Multi-scale Feature Learning on Pixels and Super-pixels for Seminal Vesicles MRI Segmentation

Qinquan Gao\textsuperscript{1}, Akshay Asthana\textsuperscript{1}, Tong Tong\textsuperscript{1}, Daniel Rueckert\textsuperscript{1}, Philip "Eddie" Edwards\textsuperscript{2}

\textsuperscript{1}Department of Computing, Imperial College London, United Kingdom; \textsuperscript{2}Department of Surgery and Cancer, Imperial College London, United Kingdom

ABSTRACT

We propose a learning-based approach to segment the seminal vesicles (SV) via random forest classifiers. The proposed discriminative approach relies on the decision forest using high-dimensional multi-scale context-aware spatial, textual and descriptor-based features at both pixel and super-pixel level. After affine transformation to a template space, the relevant high-dimensional multi-scale features are extracted and random forest classifiers are learned based on the masked region of the seminal vesicles from the most similar atlases. Using these classifiers, an intermediate probabilistic segmentation is obtained for the test images. Then, a graph-cut based refinement is applied to this intermediate probabilistic representation of each voxel to get the final segmentation. We apply this approach to segment the seminal vesicles from 30 MRI T2 training images of the prostate, which presents a particularly challenging segmentation task. The results show that the multi-scale approach and the augmentation of the pixel based features with the super-pixel based features enhances the discriminative power of the learnt classifier which leads to a better quality segmentation in some very difficult cases. The results are compared to the radiologist labeled ground truth using leave-one-out cross-validation. Overall, the Dice metric of 0.7249 and Hausdorff surface distance of 7.0803 mm are achieved for this difficult task.

Keywords: Random Forest Classifier, Graph Cut, Superpixel, Multi-scale feature

1. INTRODUCTION

Automatic and accurate delineation of the prostate structure from MR image is important in providing models for diagnosis, cancer detection and planning of interventions. The seminal vesicles (SV) are the glands close to the prostate that are frequently included in the clinical target volume of EBRT treatments.\textsuperscript{1} There has been a lot of research on prostate segmentation but comparatively little focusing on the surrounding structures, especially on the SV. The SV is a small structure that has large intensity variation and irregular shape in the MRI images. It is extremely difficult to register two individuals together and a segmentation approach based only on the intensity information is unlikely to succeed. During our initial experiments, several multi-atlas based methods including non-rigid registration based,\textsuperscript{2} patch-based,\textsuperscript{3} and sparse-coding based\textsuperscript{4} methods failed to segment the SV region. Moreover, the patient-specific deformable model based approach\textsuperscript{5} require manual annotation of the corresponding landmarks which is an extremely tedious and error-prone task. It is very difficult to place accurate corresponding landmarks on the irregular shape of the SV of different training images. The automatic and accurate delineation of the SV remains a challenge as most methods in the published literature failed to obtain a good segmentation.

Alternatively, automatic delineation of the SV from MR images can be treated as a binary classification problem to distinguish the SV structure from the background. Recently, the random forest has been shown to be an efficient and robust classifier and has become very popular for similar segmentation tasks in computer vision.\textsuperscript{6} Using random feature selection to split each node yields error rates that compare favorably to Adaboost and max-margin classifiers such as SVMs, but are more robust with respect to noise. Random forests have been successfully applied to medical image segmentation.\textsuperscript{7,8} These methods are based on learning discriminative features for voxel-wise label prediction, which can handle irregular structure segmentation. Inspired by these methods, we explore the use of random forest classifiers to solve the difficult SV segmentation problem.

E-mail: \{q.gao10,a.asthana,t.tong11,Eddie.Edwards,d.rueckert\}@imperial.ac.uk
In this paper, we propose a nearly automatic learning-based method for SV segmentation. Given a test image, a global random forest classifier which is based on pixel-wise discriminative descriptors is trained using a selected set of most similar images. This classifier is used to obtain the first SV segmentation. Then, to further refine the segmentation, a graph cut is applied to optimize the energy function which is based on a Markov random field (MRF) with the random forest predicted probability as a prior. The performance of the proposed method compares favorably to existing methods. However, the quality of boundary delineation is still not good. Superpixel based classifiers have been shown to outperform the previously published state-of-the-art pixel based method on natural image segmentation. Since the SV region contains many bubble-like structures, we propose a combination of superpixel-level features and the pixel level features for more robust boundary classification. Moreover, in recently published face recognition work, it has been shown that the high-dimensional features result in higher performance. Inspired by this, we explore the use of high-dimensional features by adopting a multi-scale feature scheme for random forest classifier training. The results show that the multi-scale features on both pixel and superpixel level improve the segmentation accuracy.

