An automatic multi-atlas based prostate segmentation using local appearance-specific atlases and patch-based voxel weighting

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

\textsuperscript{1} Department of Computing  
Imperial College London, United Kingdom  
q.gao10@imperial.ac.uk

\textsuperscript{2} Department of Surgery and Cancer  
Imperial College London, United Kingdom  
Eddie.Edwards@imperial.ac.uk

\textbf{Abstract.} Prostate segmentation facilitates prostate cancer detection and can help to diagnose the pathological stage of disease. Segmented anatomical models may also help to improve the outcome of robotic-aided laparoscopic prostatectomy (RALP) by augmented reality image guidance. In this paper, we present a fully automated segmentation pipeline for multi-center and multi-vendor MRI prostate segmentation using a multi-atlas approach with local appearance-specific voxel weighting. Segmenting prostates with a large variation of shape and intensity still remains a significant challenge. In this work, the atlases with the most similar global appearance are classified into the same categories. Sum-of-square local intensity difference after affine registration is used for atlas selection and after non-rigid registration, a local patch-based atlas fusion is performed using voxel weighting based on the local patch distance. Such multi-atlas segmentation is a widely used method in brain segmentation. We thoroughly evaluated the method on 50 training images by performing a leave-one-out study. Dice coefficient and volumetric overlapping accuracy are used to quantify the difference between the automatic and manual segmentation. Compared to the manual gold standard segmentation, our proposed method produced favorable outcomes in these highly variable data sets, with an average Dice coefficient $0.8467 \pm 0.0435$. The result shows that the algorithm presented could be used to aid the delineation of the prostate from diverse MRI images, which may be useful in a number of clinical applications.

Keywords: Multi-atlas segmentation, atlas segmentation, image registration, patch-based segmentation, local weighting

1 Introduction

Accurate segmentation and location of the prostate is crucial for prostate cancer detection and staging, surgical planning and image-guided robotic-aided laparoscopic prostatectomy (RALP) with augmented reality (AR). Currently,
the majority of segmentation is done by well-trained radiologists based on the anatomical knowledge of the appearance of relevant physical structures in the preoperative scans. This is very time consuming to achieve manually, especially for a large number of segmentation. Therefore, there is a pressing need for fast, automatic segmentation methods for clinical applications. There is much previous work based on the well known methods of statistical shape modeling and probabilistic atlas priors learned from training data \cite{3}.

In a recent study, multi-atlas segmentation provides the best accuracy compared to a number of algorithms for the segmentation of subcortical structures \cite{1}. However, as there are large shape variations and intensity differences in the appearance of the prostate in scans with different acquisition protocols, an accurate, generic, robust and fully automated segmentation remains a significant research challenge. In order to tackle this problem, in multi-atlas segmentation, the atlas in the database which is the most similar to the query image is used. Several methods have been investigated and compared to improve this selection by Lotjonen \cite{5}. Also, during the multi-atlas fusion process, it is reasonable to expect the atlases whose reference images are more similar to the target image should contribute more to the segmentation \cite{10}. Considering all these factors, Klein \cite{4} proposed a multi-atlas matching for prostate segmentation using localized mutual information, which achieves a good prostate segmentation. However, due to local structure variation, a global weighting strategy for label fusion may not achieve the most accurate delineation of the local boundary.

In this paper, we introduce a multi-atlas segmentation using local appearance-specific atlases, which is more robust to inter-subject variation. The atlas database was classified into different categories and the most similar atlases in the region of interest are selected for multi-atlas registration by comparing the sum of square intensity distance after affine registration. The selected atlases are non-rigidly aligned to the target image and a patch-based local voxel weighting strategy is introduced. This strategy was recently proposed for use in patch-based brain segmentation \cite{2}. In the final atlas fusion process we use a local weighting of each atlas depending on the mapping agreement from atlas to target. The proposed method was evaluated on the 50 training data, which is a representative set of the types of diagnostic MR images acquired in a clinical setting.

2 Method

In multi-atlas based segmentation, the most similar atlases create more accurate transformations to the target image and therefore provide a better label estimation. In this paper we aim for a robust and accurate segmentation of multi-center and multi-vendor MRI prostate scans. To this end we introduce appearance-specific atlas selections and a patch-based local weighting strategy for atlas fusion. An initial denoising and intensity inhomogeneity correction is performed on all images. Atlases are classified into two categories: normal MRI scans $A_n$ and scans taken with a transrectal coil $A_m$. This is easily achieved by examining the intensity variation around the rectum since the transrectal coil...
produces significant physical distortion but also has a characteristic bright appearance in the local region near the coil itself. The sub-atlas database whose atlas appearance is closest to the new target is chosen as the initial atlas database. After that, the top N similar atlases are further chosen for atlas registration by measuring intensity difference in the region of interest around prostate. After all the selected atlases are non-rigidly registered to a target image, the resulting transformation is used to propagate the anatomical structure labels of the atlas into the space of the target image. Finally, using patch-based voxel weighting, the label that the majority of all warped labels predict for each voxel is used for the final segmentation of the target image. The pipeline of multi-atlas segmentation of the prostate is divided into the following parts: atlas database construction, appearance-specific atlas selection, multi-atlas pairwise registration, and atlas propagation and fusion with local voxel weighting, shown in Figure 4.

