

Spectral Clustering of Shape and Probability Prior Models for Automatic Prostate Segmentation

S. Ghose^{1,2}, J. Mitra^{1,2}, A. Oliver², R. Martí², X. Lladó²,
J. Freixenet², J. C. Vilanova³, J. Comet⁴, D. Sidibé¹ and F. Meriaudeau¹

Abstract—Imaging artifacts in Transrectal Ultrasound (TRUS) images and inter-patient variations in prostate shape and size challenge computer-aided automatic or semi-automatic segmentation of the prostate. In this paper, we propose to use multiple mean parametric models derived from principal component analysis (PCA) of shape and posterior probability information to segment the prostate. In contrast to traditional statistical models of shape and intensity priors, we use posterior probability of the prostate region determined from random forest classification to build, initialize and propagate our model. Multiple mean models derived from spectral clustering of combined shape and appearance parameters ensure improvement in segmentation accuracies. The proposed method achieves mean Dice similarity coefficient (DSC) value of 0.96 ± 0.01 , with a mean segmentation time of 0.67 ± 0.02 seconds when validated with 46 images from 23 datasets in a leave-one-patient-out validation framework.

Index Terms—Prostate segmentation, random forest, statistical shape and posterior probability models, spectral clustering.

I. INTRODUCTION

Accurate prostate segmentation in TRUS may aid in biopsy needle placement and multimodal image fusion between TRUS and magnetic resonance imaging (MRI) to improve malignant tissue extraction during biopsy [1]. However, accurate computer aided prostate segmentation from TRUS images is a challenging task due to low contrast of TRUS images, speckle, and shadow artifacts, inter-patient prostate shape, size and deformation variations and heterogeneous intensity distribution inside the prostate gland. To deal with the prostate segmentation challenges we propose a novel prostate segmentation method in which appearance and spatial context based information from the training images are used to achieve a probabilistic classification of the prostate. Thereafter, the statistical shape and appearance model derived from PCA of prostate shape and posterior probabilistic values of the prostate region of the training

TRUS images propagates in a multi-resolution framework to segment a test image. However instead of a single mean model, multiple mean models are created using spectral clustering from the combined shape and appearance parameters to capture larger variations. Finally, the mean model that has the lowest fitting error with the test image provides accurate segmentation of the prostate. The performance of our method is compared with the traditional AAM [3] and also with our previous works [11], [14], [13]. The key contributions of this work are: 1) The use of random forest classification framework to obtain a soft classification of the prostate. 2) Use of spectral clustering on prostate shape and probability priors to group similar prostates creating multiple mean models.

The remaining paper is organized in the following manner. Section II provides a description of the proposed segmentation framework, followed by the results and discussions in Section III. Finally, the paper is concluded in Section IV with the conclusions and future works.

II. PROPOSED SEGMENTATION FRAMEWORK

The proposed method is developed on three major components: A) supervised learning framework of random forest to determine posterior probability of a pixel being prostate, B) adapting statistical models of shape and intensity priors to incorporate the posterior probabilities of the prostate region for training, initialization and propagation of the parametric model and C) use of spectral clustering to group prostate multiple mean models having similar shape and appearances. These components of the proposed segmentation method are described in the following subsections.

A. Random forest based probabilistic classification

In traditional AAM [3], the appearance model is built from the intensity distribution of the prostate region. However, depending on that ultrasound image acquisition parameters and the machine manufacturer, the intensities of the prostate may vary significantly leading to an inaccurate appearance model. Therefore, to reduce inter-dataset intensity variations and intensity variations inside the prostate region, we propose to determine the posterior probability of the image pixels being prostate in a supervised learning framework of random forest and use PCA of the posterior probabilities of the prostate region to build our appearance model. Decision trees are discriminative classifiers which are known to suffer from over-fitting. However, a random decision forest or random

This work was supported by VALTEC 08-1-0039 of Generalitat de Catalunya, Spanish Science and Innovation grant nb. TIN2011-23704, Spain and Conseil Régional de Bourgogne, France.