2. METHOD

We propose a learning based approach for segmentation of the seminal vesicles. The method (Figure 1) consists of two parts: probabilistic segmentation obtained by a well trained random forest classifier and segmentation refinement using the prior probability and graph cuts. The choice of classifier for the proposed approach is dictated by the fact that the random decision forest can handle the over fitting problem by using many independent decision trees on a random subset of the training data and randomly selecting the features for each training node, thereby achieving fast and accurate prediction. Note that other non-linear methods (such as support vector machines) can be seamlessly merged into the proposed framework. In order to construct powerful discriminative features to train a classifier, we propose the use of high-dimensional multi-scale features based on the pixel and super-pixel level, which include spatial, appearance and descriptor-based features. We apply this approach to the T2 SV MRI training images.

2.1 Data Preprocessing

The 30 T2-weighted MR training images of the seminal vesicles exhibit large shape and intensity variation. To minimize this variability, we performed initial processing steps as mentioned in previous work. All the images were preprocessed to transform them to a template space and then selected the atlas for each image.
were firstly denoised followed by bias correction to reduce noise and make the intensity more homogeneous. The preprocessed images were then globally transformed into the template space by affine registration. Seminal vesicles are quite close to the prostate but have irregular shapes. In order to improve the registration, eleven landmarks (2D view shown in Fig. 2) are manually placed on the prostate to guide initial affine registration which is followed by intensity-based affine registration. In the template space, the intensity of images were normalized to [0, 100] within the dilated masked region of interest (ROI) making the contrast and luminance relatively consistent.

2.2 Probabilistic Segmentation via Random Forest Classifier

For initial segmentation, a random forest classifier is trained on the SV region to achieve probabilistic prediction for each voxel. Each voxel in the masked region is treated as a training sample. To compute a discriminative feature, we employ three types of spatial and context-aware features, which are spatial, multi-scale textual and multi-scale descriptor-based features extracted at pixel and super-pixel levels. Every instantiated feature \( F(v_i) \), where voxel \( v_i \in ROI \), with its unique parameters corresponds to one dimension of the feature space \( \Gamma \) used by the decision trees. Since there is a large spacing in the z direction, the multi-scale descriptor-based features are computed in the 2D space slice-by-slice.

### Spatial Feature

\( F(v_i, S) \) consists of spatial position of each voxel \( v_i \), the distance \( d_{v_p, v_i} \) to the prostate center \( v_p \), which is estimated as the average of the manual input landmarks around the prostate, and the angles of the vector \( \beta(Vv_p \rightarrow v_i) \) in the x, y, z direction respectively, shown in Fig. 2. Our results indicate that the spatial features are significant to the success of the global classification.

### Textual Feature

\( F(v_i, T) \) is used to capture the intensity variation within the pixel neighborhood. In each 2D slice, the gradient and hessian matrix are computed. This feature consists of the difference between intensity means and the standard deviation of a rectangular patch around point \( v_i \) and a randomly selected point \( v_j \) in the neighborhood with a searching window 23 \( \times \) 23. To make the process multi-scale, patch sizes are set to \{3 \( \times \) 3, 5 \( \times \) 5, 7 \( \times \) 7\} and the number of randomly selected neighborhood points is fixed to 30.

### Descriptor Feature

\( F(v_i, H) \) is based on the histograms of oriented gradient (HOG) features as these are robust to intensity variation and provide useful information for object segmentation. For each voxel, we crop a patch in its 2D slice and multi-scale HOG descriptors are computed using the bin size of 9, the clipping values of 0.2 for normalization and the following sets of parameters (Table 1):

