binary segmentation results are (with specificity, sensitivity, and accuracy, depending on the threshold). For a given test data, we use the following definitions to calculate sensitivity value Sens and specificity value Spec, respectively:

$$\text{Spec} = \frac{\sum_{l=1, x \in S} (\text{DTN})_l}{\sum_{l=1, x \in S} (\text{TN})_l}, \tag{30}$$

where DTN can be formulated as

$$\text{DTN}_l = \left(\left(\sum_{j=1}^{m} h_j x_l^j + h_0 \right) < T \right) \cap (\text{TN})_l, \tag{31}$$

$$\text{Sens} = \frac{\sum_{l=1, x \in S} (\text{DTP})_l}{\sum_{l=1, x \in S} (\text{TP})_l}, \tag{32}$$

and DTP can be formulated as

$$(\text{DTP})_l = \left(\left(\sum_{j=1}^{m} h_j x_l^j + h_0 \right) > T \right) \cap (\text{TP})_l. \tag{33}$$

During our experiments, we first compare the localization performance of different image combinations. We first form the feature vectors by using the mean values of the cancer

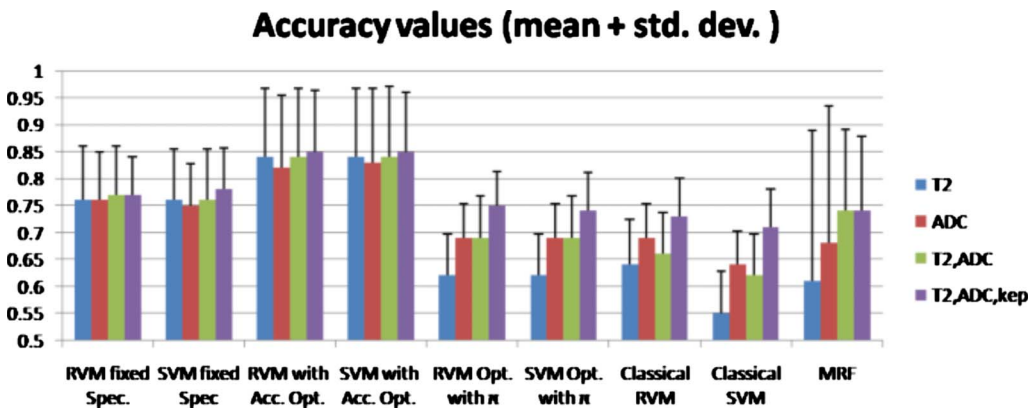

FIG. 6. The average accuracy and standard deviations for different algorithms and image sets.

and normal regions on each image for training. After training the SVM and RVM, we applied proposed thresholding algorithms and finally applied MRF on the same data set. Finding the optimal C value is currently an open research field for the SVM community in machine learning, and often cross validation is used for selecting this parameter. In our experiments, we have performed experiments for a range of C values and chosen the C value that performed well for all patients on average.

III.B. Experimental results

We show the segmentation results for an example patient in Fig. 4 for various combinations of multispectral MRI data and for all automated methods. In this figure, the first method corresponds to the accuracy maximizing, the second method corresponds to weighted accuracy maximizing, and the third method corresponds to fixing the specificity at a certain value (0.80). As shown in Fig. 4, when only $T2$ mapped values are used for classification, specificity decreases as it classifies excessive number of normal pixels as tumor. Combining all three $T2$, ADC, and k_{ep} images results in the most accurate segmentation.

Results that are presented in this section are obtained after using a 5×5 median filtering on the data set and show the results for different combinations of MR images. However the effect of using median filtering is also studied on Fig. 8.

Figure 5 shows the average specificity and sensitivity values obtained from RVM, SVM, and fuzzy MRF algorithms for 20 patients (using leave one out scheme) with the corresponding standard deviations. For SVM and RVM, we tested all three proposed approaches for thresholding. The average accuracy values with the corresponding variances are shown on Fig. 6 and the average AUC values with their standard deviation are similarly illustrated on Fig. 7.