¹ S. Ghose, J. Mitra, D. Sidibé & F. Meriaudeau are with Le2i CNRS-UMR 6306, Université de Bourgogne, Le Creusot, France. soumyaghose@gmail.com, jhimli.mitra@u-bourgogne.fr, dro-desire.sidibe@u-bourgogne.fr, fabrice.meriaudeau@u-bourgogne.fr

² R. Martí, A. Oliver, X. Lladó & J. Freixenet are with the Computer Vision and Robotics Group, University of Girona, Girona, Spain. marly@eia.udg.edu, aoliver@eia.udg.edu, llado@eia.udg.edu, jordif@eia.udg.edu

³ J. C. Vilanova is with the Girona Magnetic Resonance Center, Girona, Spain.

⁴ J. Comet is with the Hospital Dr. Josep Trueta, Girona, Spain.

forest achieves better generalization by growing an ensemble of many independent decision trees on a random subset of the training data and by randomizing the features made available at each node during training [2].

During **training**, to minimize the pose and intensity variations, our datasets are rigidly aligned and the inter-patient intensity variations are normalized. The data consists of a collection of $V = (X, F)$, each centered at 3×3 neighborhood of pixels. Where, $X = (x, y)$ denotes the pixel position and the feature vector F constitutes of the mean and standard deviation of the 3×3 pixel neighborhood. Each tree τ_i in random forest receives the full set V , along with the label and the root node and selects a test to split V into two subsets to maximize the information gain. A test constitutes of a feature (like the mean of a pixel neighborhood) and a feature response threshold. The left and the right child nodes receive their respective subsets of V and the process is repeated at each child node to grow the next level of the tree. Growth is terminated when either the information gain is minimum or the tree has grown to a maximum depth specified. Each decision tree in the forest is unique as each tree node selects a random subset of features and threshold.

During **testing**, the test image is rigidly aligned to the same frame of the training datasets and its intensities are normalized. The pixels are routed to one leaf in each tree by applying the test (selected during training). Each pixel of the test dataset is propagated through all the trees by successive application of the relevant binary test to determine the probability of belonging to class c . When reaching a leaf node l_τ , where $\tau \in [1, \dots, \Gamma]$, the posterior probabilities ($P_\tau(c|V)$) are gathered in order to compute the final posterior probability of the pixel defined by $P(c|V) = \frac{1}{\Gamma} \sum_{\tau=1}^{\Gamma} P_\tau(c|V)$.

B. Statistical Shape and Appearance Model

The process of building the parametric statistical model of shape and appearance variations involves the task of building a shape model, an appearance model, and consecutively a combined model of shape and appearance priors. To build the shape model, a PDM [3] is built by equal angle sampling of the prostate contours to determine the landmarks automatically. The PDM of the contours are aligned to a common reference frame by generalized Procrustes analysis [5]. PCA of the aligned PDMs identifies the principal modes of shape variations. Posterior probabilistic information (of pixels being prostate) of the segmented region are warped into correspondence using a piece-wise affine warp and are sampled from a shape-free reference similar to that of AAM [3]. PCA of the posterior probabilities from Section II-A is used to identify their principal modes of variation. The model may be formalized in the following manner. Let s and t represent the shape and posterior probability models, then

$$s = \bar{s} + \Phi_s \theta_s, \quad t = \bar{t} + \Phi_t \theta_t \quad (1)$$

where \bar{s} and \bar{t} denote the mean shape and posterior probability information respectively, then Φ_s and Φ_t contain the first p eigenvectors (obtained from 98% of total variations) of

the estimated joint dispersion matrix of shape and posterior probability information and θ represent the corresponding eigenvalues. The model of shape and posterior probability variations are combined in a linear framework as,

$$b = \begin{bmatrix} W\theta_s \\ \theta_t \end{bmatrix} = \begin{bmatrix} W\Phi_s^T(s - \bar{s}) \\ \Phi_t^T(t - \bar{t}) \end{bmatrix} \quad (2)$$

where W denotes a weight factor (determined as in AAM [3]) coupling the shape and the probability space. A third PCA of the combined model ensures the reduction in redundancy of the combined model, and is given as,

$$b = E\alpha \quad (3)$$

where E is the matrix of eigenvectors and α the appearance parameters.