<table>
<thead>
<tr>
<th>Patch Size</th>
<th>Cell Size</th>
<th>Block Size</th>
<th>Patch Size</th>
<th>Cell Size</th>
<th>Block Size</th>
<th>Patch Size</th>
<th>Cell Size</th>
<th>Block Size</th>
<th>Patch Size</th>
<th>Cell Size</th>
<th>Block Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 ( \times ) 8</td>
<td>4</td>
<td>2</td>
<td>16 ( \times ) 16</td>
<td>4</td>
<td>3</td>
<td>24 ( \times ) 24</td>
<td>6</td>
<td>3</td>
<td>32 ( \times ) 32</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>8 ( \times ) 8</td>
<td>8</td>
<td>1</td>
<td>16 ( \times ) 16</td>
<td>8</td>
<td>2</td>
<td>24 ( \times ) 24</td>
<td>12</td>
<td>2</td>
<td>32 ( \times ) 32</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>16 ( \times ) 16</td>
<td>16</td>
<td>1</td>
<td>24 ( \times ) 24</td>
<td>24</td>
<td>1</td>
<td>32 ( \times ) 32</td>
<td>32</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Parameters for HOG calculation

An example of the sparse HOG descriptor with different parameter sets is visualized in Fig. 2. It presents useful information to distinguish the SV region from the background.
Figure 3. The sparse HOG examples with different parameter sets: 1) patch size 16 × 16, cell size: 16 and block size 1; 2) patch size 16 × 16, cell size: 8 and block size 2; 3) patch size 16 × 16, cell size: 4 and block size 3.

**Super-Pixel Feature** $F_s(v_i, H)$ is based on the intensity histogram of the super-pixel normalized with respect to the neighboring super-pixels. For this, each 2D slice is pre-segmented into super-pixel patches using SLIC super-pixel method, as shown in Figure 2. Then, a histogram is computed for each super-pixel using the intensity values of each pixel within the super-pixel. To compute the final feature $F_s(v_i, H)$ for each super-pixel, a normalized histogram of super-pixel neighborhood is computed using the $K$ nearest super-pixels. Let $h_i$ denote the histogram of super-pixel $S_i$, $h_{i,j}$ the histogram of neighborhood super-pixel $S_{i,j}$, then:

$$H_i = h_i + \sum_{k=1}^{K} h_{i,j}$$ (1)

$$F_s(v_i, H) = \frac{H_i}{||H_i||_1}$$ (2)

To capture all the possible ranges of super-pixel feature, $K$ is set as $K = \{3, 4, 5, 6\}$ and the bin size is set to 16. After that, $F_s$ is assigned to each pixel in the super-pixel $S_i$ equally. The motivation behind constructing this feature is that it provides a very reliable source to capture useful boundary information.

**Final High-Dimensional Feature** $F(v_i)$ is computed by concatenating all the above discussed features $F(v_i) = [F(v_i, S); F(v_i, T); F(v_i, H); F_s(v_i, H)]$. A balance training of random forest classifier is carried out by random selection of the same number of negative training samples as the available positive samples. For a test image, the predicted label probabilities $P_f$ are gathered from each tree in the forest in order to compute the final posterior probability of the voxel, defined by $P_f = \frac{1}{N} \sum_{i=1}^{N} P_{f,i}$. To account for potential prediction error, the probability map is smoothed by a Gaussian with $\sigma = 0.5mm$.

### 2.3 Graph-cut segmentation refinement

For the probabilistic segmentation, an energy function based on Markov random fields (MRF) optimized by graph-cuts is applied to refine the segmentation. Let $I$ is the target image, $f_i$ is to define the label to each voxel $i$. A MRF based energy function can be formulated as follows:

$$E = \lambda \sum_{i \in I} D_i(f_i) + \sum_{\{i,j\} \in N} V_{i,j}(f_i, f_j)$$

where, $D_i(f_i)$ is the data term which measures disagreement between the a priori probability from the random forest classifier and the observed data. The data term is estimated using Monte Carlo expectation-maximization (MCEM) segmentation, which generates a probability map for each class for each voxel. The smoothing term $V_{i,j}(f_i, f_j)$ between the neighborhood voxels in a 3D patch is defined as:

$$V_{i,j}(f_i, f_j) = w \times \frac{1}{ln(1 + (I_i - I_j)^2 + \varepsilon)}$$ (3)

where $I_i$ is the intensity for voxel $i$, $\varepsilon$ is a very small value chosen to avoid division by zero. Here, $w$ is defined by $\frac{1}{D}$ where $D$ is the distance between voxels $i$ and $j$ and $\lambda$ is the trade-off between the data and smoothing term. A graph cut is then applied to optimize the energy function and achieve the final segmentation.
3. EXPERIMENTS AND RESULTS