It can be seen from Figs. 5 and 6 that the average *specificity* value is increased, as well as the average accuracy value, as more features derived from multispectral images are used for both SVM and RVM algorithms. For all algorithms, the highest accuracy values are obtained when all three types of $T2$, ADC, and k_{ep} images are combined for localization. Accuracy increases from 0.70 to 0.80 when $T2$, ADC, and k_{ep} are used compared to $T2$ only with weighted accuracy RVM. We observe that accuracy optimization in general provides improved accuracy results when compared to weighted accuracy results as expected. However, Fig. 5 shows that the sensitivity is low when this method is used, although the accuracy is the largest. Therefore, in view of Figs. 5 and 6 together, we conclude that threshold selection with weighted accuracy could be preferred when high sensitivity is desired, and accuracy maximization specificity fixing can be preferred when high specificity is desired. We also find that although our unsupervised technique can perform better than generic vector machines in certain cases; it performs poorly compared to vector machines with adjusted threshold.

Figure 7 shows the AUC values with the standard deviations. We observe that AUC values are higher for both SVM and RVM compared to MRF method, and also observe that expanding the MR data set results in higher AUC values, indicating improved segmentation performance. Comparing the accuracy values, observe that RVM slightly works better than SVM.

When we compare standard deviations in Figs. 5–7, we observe that MRF has higher standard deviation than supervised methods, indicating that supervised methods provide more stable segmentation results.

Finally, Fig. 8 shows the effect of using the median filtering when the feature vectors are formed of using $T2$, ADC,

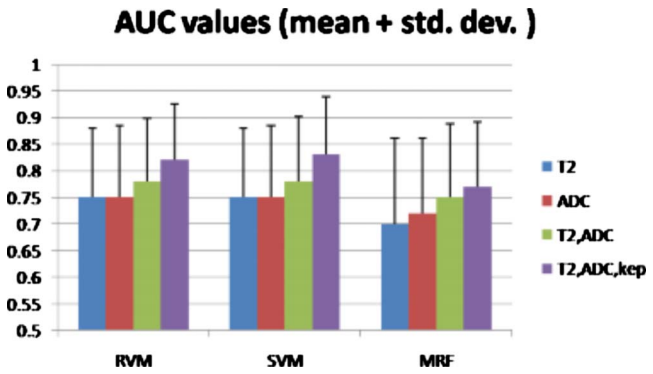

FIG. 7. Average AUC and standard deviation values.

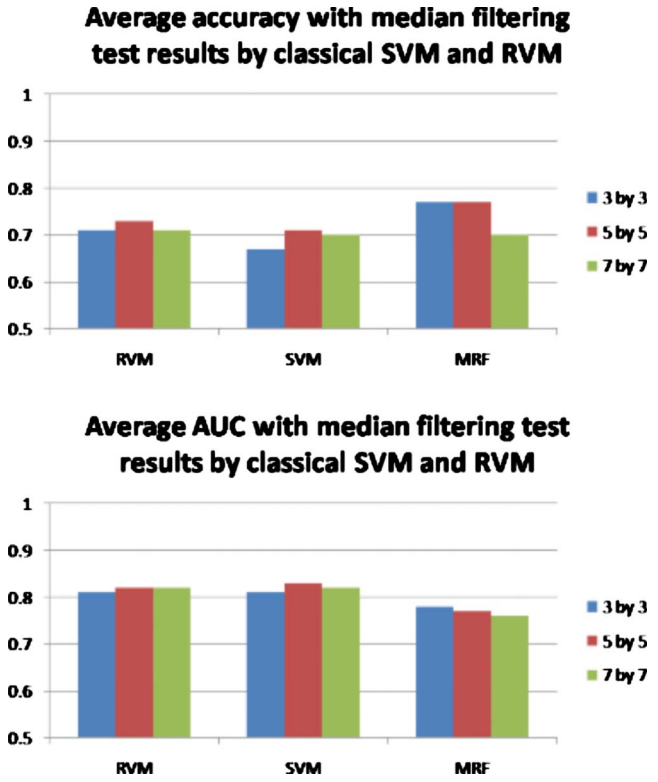

FIG. 8. Average median filtering test results by using classical RVM and SVM.