![Fig. 1. The pipeline of multi-atlas segmentation of the prostate](image)

2.1 Image preprocessing for atlas database construction

In the training set, 50 transverse T2-weighted MR images of the prostate are provided, which are representative of the types of MR images acquired in clinically for diagnosis. The data is multi-center and multi-vendor and has different acquisition protocols, such as differences in slice thickness and the presence of an endorectal coil in some images. Many of these images exhibit significant intensity inhomogeneity. During atlas database construction, variability caused by image formation is minimized by performing denoising, inhomogeneity correction, and
an inter-subject intensity normalization. To remove the intensity bias introduced by the Rician nature of noise, a Rician adaption of non-local means [11] was used for denoising. The well-known N3 intensity nonuniformity correction [9] was applied to reduce intensity inhomogeneity. Finally, all the atlases are transformed into the template space by affine registration. The intensity of images were then normalized together in the template space using the method proposed by Nyul and Udupa [7]. This makes the contrast and luminance of each tissue type more consistent across the training images in the database. After all these procedures, an atlas database was constructed with the preprocessed MR images and their corresponding manual segmentations, which we represent as \( A(I_i, L_i) \), where \( I_i \) represents a MR image, \( L_i \) is a segmentation.

2.2 Local appearance-specific atlas selection

In the atlas database, the endorectal coil influences the appearance of the scan significantly in both shape and intensity, which often leads to faulty registrations. In order to tackle this problem we compare the intensity difference around the region of rectum and classify the atlas database into two sub-databases, represented as \( A = \{ A(I_i, L_i) \} = \{ \{A_N(I_i, L_i)\}, \{A_M(I_j, L_j)\} \} \): atlases with normal MRI scans \( \{A_N(I_i, L_i)\} \) and atlases taken with an endorectal coil \( \{A_M(I_j, L_j)\} \). The most suitable sub-database will be automatically chosen for a new unseen target image.

Within the chosen sub-database, the top \( N \) atlases with the most similar appearance around the prostate region are selected for final multi-atlas segmentation. Atlas selection in these two steps are base on the L2 norm: the sum of squared intensity differences \( \triangle(A_i, L) \) between atlas \( A_i \) and target image \( L \), defined over a region of interest \( R \), measuring local image appearance:

\[
\triangle(L, A_i) = \sum_{j \in N} ||L(x_j) - A_i(x_j)||^2 \tag{1}
\]

2.3 Patch-based voxel weighting

During the multi-atlas fusion process, would like the atlases whose reference images are more similar to the target image to contribute more to the segmentation. Also, the accuracy of the transformation from atlas to target is crucial for accurate label estimations. In this paper, we propose a more robust and improved weighting strategy for atlas label fusion that combines a mapping agreement weighting with the patch-based weighting for each voxel label. This weight is based on the similarity of a patch surrounding voxel \( x_i \) and patches in a local neighborhood of all non-rigid aligned atlas images \( A_s \). The segmentation problem proceeds as follows:

\[
\Gamma(x_i) = \frac{\sum_{s=1}^{N} w_{T(A_s \rightarrow L)} \sum_{j \in V} \frac{w(x_i, x_{s,j}) L_{s,j}}{\sum_{s=1}^{N} w_{T(A_s \rightarrow L)} \sum_{j \in V} w(x_i, x_{s,j})}}{\sum_{s=1}^{N} w_{T(A_s \rightarrow L)} \sum_{j \in V} w(x_i, x_{s,j})} \tag{2}
\]
where $I_i$ is the estimated label for voxel $i$, $L_{s,j}$ is the label from the expert for voxel $x_j$ in atlas $s$, $V$ is the search window size, $w(x_i, x_{s,j})$ is the weight assigned to label $L_{s,j}$ by comparing the patch surrounding $x_i$ to that surrounding $x_{s,j}$, as follows:

$$w(x_i, x_{s,j}) = \exp\left(\frac{-\Delta(P_{L_i}, P_{A_{s,j}})}{h}\right)$$  \hspace{1cm} (3)

where $h$ is a decay parameter, which is set to the minimum patch distance as proposed in [2].