The whole segmentation algorithm is implemented in MATLAB on a quad-core 3.20 GHz processor with 64 GB RAM and applied to 30 SV MR T2 images from NCI. A leave-one-out strategy has been adopted for experiments and random forest classifier is trained for each individual separately using the 5 most similar images from the two data sets. The similar images are selected based on normalized mutual information on the masked SV region. In order to find the best segmentation method, three main experiments were conducted: (1) Using single scale descriptor features only with the dimension of 480 (i.e. HOG is computed on a patch size of $16 \times 16$ with 3 sets of parameters shown in Table 1); (2) Using single scale descriptor features and super-pixel features with the dimension of 496 (i.e. HOG is computed using the same parameters as experiment 1 and super-pixel features are computed using 4-nearest neighborhood super-pixels); (3) Using multi-scale descriptor and super-pixel features with the dimension of 1327 (i.e. both features are calculated at multi-scale level as discussed in Section 2.2). Note that the spatial and textual features are used for all the three experiments. In the random forest training, the parameters are empirically set as follows: Experiments (1) and (2) uses 200 trees and 150 randomly selected feature dimensions; Experiment (3) uses 200 trees and 300 randomly selected feature dimensions. Manual intervention is required to place 11 landmarks into the prostate region for spatial information calculation and to refine the region of interest when there is a registration error. In the graph-cut based segmentation step, 2 Gaussian distributions are applied to model the background and foreground intensity distributions, with the trade off $\lambda = 5$ and $\epsilon = 0.001$ to get the best refinement.

<table>
<thead>
<tr>
<th>Results</th>
<th>Experiment (1)</th>
<th>Experiment (2)</th>
<th>Experiment (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dice Metric</td>
<td>0.695 ± 0.076</td>
<td>0.708 ± 0.069</td>
<td>0.725 ± 0.06</td>
</tr>
<tr>
<td>Hausdorff Distance</td>
<td>8.279 ± 2.69</td>
<td>7.72 ± 2.57</td>
<td>7.08 ± 2.23</td>
</tr>
</tbody>
</table>

Table 2. Evaluation of segmentation results for Experiments (1)-(3).

The results are shown in Table 2, from which, we can see that by adding superpixel level features, the accuracy of segmentation improved slightly. With the high dimensional multi-scale features at both pixel and super-pixel level, the overall performance increases significantly with dice metric from 0.695 to 0.725 and Hausdorff distance from 8.279 to 7.08. From the sample result in Fig. 4, we can see that by augmenting the pixel-based features with the super-pixel features, the segmentation improves especially around the difficult boundary regions. These are modest results for seminal vesicles segmentation but compare favorably to the methods which failed (with Dice metric < 0.4).

Figure 4. Sample Segmentation Results: a) Ground-Truth; b) Experiment (1); c) Experiment (2); d) Experiment (3)

However, the high-dimensional features increase the computation burden, which needs around 3 hours to train a classifier. During the experiments, we also tried to reduce the dimensions by using PCA but this resulted in significant decrease in the performance. We believe that this is due to the complexity of the multi-scale features and more sophisticated non-linear approaches might be required to efficiently reduce the dimensions of the features without losing discriminative information.

4. CONCLUSION AND FUTURE WORK

We proposed a nearly automatic approach for SV segmentation using random forest classifiers and graph cuts. The results show that multi-scale features improve the segmentation accuracy significantly and the superpixel based features introduce more discriminative characteristics to the features, particularly around the boundary,
which makes the proposed method very robust and feasible for clinical application. For future work, non-linear feature learning will be introduced to extracted more discriminative features to improve the quality of classification and also reduce the computation time. We aim to exploit such higher dimensional features and subspace learning methods to improve segmentation for various other images and target organs.

Acknowledgment

This research was funded by Cancer Research UK under project A8087/C24250 and the scholarship from China Scholarship Council. We are grateful to National Cancer Institute (US) for providing the seminal vesicles image data. The work of Akshay Asthana is funded by Marie Curie International Incoming Fellowship under the FP7-PEOPLE-2011-IIF Grant agreement no. 302836.

REFERENCES