In our model, we use the optimization framework similar to that proposed by Cootes et al. [3]. The objective function of our model is similar to AAM. However, instead of minimizing the sum-of-squared differences of intensities between the mean model and target image, we minimize the sum-of-squared differences of the posterior probabilities of the mean model and the target image. The prior knowledge of the optimization space is acquired by perturbing the combined model with known model parameters and perturbing the pose (translation, scale and rotation) parameters. Linear relationships between the perturbation of the combined model (δc) and the residual posterior probability values (δt) (obtained from the sum-of-squared differences between the posterior probabilities of the mean model and the target image), and between the perturbation of the pose parameters (δp) and the residual posterior probability values are acquired in multivariate regression frameworks as,

$$\delta c = R_c \delta t, \quad \delta p = R_p \delta t \quad (4)$$

R_c and R_p refer to the correlation coefficients. Given a test image, the posterior probability values of the pixels being prostate is determined with random forest soft classification. The sum-of-squared differences of the posterior probability values with the mean model is used to determine the residual value δt . The combined model (δc) and the pose parameters (δp) are then updated using Eq. (4) to generate a new shape, and combined model and consequently the new posterior probabilities. The process continues in an iterative manner until the difference of the mean model with the target image remains unchanged.

C. Spectral clustering and multiple mean models

Traditionally a statistical shape and appearance model assumes the shape and the appearance spaces to be Gaussian. However in real cases, inter-patient prostate intensity distribution, prostate shapes and sizes may vary significantly. In such circumstances, a single mean model is inefficient to capture the variations of shape and appearance spaces and it would be inappropriate to approximate them with a single Gaussian distribution. To address this problem, we propose to use multiple mean models. Therefore, the objective is to employ a grouping scheme that considers the prostates with similar shape and appearance parameters to build a set of mean models and produce an accurate optimization

space as described in Section II-B. Furthermore, the number of mean models should change dynamically depending on combined shape and appearance parameters of the training datasets. To address both the issues, we propose to use spectral clustering in the combined space of shape and appearance parameters. Spectral clustering relies on the eigen structure of a similarity matrix to partition points into disjoint clusters with high intra-cluster similarity and also high inter-cluster dissimilarity. Moreover, the number of clusters could be determined dynamically from the principal components in the eigen space. During training, the combined shape and appearance eigenvectors are obtained by Eq. (2) and Eq. (3). Cosine similarities of the combined vectors (E) of \mathcal{P} eigenvectors are used to construct a $\mathcal{P} \times \mathcal{P}$ similarity or affinity matrix \mathcal{W} that measures the similarity between the \mathcal{P} points. $\mathcal{W}_{i,j}$ is large when the points indexed by i and j are likely to be in the same cluster. The problem may be defined in terms of a complete graph with vertices $v = 1, \dots, \mathcal{P}$ and an affinity matrix with weights $\mathcal{W}_{i,j}$, for $i, j \in v$. The objective is to determine K disjoint clusters $A = (A_k)_{k=1, \dots, K}$, where $\bigcup_k A_k = v$, that optimizes the cost function of K -way normalized cut defined as,

$$C(A, \mathcal{W}) = \sum_{k=1}^K \left(\sum_{i \in A_k, j \in v \setminus A_k} \mathcal{W}_{i,j} \right) / \left(\sum_{i \in A_k, j \in v} \mathcal{W}_{i,j} \right) \quad (5)$$

The K -way normalized cut $C(A, \mathcal{W})$ may be simplified as $D^{-1/2} \mathcal{W} D^{-1/2}$ where $D = \text{diag}(\mathcal{W})$ [15]. The algorithm may be summarized as,

- 1) **Input:** Similarity matrix $\mathcal{W} \in \mathbb{R}^{\mathcal{P} \times \mathcal{P}}$
- 2) Compute first k eigenvectors of $D^{-1/2} \mathcal{W} D^{-1/2}$ where $D = \text{diag}(\mathcal{W})$ to build the matrix U .
- 3) Re-normalize the matrix U .
- 4) Perform k-means clustering on the normalized U .
- 5) **Output:** Similar disjoint clusters k .