and k_{ep} images. This figure clearly shows that the average AUC and accuracy values are increased when the median filtering is used with the proposed thresholding algorithm. Based on the AUC results, 5×5 median filtering is applied to all data.

III.B.1. Statistical analysis

We perform a statistical analysis to find out if the differences in various image combinations and segmentation methods are significant or not. For that purpose we use a paired *t*-test and calculate the *p*-values. In Table II, we provide the *p*-values for the difference of AUC values in four image combinations using SVM, RVM, and MRF. We can see that the combination of *T*2, ADC, and k_{ep} is significantly better than *T*2, ADC; combination of *T*2 and ADC is significantly better than *T*2 alone and ADC alone; while the difference between *T*2 and ADC is not significant. In Table III, we can see that SVM and RVM are significantly better than MRF, and there is no significant difference between SVM and RVM.

IV. CONCLUSION AND DISCUSSION

In this study, we have used large margin classifiers' to localize the prostate cancer on multispectral MRI data sets and compare them to an unsupervised algorithm we recently developed. We have shown that using more than one type of MR image can increase the accuracy of prostate cancer localization and supervised methods perform better when compared to the unsupervised method.

We also developed methods for selecting a threshold to optimize a certain performance criterion and how different strategies of threshold selection can be used depending on the application. Experimental results show that using proposed threshold algorithms can tune SVM and RVM algorithms for prostate cancer segmentation yielding a desired improved performance with respect to a criterion that the user selects. We also proposed using a thresholding algorithm for the unsupervised method we have developed earlier, making it fully automatic.

We concluded that using RVM yield slightly better results when compared to the SVM algorithm, but not significantly. The main advantage of using RVM is that RVM does not require an additional optimization parameter such as the *C* parameter in SVM for an optimized classification result. Moreover, RVM learns the data with significantly lower number of relevance vectors compared to the number of support vectors for SVM.

Our results also show that using quantitative *T*2, ADC, and k_{ep} maps derived from multiecho FSE-MRI, DWI, and DCE-MRI, respectively, significantly increases the accuracy compared to the *T*2 alone, ADC alone, and *T*2, ADC combined.

One of the important topics to be considered in the future is the registration of different images, and automated transfer of the normal and cancer regions from histological image to the MRI images. The image registration step can affect the results for pixel based training since this process is manual. This is one of the reasons we use mean values for training: To decrease the registration error and variability. The manual segmentation of the PZ does not directly affect the classification or the segmentation results, since the training is performed on tumor and normal masks, which are independent of the PZ. Variation in PZ would only affect the performance criteria that are calculated since it changes the region for which these are calculated. A detailed analysis of manual registration and PZ is not directly related to the main context of this paper and outside the scope, but is an important future work to be performed.

Unsupervised methods showed better performance on

TABLE II. Comparison of AUC with different MR image combinations using *p*-values. "1" denotes *T*2, "2" ADC, and "3" k_{ep}. Combination of *T*2, ADC, and k_{ep} is significantly better than *T*2, ADC, and *T*2, ADC.

SVM	2	1,2	1,2,3	RVM	2	1,2	1,2,3	MRF	2	1,2	1,2,3
1	0.003	0.0009	0.003		0.17	0.05	0.003		0.47	0.017	0.009
2		0.056	0.003			0.089	0.004			0.087	0.002
1,2			0.009				0.017				0.019

TABLE III. Comparison of AUC with different methods using *p*-values. SVM and RVM are significantly better than MRF; there is no significant difference between SVM and RVM.

