A new prostate segmentation approach using multispectral Magnetic Resonance Imaging and a statistical pattern classifier

Bianca Maan[a], Ferdi van der Heijden[a] and Jurgen J. Fütterer[b]

[a] MIRA Institute for Biomedical Technology and Technical Medicine, University of Twente, the Netherlands;

[b] Department of Radiology, Radboud University Nijmegen Medical Centre, the Netherlands

ABSTRACT

Prostate segmentation is essential for calculating prostate volume, creating patient-specific prostate anatomical models and image fusion. Automatic segmentation methods are preferable because manual segmentation is time-consuming and highly subjective. Most of the currently available segmentation methods use *a priori* knowledge of the prostate shape. However, there is a large variation in prostate shape between patients.

Our approach uses multispectral magnetic resonance imaging (MRI) data, containing T1, T2 and proton density (PD) weighted images and the distance from the voxel to the centroid of the prostate, together with statistical pattern classifiers. We investigated the performance of a parametric and a non-parametric classification approach by applying a Baysian-quadratic and a k-nearest-neighbor classifier respectively. An annotated data set is made by manual labeling of the image. Using this data set, the classifiers are trained and evaluated.

sThe following results are obtained after three experiments. Firstly, using feature selection we showed that the average segmentation error rates are lowest when combining all three images and the distance with the k-nearest-neighbor classifier. Secondly, the confusion matrix showed that the k-nearest-neighbor classifier has the sensitivity. Finally, the prostate is segmented using both classifier. The segmentation boundaries approach the prostate boundaries for most slices. However, in some slices the segmentation result contained errors near the borders of the prostate. The current results showed that segmenting the prostate using multispectral MRI data combined with a statistical classifier is a promising method.

Keywords: prostate cancer, prostate segmentation, magnetic resonance imaging, multispectral MRI, statistical pattern classification, Bayes-quadratic classifier, k-nearest-neighbor classifier

1. INTRODUCTION

1.1 Background

Prostate cancer is one of the most widespread cancer types, and one of the leading causes of cancer death for the American male population. Fortunately, the survival rate is high when diagnosis is early. The current standard for the detection of prostate cancer is transrectal ultrasound (TRUS) guided needle biopsy, which has a high false negative rate associated with it.[1] Magnetic Resonance Imaging (MRI) is an attractive alternative for guiding and monitoring a biopsy intervention. Identification of prostate boundaries is required for calculating the prostate volume, image fusion and creating patient-specific prostate anatomical models. These models are used to predict the prostate's movement behavior during the biopsy procedure, and to plan the needle path.

Because manual segmentation of the prostate is time-consuming and highly subjective, automatic segmentation methods are preferable. However, segmenting the prostate in MR images is challenging due to the large variations of prostate shape between subjects, the lack of clear prostate boundaries and the similar intensity profiles of the prostate and surrounding tissues.

Further author information: (Send correspondence to Bianca Maan.)
B. Maan.: E-mail: b.maan@ewi.utwente.nl, Telephone: +31 53 489 2897

Medical Imaging 2012: Image Processing, edited by David R. Haynor, Sébastien Ourselin,
Proc. of SPIE Vol. 8314, 83142Q · © 2012 SPIE · CCC code: 0277-786X/12/$18 · doi: 10.1117/12.911194

Only a few prostate segmentation methods for MR images currently exist. Klein et al.[2] proposed a method based on nonrigid registration of a set of prelabeled atlas images, against the target patient's image, using mutual information. Subsequently, the segmentation is obtained as the average of the best matched registered atlas sets. The algorithm presented by Toth et al.[3] starts by isolating the region of interest (ROI) from magnetic resonance spectroscopy (MRS) data. MRS allows noninvasive and in vivo exploration of the metabolite concentrations within the tissue. Subsequently, an active shape model (ASM) is applied within the ROI to obtain the final segmentation. Vikhal et al.[4] used *a priori* knowledge of prostate shape to detect and refine the prostate contour. The procedure is initiated by the clinician and based on edge detection techniques. Clearly, all these approaches rely on *a priori* knowledge of prostate shape and use only a single MR image.