The number of clusters k is determined by the number of the largest k -eigen vectors (obtained from 98% of total variations) of the normalized Laplacian $C(A, \mathcal{W})$ of the affinity matrix. The k-means clustering therefore groups the similar prostates to form k mean models.

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

We have validated the accuracy and robustness of our method with 46 axial mid gland TRUS images of the prostate with a resolution of 348×237 pixels from 23 prostate datasets in a leave-one-patient-out evaluation strategy. Manual segmentations performed by an expert radiologist were validated by an experienced urologist to prepare the ground truth. For the random forest based classification, we have fixed the number of trees to 100, the tree depth to 30 and the lower bound of information gain to 10^{-7} . These parameters were chosen empirically as they produced promising results with the test images. During validation, the test dataset is removed and multiple mean models are built with the aid of spectral clustering. The number of mean models varied between 5 to 7 depending on the training datasets. Given an

TABLE I
PROSTATE SEGMENTATION QUANTITATIVE COMPARISON (HD, AND
MAD ARE IN MM.)

Method	DSC	HD	MAD
AAM [3]	0.94 ± 0.03	4.92 ± 0.96	2.15 ± 0.94
Ghose [14]	0.95 ± 0.06	2.53 ± 1.94	0.87 ± 1.23
RF-AAM	0.95 ± 0.06	3.76 ± 2.03	1.40 ± 0.91
Our Method	0.96 ± 0.01	2.51 ± 0.93	0.84 ± 0.31

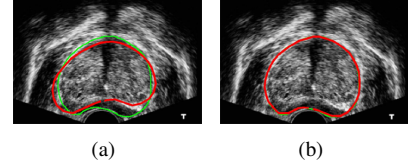


Fig. 1. Improvement in segmentation accuracies with spectral clustering and multiple mean models. (a) shows segmentation without multiple mean models and (b) shows segmentation with multiple mean models.

image of the test dataset, all the mean models are applied for segmentation in parallel. Fitting or registration errors between the mean models after final deformation and the test image are computed by normalizing the pixel-wise intensity differences. Finally, the segmentation by the mean model that provided the least fitting error is selected as optimum segmentation. We have used most of the popular prostate segmentation evaluation metrics like DSC, 95% Hausdorff Distance (HD), and mean absolute distance (MAD) to evaluate our method. Furthermore, the results are compared with the traditional AAM proposed by Cootes et al. [3] and to our previous work [14]. It is observed from Table I that a probabilistic representation of the prostate appearance in TRUS images significantly improves segmentation accuracy when compared to traditional AAM. We achieved a statistically significant improvement in t -test p -value < 0.0001 for DSC, HD and MAD compared to traditional AAM [3]. As observed in Table I that [14] and RF-AAM (that uses posterior probabilities from random forest classification) provide better results compared to AAM justifying the use of posterior probability of the prostate region instead of intensity. However, our proposed method provides excellent results when compared to [14] and RF-AAM, suggesting that the use of both posterior probability and multiple mean models are essential to improve segmentation accuracies as illustrated in Fig. 1. In Fig. 2 we illustrate the necessity of a deformable model based segmentation after random forest classification. We observe that random forest classification produces misclassified regions as shown in Fig. 2(a). Our statistical shape and probability prior models working on initial segmentation improve segmentation accuracies as observed in Fig. 2(c). To provide qualitative results of our method we present a subset of results in Fig. 3. The first row shows the results achieved with AAM [3] and the second row shows the results achieved with our method. A quantitative comparison of different prostate segmentation methodologies is difficult in absence of a public dataset and standardized evaluation metrics. Nevertheless, to have an overall qualitative estimate of the functioning of our method we have compared with some of the existing works in Table II. Analyzing the results we observe that our mean DSC value is better compared