	RVM	MRF
SVM	0.12	0.019
RVM		0.05

some images, but not on average. Also, we can see that the standard deviation is larger on Figs. 5 and 6 with unsupervised methods. That is, supervised methods are more robust in terms of standard deviation when compared to unsupervised methods.

Since the study[8] reports that there is no optimal model for pharmacokinetic parameters, we used Brix model in this study because it does not require arterial input function, and is a classical compartmental model.

Choosing an alternative/improved kinetic model, designing a fully automated registration and PZ extraction method, and comparing their performances can be very useful and are good candidates for future work. The selection of the optimal MR image combinations is also an important problem for multispectral MRI segmentation for prostate cancer in order to reduce the data size, as well as to increase the segmentation accuracy. Therefore, an automated method that selects the most important image types with relevant information is an important topic for a future study.

One other interesting area of research would be to fine-tune the parameters of the protocol such that localization accuracy can be optimized. All methods developed in this paper certainly depend on the capability of a certain MR technique to produce contrast between tumor and normal regions. A revised imaging protocol has the potential to improve this capability and can be investigated in the future.

[a]Electronic mail: sozer1@iit.edu

[1]American Cancer Society, *Cancer Facts and Figures 2007* (American Cancer Society, Atlanta, 2007).
[2]W. Catalona *et al.*, "Measurement of prostate-specific antigen in serum as a screening test for prostate cancer," N. Engl. J. Med. **324**(17), 1156–1161 (1991).
[3]S. E. Eggener *et al.*, "Focal therapy for localized prostate cancer: A critical appraisal of rationale and modalities," J. Urol. (Baltimore) **178**(6), 2260–2267 (2007).
[4]F. Lee *et al.*, "Transrectal ultrasound in the diagnosis of prostate cancer: Location, echogenicity, histopathology, and staging," Prostate **7**, 117–129 (1985).
[5]H. Hricak, P. L. Choyke, S. C. Eberhardt, S. A. Leibel, and P. T. Scardino, "Imaging prostate cancer: A multidisciplinary perspective," Radiology **243**, 28–53 (2007).
[6]M. D. Rifkin, E. A. Zerhouni, C. A. Gatsonis, L. E. Quint, D. M. Paushter, J. I. Epstein, U. Hamper, P. C. Walsh, and B. J. McNeil, "Comparison of magnetic resonance imaging and ultrasonography in staging early prostate cancer. Results of a multi-institutional cooperative trial," N. Engl. J. Med. **323**, 621–626 (1990).
[7]S. E. Seltzer, D. J. Getty, C. M. Tempany, R. M. Pickett, M. D. Schnall, B. J. McNeil, and J. A. Swets, "Staging prostate cancer with MR imaging: A combined radiologist-computer system," Radiology **202**, 219–226 (1997).
[8]R. Srikanchana, D. Thomasson, P. Choyke, and A. Dwyer, "A comparison of pharmacokinetic models of dynamic contrast enhanced MRI," in Proceedings of the Symposium on Computer-Based Medical Systems (CBMS'04), 2004.
[9]G. Brix, W. Semmler, R. Port, L. Schad, G. Layer, and W. Lorenz, "Pharmacokinetic parameters in CNS Gd-DTPA enhanced MR imaging," J. Comput. Assist. Tomogr. **15**, 621–628 (1991).
[10]A. Tanimoto, J. Nakashima, H. Kohno, H. Shinmoto, and S. Kuribayashi, "Prostate cancer screening: The clinical value of diffusion-weighted imaging and dynamic MR imaging in combination with *T*2-weighted imaging," J. Magn. Reson Imaging **25**, 146–152 (2007).