It is also possible to use multispectral MRI — comprising multiple MR sequences — to improve segmentation results. MRI mainly uses three properties of hydrogen atoms to acquire images: the relaxation times T1 and T2 and the proton density (PD). Although it is possible to separate these three components of the MR signal, in practice MR images weight and represent all three properties. A particular set of pulse sequence parameters will provide the best contrast between different tissue types. A series of these images can be combined to create a multispectral MRI data set from structural images. The multispectral data set based on weighted T1, T2 and PD images is already used for brain segmentation.[5,6]

1.2 Contribution

The aim of this study is to investigate whether multispectral MRI can also be used for whole prostate segmentation. In this paper, a four-dimensional feature space consisting of T1, T2 and PD weighted voxel intensities and a geometric distance feature is used. Our segmentation approach is based on statistical pattern classification. We investigated the application of a Baysian-quadratic classifier, assuming a normal density likelihood, and a k-nearest-neighbor classifier, which are representations of a parametric and a non-parametric approach respectively.

Using 5 datasets we performed feature selection and calculated confusion matrices to find the best combination of features and the best classifier.

1.3 Outline

The remainder of this paper is organized as follows. Section 2 describes the methods and Section 3 presents the experimental results. After that, Section 4 draws the conclusions.

2. METHODS

2.1 Data description

Five multispectral data sets consisting of T1, T2 and PD weighted axial images are utilized in this study. These datasets are completely anonymized and retrieved at the Radboud University Nijmegen Medical Centre. The five datasets are divided in three groups (A, B and C) according to their different sequences and tissue contrast. Groups B and C contain two datasets while group A only contains one dataset. The details about the sequences are listed in Table 1.

The preprocessing consisted of two steps:

- Rigid registration of the three image sets using a multiresolution approach and a mutual information metric. The images are registered using the "fast rigid registration" module in the open-source 3D slicer software.[7] Subsequently the first and the last slice, as well as a 20 pixels border around each slice are discarded to remove the empty pixels which originate during the registration.
- 3D Filtering using Gaussian kernels ($\sigma = 1$, kernelsize = 6.0 pixels) to remove noise in the image.

Table 1. Sequence details. This study uses five datasets which can be divided in three groups according to their sequences and tissue contrast.

	Dataset:	**A**	**B**	**C**
General information	Number of data sets	1	2	2
	MR	Siemens Skyra	Siemens Skyra	Siemens Trio TIM
T2-sequence	Voxel Dimensions [mm]	0.5*0.5*3	0.5*0.5*3	0.5625*0.5625*3
	Repetition Time [ms]	5560	6250	4480
	Echo Time [ms]	104	104	103
	Number of slices	19	21	19
T1-sequence	Voxel Dimensions [mm]	1.5*1.5*4	1.5*1.5*4	1.797*1.797*4.4
	Repetition Time [ms]	36	36	32
	Echo Time [ms]	1.41	1.41	1.47
	Number of slices	16	16	12
PD-sequence	Voxel Dimensions [mm]	1.5*1.5*4	1.5*1.5*4	1.797*1.797*4.4
	Repetition Time [ms]	800	800	800
	Echo Time [ms]	1.53	1.53	1.47
	Number of slices	16	16	12

Table 2. The total number of voxels included in the training set, and the percentage of voxels within the training set which are annotated as prostate transition zone, prostate peripheral zone and prostate as a whole.

		Prostate voxels		
Dataset	**Total**	**Transition** [%]	**Peripheral** [%]	**Total** [%]
A	389036	9.83	0.43	10.26
B1	314331	13.40	2.95	16.35
B2	379562	2.46	0.31	2.77
C1	343181	9.28	2.26	11.54
C2	223862	9.16	1.21	10.36

2.2 Training set

Subsequently, one training set is made per data set by manual annotation (by B.M.) of the T2 images, performed by drawing contours enclosing nine tissue regions; prostate (transition zone), prostate (peripheral zone), bladder, fat, bone, gluteus maximus, other muscles, rectum and the seminal vesicles. The prostate transition zone and peripheral zone are annotated separately because these regions have different MRI intensities. After annotation, we have a labeled data set (the training set) which is a subset of the voxels in the images. Each sample in the training set originates from one voxel in the data set.