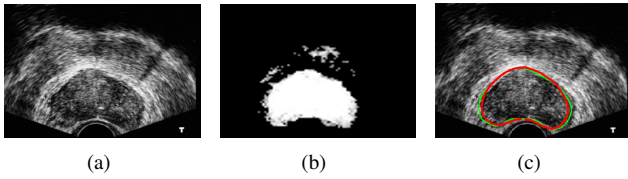


Fig. 2. Illustration of the requirement of shape and probability prior model. (a) is the images to be segmented, (b) shows the random forest classification with mis-classified regions and (c) shows the final segmentation achieved.

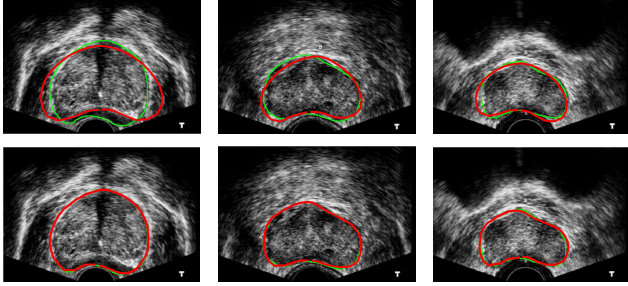


Fig. 3. The green contour gives the ground truth and the red contour gives the obtained result. Row 1 show the results achieved with AAM and row 2 show results of our model for the corresponding prostates.

to area overlap accuracy values of [8], [9], [13], and very close to the area overlap error value of [10]. However, it is to be noted that we have validated with more images compared to [10]. Our MAD value is better compared to [8], [10], [9] and [13]. From these observations we may conclude that qualitatively our method performs well in overlap and contour accuracy measures. Our method is implemented in Matlab 7 on an Intel Core i5, 2.8 GHz processor and 8 GB RAM. The mean segmentation time of the method is 0.67 ± 0.02 seconds with an unoptimized Matlab code. Even with an unoptimized Matlab code in Table II we observe that our mean segmentation time is less when compared to [8], [10], [12], [13], [11] and only more compared to [1]. However, [1] used an optimized C++ code to achieve their results.

Comparing our proposed model with our previous work [11] we observe that we achieve similar contour and area overlap accuracies. However, our proposed method achieves prostate segmentation in significantly less time compared to [11]. The mean segmentation time of [11] with current machine configuration is 4.33 ± 0.21 seconds while the proposed method using supervised learning with random forest takes 0.67 ± 0.02 seconds without compromising the segmentation accuracies. Mean segmentation time is often a critical element in selecting one segmentation method over the other for near real-time multimodal image fusion between TRUS and MRI to improve malignant tissue sampling during biopsy. In this context, we may claim that our present method shows improvement over our previous work in [11]. In contrast to [11] where the number of mean models was dependent on a user defined threshold of fitting error, our present work uses spectral clustering to determine the number of mean models.

IV. CONCLUSIONS AND FUTURE WORKS

A novel approach of multiple statistical models of shape and posterior probability information of prostate region with