[11]P. Gibbs, D. J. Tozer, G. P. Liney, and L. W. Turnbull, "Comparison of quantitative *T*2 mapping and diffusion-weighted imaging in the normal and pathologic prostate," Magn. Reson. Med. **46**, 1054–1058 (2001).
[12]M. Haider *et al.*, "Combined *T*2-weighted and diffusion weighted MRI for localization of prostate cancer," AJR, Am. J. Roentgenol. **189**, 323–328 (2007).
[13]I. Chan, W. Wells, R. V. Mulkern, S. Haker, J. Zhang, K. H. Zou, S. E. Maier, and C. M. Tempany, "Detection of prostate cancer by integration of line-scan diffusion, *T*2-mapping and *T*2-weighted magnetic resonance imaging; a multichannel statistical classifier," Med. Phys. **30**(9), 2390–2398 (2003).
[14]L. O. Hall *et al.*, "A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain," IEEE Trans. Neural Netw. **3**, 672–682 (1992).
[15]S. Ruan, S. Lebonvallet, A. Merabet, and J. M. Constans, "Tumor segmentation from a multispectral MRI images by using support vector machine classification," in Biomedical Imaging: From Nano to Macro, ISBI 2007, 4th IEEE International Symposium on ISBI 2007, 2007, pp. 1236–1239.
[16]S. Ozer, M. A. Haider, D. L. Langer, T. H. van der Kwast, A. J. Evans, M. N. Wernick, J. Trachtenberg, and I. S. Yetik, "Prostate cancer localization with multispectral MRI based on relevance vector machines," IEEE ISBI 2009 73–76 (2009).
[17]X. Liu, I. S. Yetik, D. L. Langer, M. A. Haider, Y. Yang, and M. N. Wernick, "Prostate cancer segmentation with simultaneous estimation of Markov random field parameters and class," IEEE Trans. Med. Imaging **28**(6), 906–915 (2009).
[18]V. Vapnik, *Statistical Learning Theory* (Wiley-Interscience, New York, 1998).
[19]E. Osuna, R. Freund, and F. Girosi, "Support vector machines: Training and applications," Technical Report A.I. Memo No. 1602 (MIT Artificial Intelligence Laboratory, 1997).
[20]M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," J. Mach. Learn. Res. **1**, 211–224 (2001).
[21]C. M. Bishop and M. E. Tipping, "Variational relevance vector machines," in Proceedings of Uncertainty in Artificial Intelligence, 2000, pp. 46–53.
[22]L. Wei, Y. Yang, R. M. Nishikawa, M. N. Wernick, and A. Edwards, "Relevance vector machine for automatic detection of clustered microcalcifications," IEEE Trans. Med. Imaging **24**(10), 1278–1285 (2005).
[23]S. Ozer and C. H. Chen, "Generalized Chebyshev kernels for support vector classification," International Conference on Patter Recognition, (ICPR 2008), pp. 1–4, 2008.
[24]J. Zeng and Z. Q. Liu, "Type-2 fuzzy Markov random fields to handwritten character recognition," in 18th International Conference on Pattern Recognition (ICPR'06), Vol. **1**, pp. 1162–1165 (2006).
[25]S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," IEEE Trans. on Pattern Analysis and Machine Intelligence **PAMI-6**(6), 721–741 (1984).
[26]N. Otsu, "A threshold selection method from gray-level histograms," IEEE Trans. Syst. Man Cybern. **9**(1), 62–66 (1979).
[27]J. E. McNeal, E. A. Redwine, F. S. Freihas, and T. A. Stamey, "Zonal distribution of prostatic adenocarcinoma, correlation with histologic pattern and direction of spread," Am. J. Surg. Pathol. **12**, 898–906 (1988).
[28]C. Chang and C. Lin, "LIBSVM: A library for support vector machines," 2001, http://www.csie.ntu.edu.tw/~cjlin/libsvm.
[29]J. Platt, "Probabilistic outputs for support vector machines and comparison to regularized likelihood methods," in *Advances in Large Margin Classifiers*, edited by A. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans (MIT Press, Cambridge, 2000).