For example, the training set of data set A contains 389036 samples divided in nine classes; in total 39915 voxels are annotated as prostate which corresponds to 10.26% of the training set. The number of annotated voxels per training set are listed in Table 2.

2.2.1 Additional feature

Besides the three MRI features we propose the addition of a geometric feature, namely the distance between the voxels and the centroid of the prostate. In this way, we add the prior knowledge that the prostate is always situated near the geometric center of the data set. With the addition of the distance feature, the radiometric (MRI) domain is combined with a geometric domain. When the noise-free geometric distance feature is added to the three MRI features, the result of the classifiers will mainly depend on the geometric feature. Therefore we need to control the uncertainty of the geometric feature by adding noise to this feature in the training set.

Firstly, the centroid is determined using the intersect of the manually segmented prostate (by B.M.) on the T2, T1 and PD data sets. Subsequently, noise is added to the (x,y,z)-position of the voxels with $\mu = 0$, $\sigma = 10.0$ times the within slice resolution. Subsequently the distance between the voxels and the centroid is calculated.

2.2.2 General overview

Finally, we obtained a training set which is divided in nine classes. Each sample in the training set originated from one voxel and is represented by a 4D vector comprising the intensities T1, T2 and PD, and the distance between the voxel and the centroid of the prostate. This vector is schematically described in Figure 1.

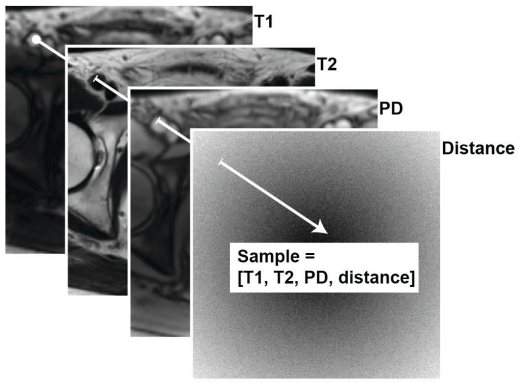

Figure 1. A training set can be obtained by manual annotation of the images. The T1, T2 and PD intensities together with the distance between the voxel and the centroid of the prostate are combined in one feature vector per voxel.

2.3 Classification experiments

Several classification experiments are performed under varying conditions of feature selection and classifiers to find the most promising settings for further investigations. Experiments are carried out in MATLAB using PRTools.[8,9] The following classifiers are used:

- Baysian-quadratic normal classifier (qdc)
- k-nearest-neighbor classifier (knnc).

To normalise the scale ratio between the features, the knnc is applied to a scaled data set with zero mean and unit variance in all feature directions. The qdc is applied directly to the unscaled data set because normal density based classifiers already use their variance estimation, and are thus insensitive to scaling of the features.

To prevent overtraining, a 5-fold cross-validation is performed over 10 repetitions to estimate the classification error of each classifier for different combinations of the four features. The distance feature is included in all feature combinations; the cross-validation is performed for seven unique combinations of the remaining three features. Selection of the best feature combination (feature selection) is done based on the outcome of the cross-validation. However, due to the long computation times when using the whole training set, only 1/5th of the set is used. For all feature combinations the same subset of the training set is used, only the combination of features is changed. For each feature combination the optimal knnc is determined by optimizing k with respect to the leave-one-out error of the subset.

In addition to cross-validation, the confusion matrix has also been assessed, which gives the frequency of both correctly and incorrectly classified voxels by tissue-type. Training and evaluation is done on two-disjoint, equal-sized subsets of the training set.

Finally, prostate segmentation is performed using both classifiers and the most promising feature combination. Training is performed using half of the training set.

