Spectral Clustering of Shape and Probability Prior Models for Automatic Prostate Segmentation in Ultrasound Images
Soumya Ghose, Jhimli Mitra, Arnau Oliver, Robert Marti, Xavier Llado, Jordi Freixenet, Joan Vilanova, Josep Comet, Désiré Sidibé, Fabrice Mériaudeau

To cite this version:

HAL Id: hal-00710953
https://hal.archives-ouvertes.fr/hal-00710953
Submitted on 22 Jun 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Spectral Clustering of Shape and Probability Prior Models for Automatic Prostate Segmentation in Ultrasound Images

S. Ghose¹,², J. Mitra¹,², 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 forest achieves better generalization by growing an ensemble of many independent decision trees on a random subset of the

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, drodesire.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, jfreix@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.
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 $3 \times 3$ 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 $3 \times 3$ pixel neighborhood. Each tree $\tau_i$ in random forest receives the full set $V$, along with the label and the root node 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 applying the test (selected during training). Each pixel of the normalized. The pixels are routed to one leaf in each tree by same frame of the training datasets and its intensities are specified. Each decision tree in the forest is unique as each tree node selects a random subset of features and threshold.

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 relevant binary test to determine the probability of belonging to class $c$. When reaching a leaf node $l_r$, where $\tau \in [1, \ldots, \Gamma]$, the posterior probabilities ($P_t(c|V)$) are gathered in order to compute the final posterior probability of the pixel defined by $P(c|V) = \frac{1}{\sum_{r=1}^{\Gamma} P_t(c|V)}$.

\begin{equation}
\text{where } \pi = \Phi_s \theta_s, \quad t = \overline{1} + \Phi_t \theta_t
\end{equation}

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 \( v = 1, ..., \mathcal{P} \) and an affinity matrix with weights \( \mathcal{W}_{i,j} \), for \( i, j \in \mathcal{V} \). The objective is to determine \( K \) disjoint clusters \( \mathcal{A} = (A_k)_{k=1}^{\mathcal{K}} \), where \( \bigcup_k A_k = \mathcal{V} \), that optimizes the cost function of \( K \)-way normalized cut defined as,

\[
C(A, \mathcal{W}) = \sum_{k=1}^{\mathcal{K}} \left( \sum_{i \in A_k, j \in \mathcal{V} \setminus A_k} \mathcal{W}_{i,j} \right) / \left( \sum_{i \in A_k, j \in \mathcal{V}} \mathcal{W}_{i,j} \right)
\]

(5)

The \( K \)-way normalized cut \( C(A, \mathcal{W}) \) may be simplified as \( D^{-1/2}WD^{-1/2} \) 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 \( D^{-1/2}WD^{-1/2} \) 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 image of the test dataset, the image is classified with random forest and all the mean models are initialized at

![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)

the center of gravity of the classified region (as prostate) to segment the prostate 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. 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 to area overlap accuracy values of [8], [9], [13], and very close to the area overlap error value of [10]. However, it
Table II
QUALITATIVE COMPARISON OF PROSTATE SEGMENTATION. TIME IS MEASURED IN SECONDS, PXS=PIXELS, MM=MILLIMETER

<table>
<thead>
<tr>
<th>Reference</th>
<th>Area Acc.</th>
<th>Contour Acc.</th>
<th>Datasets</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Betrouni [8]</td>
<td>Overlap 93±0.9%</td>
<td>Dist-3.77±1.3 pxs</td>
<td>10 images</td>
<td>5</td>
</tr>
<tr>
<td>Shen [10]</td>
<td>Error 3.98±0.97%</td>
<td>Dist-3.2±0.87 pxs</td>
<td>8 images</td>
<td>64</td>
</tr>
<tr>
<td>Ladak [9]</td>
<td>Accuracy 90.1±3.2%</td>
<td>MAD-4.4±1.8 pxs</td>
<td>117 images</td>
<td>-</td>
</tr>
<tr>
<td>Cosio [12]</td>
<td>-</td>
<td>MAD-1.65±0.67 mm</td>
<td>22 images</td>
<td>660</td>
</tr>
<tr>
<td>Yan [1]</td>
<td>-</td>
<td>MAD-2.10±1.02 mm</td>
<td>19 datasets/301 images</td>
<td>0.3</td>
</tr>
<tr>
<td>Ghose [13]</td>
<td>DSC 0.95±0.02</td>
<td>MAD-1.26±0.51 mm</td>
<td>6 datasets/24 images</td>
<td>1.50</td>
</tr>
<tr>
<td>Ghose [11]</td>
<td>DSC 0.97±0.006</td>
<td>MAD-1.82±0.76 pxs</td>
<td>23 datasets/46 images</td>
<td>5.97</td>
</tr>
<tr>
<td>Our Method</td>
<td>DSC 0.96±0.01</td>
<td>MAD-3.00±1.12 pxs/0.84±0.34 mm</td>
<td>23 datasets/46 images</td>
<td>0.67</td>
</tr>
</tbody>
</table>

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 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