TABLE II
QUALITATIVE COMPARISON OF PROSTATE SEGMENTATION. TIME IS
MEASURED IN SECS, PXS=PIXELS, MM=MILLIMETER

Reference	Area Acc.	Contour Acc.	Datasets	Time
Betrouni [8]	Overlap $93 \pm 0.9\%$	Dist- 3.77 ± 1.3 pxs	10 images	5
Shen [10]	Error $3.98 \pm 0.97\%$	Dist- 3.2 ± 0.87 pxs	8 images	64
Ladak [9]	Accuracy $90.1 \pm 3.2\%$	MAD- 4.4 ± 1.8 pxs	117 images	-
Cosio [12]	-	MAD- 1.65 ± 0.67 mm	22 images	660
Yan [1]	-	MAD- 2.10 ± 1.02 mm	19 datasets/ 301 images	0.3
Ghose [13]	DSC 0.95 ± 0.02	MAD- 1.26 ± 0.51 mm	6 datasets/ 24 images	1.50
Ghose [11]	DSC 0.97 ± 0.006	MAD- 1.82 ± 0.76 pxs	23 datasets/ 46 images	5.97
Our Method	DSC 0.96 ± 0.01	MAD- 3.00 ± 1.22 pxs/ 0.84 ± 0.34 mm	23 datasets/ 46 images	0.67

the goal of segmenting the prostate in 2D TRUS images has been proposed. Our approach is accurate, and robust to significant shape, size and contrast variations in TRUS images when compared to traditional AAM. While the proposed method is validated with prostate mid gland images, effectiveness of the method on the base and the apical slices is yet to be validated.

REFERENCES

- [1] P. Yan et al., "Discrete Deformable Model Guided by Partial Active Shape Model for TRUS Image Segmentation," *IEEE Trans. on Biomed. Engg.*, vol. 57, pp. 1158–1166, 2010.
- [2] E. Geremia et al., "Spatial decision forests for MS lesion segmentation in multi-channel MR images," in *MICCAI*, 2010, vol. 6361, pp. 111–118.
- [3] T.F. Cootes et al., "Active Appearance Models," in *Proc. of Europ. Conf. on Comp. Vision*, 1998, pp. 484–498.
- [4] T.F. Cootes et al., "The Use of Active Shape Model for Locating Structures in Medical Images," *Imag. & Vis. Comp.*, vol. 12, pp. 355–366, 1994.
- [5] J.C. Gower, "Generalized Procrustes Analysis," *Psychometrika*, vol. 40, pp. 33–51, 1975.
- [6] H. Liu et al., "Automatic Segmentation of Prostate Boundaries in Transrectal Ultrasound (TRUS) Imaging," in *Proc. of the SPIE Med. Imag.*, 2002, pp. 412–423.
- [7] K. Diaz and B. Castaneda, "Semi-automated Segmentation of the Prostate Gland Boundary in Ultrasound Images Using a Machine Learning Approach," in *Proc. of SPIE Med. Imag.*, 2008, vol. 69144, pp. 69144A–1–8.
- [8] N. Betrouni et al., "Segmentation of Abdominal Ultrasound Images of the Prostate Using A priori Information and an Adapted Noise Filter," *Comp. Med. Imag. & Graph.*, vol. 29, pp. 43–51, 2005.
- [9] H.M. Ladak et al., "Prostate Segmentation from 2D Ultrasound Images," in *Proc. of IEEE EMBS*, 2000, pp. 3188–3191.
- [10] D. Shen et al., "Segmentation of Prostate Boundaries from Ultrasound Images Using Statistical Shape Model," *IEEE Trans. on Med. Imag.*, vol. 22, pp. 539–551, 2003.
- [11] S. Ghose et al., "Multiple Mean Models of Statistical Shape and Probability Priors for Automatic Prostate Segmentation," in *MICCAI - Prostate Cancer Imaging*, 2011, vol. LNCS 6963, pp. 35–46.
- [12] F.A. Cosío, "Automatic Initialization of an Active Shape Model of the Prostate," *Med. Imag. Anal.*, vol. 12, pp. 469–483, 2008.
- [13] S. Ghose et al., "Statistical shape and texture model of quadrature phase information for prostate segmentation," *IJCARS*, vol. 7, pp. 43–55, 2012.
- [14] S. Ghose et al., "A probabilistic framework for automatic prostate segmentation with a statistical model of shape and appearance," *IEEE ICIP*, pp. 713–716, 2011.
- [15] F. Bach et al., "Learning Spectral Clustering," in *NIPS*, 2003.