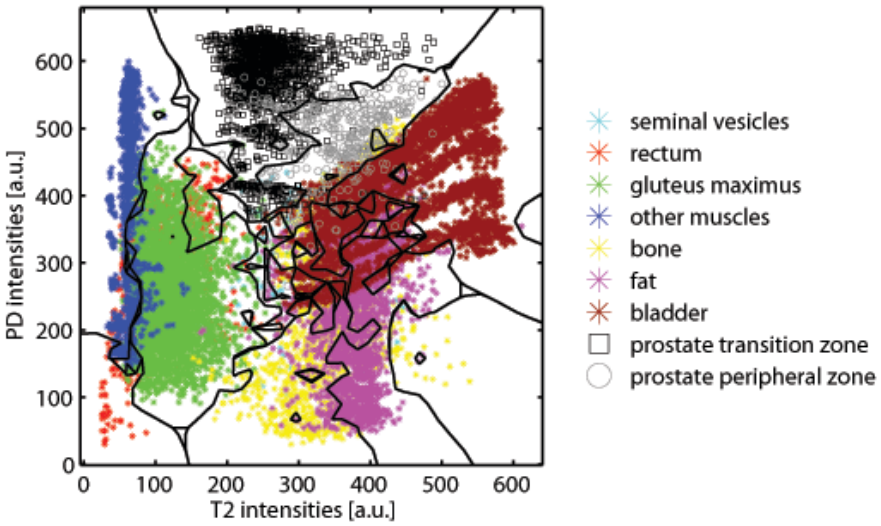

Figure 2. Scatter diagram of the T2 and PD voxel intensities together with decision boundaries determined with the knnc.

3. RESULTS AND DISCUSSION

Figure 2 shows a scatter diagram of the T2 and PD voxel intensities of training set C1. The black lines represent the classifier decision boundaries determined using the knnc. The figure shows a large overlap between the tissue classes when using a single MRI feature. For example, when using T2 images only, prostate and bone voxels have common intensities and cannot be distinguished.

However, after the addition of the PD intensities as extra feature, prostate and bone voxels can be easily distinguished because the PD intensity of these classes differ. The addition of extra MRI and geometric features can result in a better classification result because there will be less overlap between the classes.

3.1 Feature selection

Table 3 and Table 4 show a summary of the results which are obtained during the feature selection experiments. The average error rate, which is weighted by the frequencies of each class, is shown in Table 3 for all feature combinations. In accordance with the scatter diagram in Figure 2, the largest errors are obtained when using only one MRI feature, and the error decreases after the addition of extra features. Clearly, the lowest classification error is obtained using all four features in combination with the knnc.

Table 4 shows the prostate error rate, which is the average error of both prostate classes (transition and peripheral zone) weighted by their frequencies. Except data set A, the lowest classification error is obtained using all four features and the knnc. Data set A had a small peripheral zone with intensities comparable to the transition zone. For that, many voxels within the peripheral zone are classified as transition zone voxels. However, since we are interested in the segmentation of the prostate as a whole, the misclassification between these classes are not of interest.

3.2 Confusion matrix

The confusion matrix gives the frequency of both correctly and incorrectly classified voxels by tissue-type. Since we are interested in segmentation of the prostate as a whole, we combined the two prostate classes. In other words, misclassification between the two prostate classes is ignored. Table 5 shows the sensitivity (prostate correctly classified as prostate) and the specificity (other tissue correctly classified as other tissue) of both the qdc and knnc using all four features.

When combining the two prostate classes, the sensitivity of the classifiers are 97.99% (std: 0.65%) and 98.56% (std: 1.33%) using the qdc and knnc respectively. The specificity, "other tissue" correctly classified as "other tissue", of both classifiers is 99.69% (std: 0.19%) and 99.87% (std:0.08%) for the qdc and knnc respectively.

Table 3. Average classification error for all feature combinations and both qdc and knnc. The feature 'distance' is included in all feature combinations. The presented numbers are percentages.

			Feature combinations						
			T1	T2	PD	T1+T2	T1+PD	T2+PD	T1+T2+PD
Average Error	**A**	qdc	25.94	16.31	25.83	9.52	13.51	7.83	5.01
		knnc	22.62	14.78	22.36	6.71	8.60	5.34	2.57
	B1	qdc	23.45	11.51	20.08	6.03	8.81	6.52	2.81
		knnc	22.08	10.94	17.71	3.93	5.78	4.27	1.46
	B2	qdc	21.61	15.78	29.16	4.22	9.95	10.61	2.65
		knnc	19.72	15.40	27.08	3.17	7.65	8.56	1.40
	C1	qdc	26.52	19.55	31.41	10.94	18.45	13.06	5.87
		knnc	23.63	18.65	28.92	8.62	12.36	10.15	3.45
	C2	qdc	27.57	13.76	27.81	10.10	16.08	8.27	6.30
		knnc	24.67	12.24	24.71	8.16	11.14	6.48	3.45

Table 4. Prostate classification error for all feature combinations and both qdc and knnc. The feature 'distance' is included in all feature combinations. The presented numbers are percentages.

			Feature combinations						
			T1	T2	PD	T1+T2	T1+PD	T2+PD	T1+T2+PD
Prostate Error	**A**	qdc	6.66	5.64	5.80	5.67	6.38	5.15	5.56
		zknnc	6.20	5.51	5.57	5.58	5.22	4.99	5.30
	B1	qdc	8.84	11.80	8.49	7.27	7.85	5.83	5.24
		knnc	9.20	11.48	8.41	5.83	7.25	4.99	4.51
	B2	qdc	14.05	8.57	12.26	5.81	7.84	6.40	4.57
		knnc	17.19	10.77	17.90	7.08	11.58	9.47	5.60
	C1	qdc	21.79	12.22	25.28	10.15	18.63	9.35	7.45
		knnc	17.36	11.48	21.68	8.82	10.35	5.76	3.57
	C2	qdc	16.21	12.90	14.48	12.03	15.42	10.27	10.00
		knnc	15.17	13.38	14.01	10.99	11.72	8.59	5.74

Table 5. Confusion matrices: Frequency of correctly and incorrectly classified prostate voxels within the test set using qdc and knnc and all four features. The frequency is expressed as the percentage of the total number of voxels within the true label class.

			Estimated labels			
			qdc		**knnc**	
			prostate	other tissue	prostate	other tissue
True Labels	**A**	prostate	98.18	1.82	98.90	1.10
		other tissue	0.38	99.62	0.24	99.76
	B1	prostate	98.82	1.18	98.90	1.10
		other tissue	0.47	99.53	0.16	99.84
	B2	prostate	97.39	2.61	96.23	3.77
		other tissue	0.49	99.51	0.11	99.89
	C1	prostate	97.26	2.74	99.37	0.63
		other tissue	0.14	99.86	0.07	99.93
	C2	prostate	98.29	1.71	99.41	0.59
		other tissue	0.06	99.94	0.05	99.95

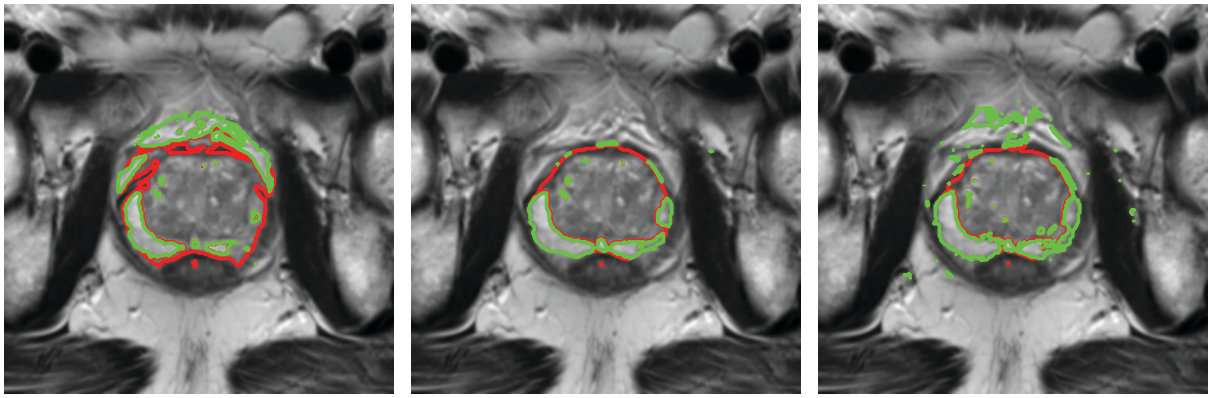

Figure 3. Prostate classification result (slice 5 data set C1). The outline of the "transition zone" in red together with the "peripheral zone" in green. Left: qdc classification result using T2 and distance feature. Middle: qdc classification result using all features. Right: knnc classification result using all features.

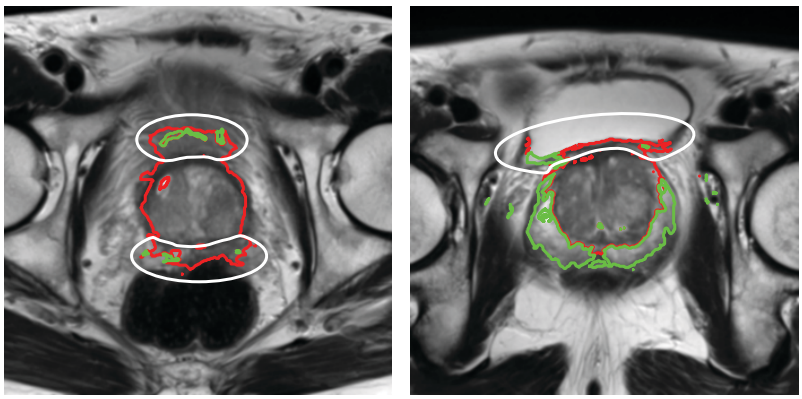

Figure 4. Prostate classification result using qdc. There is a large classification error at the boundaries of the prostate; the area inside the white elipses should not be classified as prostate. Left: Data set A slice 12. Right: Data set B1 slice 9.

3.3 Prostate segmentation

The feature selection and confusion matrix experiments in the previous section are applied to the annotated data sets. The boundary of the tissues is not included in any tissue class and does not influence the results of these experiments. Therefore we also applied the classifiers to the whole MRI data set which includes the non-annotated boundary regions. The results of this segmentation approach shows the clinical usability of the classifiers.

Figure 3 shows the classification results of data set C1 using the qdc in combination with two (T2 and distance) and four (all) features. The red line represents the "transition zone" outline and the green line represents the "peripheral zone". Clearly, using four features, the outline approaches the prostate boundary better.

Figure 3 also shows the difference between the qdc and knnc. Although in the feature selection and confusion matrix experiments the knnc performed better than the qdc, in the prostate segmentation experiment the classification results are comparable. Both classifiers have some regions of misclassified voxels. The small classification errors of the knnc result in Figure 3 can be easily removed by selection of the largest connected region. However, near the bladder some data sets have misclassified voxels attached to the segmented prostate, for examples see Figure 4. These voxels are more difficult to remove during postprocessing. Figure 5 shows a correct classification from data set B1, and a misclassification between the prostate and bladder in one slice of data set A.

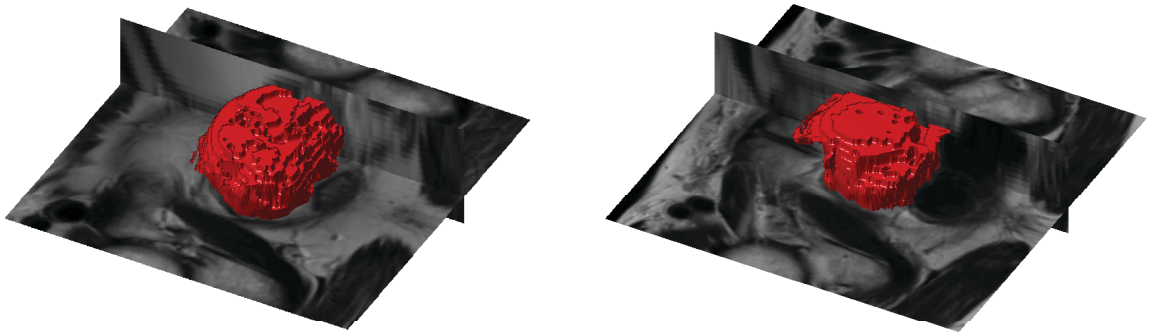

Figure 5. Three dimensional prostate classification result using qdc. Left: Data set C1 shows a correct prostate classification. Right: Data set A shows a misclassification between the prostate and the bladder, equal to Figure 4.

4. CONCLUSIONS AND FUTURE WORK

The feature selection and confusion matrix experiments show that prostate segmentation using a k-nearest-neighbor classifier combined with multispectral MRI data is a promising method. The sensitivity and specificity of the presented knnc, when tested with an annotated data set, is high (see Table 5).

However, the boundary of the tissues is not included in the annotated data set and does not influence the results of feature selection and confusion matrix experiments. The classification of the MRI data shows that the classification of the boundary voxels – especially the boundary between the prostate and the bladder – needs to be improved. Adding the misclassified voxels to the bladder class, with other words, extending the bladder class, could overcome these problems.

Further optimization of the method can be achieved by adding texture information as additional features. Currently, the prostate is segmented per voxel without taking into account any information from the neighbor voxels. For example, when one voxels belongs to the prostate, it is more certain that the neighbor voxels also belong to the prostate class. Therefore, adding information from the neighbor voxels as additional features can improve the classification.

REFERENCES

[1] Rabbani, F., Stroumbakis, N., Kava, B. R., Cookson, M. S., and Fair, W. R., "Incidence and clinical significance of false-negative sextant prostate biopsies.," *The Journal of urology* **159**, 1247–50 (Apr. 1998).

[2] Klein, S., van der Heide, U. A., Lips, I., van Vulpen, M., Staring, M., and Pluim, J. P., "Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information.," *Med Phys* **35**, 1407–1417 (Apr. 2008).

[3] Toth, R., Tiwari, P., Rosen, M., Kalyanpur, A., Pungavkar, S., and Madabhushi, A., "A multi-modal prostate segmentation scheme by combining spectral clustering and active shape models," *Proceedings of SPIE* , 69144S–69144S–12 (2008).

[4] Vikal, S., Haker, S., Tempany, C., and Fichtinger, G., "Prostate contouring in MRI guided biopsy.," *Proceedings of SPIE* **7259**, 72594A (Mar. 2009).

[5] Andersen, A. H., Zhang, Z., Avision, M. J., and Gash, D. M., "Automated segmentation of multispectral brain MR images," *Journal of Neuroscience Methods* **122**(1), 23–13 (2002).

[6] Mohamed, F. B., Vinitski, S., Faro, S. H., Gonzalez, C. F., Mack, J., and Iwanaga, T., "Optimization of tissue segmentation of brain MR images based on multispectral 3D feature maps," *Magnetic Resonance Imaging* **17**(3), 409–403 (1999).

[7] Pieper, S., Lorensen, B., Schroeder, W., and Kikinis, R., "The na-mic kit: Itk, vtk, pipelines, grids and 3d slicer as an open platform for the medical image computing community," *Proceedings of the 3rd IEEE International Symposium on Biomedical Imaging: From Nano to Macro* **1**, 698–701 (2006).

[8] Duin, R., Juszczak, P., Paclik, P., Pekalska, E., de Ridder, D., Tax, D., and Verzakov, S., "PRTools4.1, A Matlab Toolbox for Pattern Recognition," *Delft University of Technology* (2007).

[9] van der Heijden, F., Duin, R., de Ridder, D., and Tax, D. M. J., [*Classification, Parameter Estimation and State Estimation: An Engineering Approach Using Matlab*], Wiley (2004).