The other weight $w_{T(A_s \rightarrow \mathcal{L})}$ is defined by the accuracy of the non-rigid mapping from each atlas to the target image. This is measured by comparing the intensity between the warped atlas $T(A_s \rightarrow \mathcal{L})A_s^L$ and target image $\mathcal{L}^T$ under the segmented label region using normalized mutual information, denoted as:

$$w_{T(A_s \rightarrow \mathcal{L})} = NMI(T(A_s \rightarrow \mathcal{L})A_s^L, \mathcal{L}^T)$$  \hspace{1cm} (4)

It is a crucial term to reduce the negative influence of mis-registration in local regions of the image and makes final segmentation more robust.

3 Experiments and Results

The whole segmentation pipeline is implemented in C++ and CUDA with a quad 3.20 GHz CPUs and a graphic card with 96 CUDA cores and 1GB global memory, using parallel programming for the non-rigid registration and calculation of the patch weighting map for each voxel simultaneously. The proposed method was evaluated on the 50 training MRI images, which are transverse (axial) T2-weighted MR images of the prostate. For each training image, manual segmentation is provided.

A leave-one-out study has been implemented based on each of the training scans using the remaining 49 images as the atlas database. In the sub-database, the top 10 most similar atlases are chosen by comparing the local prostate appearance according to equation 1. With these selected atlases, a pairwise rigid and affine registration was applied to transform atlas images to the target image in the template space, followed by a fast three-level free-form deformation based non-rigid registration using graphic processing units (GPUs) [8] [6]. We use b-spline control point spacings of 20 mm, 10mm and 5mm for our three levels.

In the multi-atlas fusion, the patch-based voxel weighting strategy in Eq.2 is applied to get the final segmentation estimation, with a patch size of $5 \times 5 \times 5$ and search window size $9 \times 9 \times 9$.

For evaluation, we used the following metric compared with gold standard expert segmentation. We also compared our propose method with atlas fusion using global weighting based on normalized mutual information between target image and atlases:

1) Dice Coefficient (DC) : $DC = \frac{2|A \cap B|}{|A| + |B|}$.

2) Volumetric overlapping accuracy (VOA): $VOA = \min\left(\frac{|A \cap B|}{|A| \times 100\%}, \frac{|A \cap B|}{|B| \times 100\%}\right)$.
Table 1 shows the average value of these metrics with their standard deviation among all the 50 MRI scan segmentations compared with manual segmentation. Final segmentation results for the prostate from MRI images exhibit an average DC of $0.8467 \pm 0.0435$ and an average VOA of $0.8259 \pm 0.0630$. The algorithm is fully automated and most training data can be well segmented, but there are still a few examples where the segmentation is poor, as shown in Figure 2. These rare cases reduce the average value of the Dice coefficient significantly.

<table>
<thead>
<tr>
<th>Segmentation Method</th>
<th>DC (average+sd) [%]</th>
<th>VOA (average+sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global weighting based atlas fusion</td>
<td>0.82318 ± 0.0456</td>
<td>0.8015 ± 0.0555</td>
</tr>
<tr>
<td>Our proposed method</td>
<td>0.8467 ± 0.0435</td>
<td>0.8259 ± 0.0630</td>
</tr>
</tbody>
</table>

Table 1. Average metric compared with gold standard, including: volumetric overlapping accuracy (VOA) and Dice coefficient (DC).

As can be seen in Table 1, the atlas fusion using patch-based voxel weighting outperforms the method that directly applies a global weighting strategy, with higher Dice metric and VOA. Though the increase in the Dice metric is not huge, the segmentation is significantly better, with smoother boundaries of the segmented prostate surface, as can be seen in Figure 3.

4 Discussion and Conclusion

In this paper, we propose a novel and automatic multi-atlas segmentation using local appearance-specific atlases and patch-based local weighting. Among the 50 training data with diverse intensity variation and prostate shape, most of the segmentations performed well and the method is fully automatic. However, for some of the segmentation estimations, a small segmentation error will cause
significant Dice coefficient change, since there is only a small part of the prostate structure contained in the MRI scan. This makes registration very difficult and there are matches to the wrong structures, which makes the average Dice metric significantly smaller. Results are much better for images of the whole prostate and it is to be expected that most diagnostic prostate images will cover at least the majority of the prostate itself since this is the target organ of the scan. Another issue is that for some patients, the prostates are very large, probably due to pathological growth. Matching the atlas images to the patient image is thus more difficult. In order to tackle this problem, landmarks could be introduced manually to get a better initial transformation.

The run-time of our algorithm is affected by the number of atlases used, which needs more computation time to get the affine and non-rigid transformation from atlas space to target space. In our experiment, 10 atlases were used, which accounts for around 30 minutes overall run-time for one segmentation, including the data preprocessing step. This is a quite reasonable processing time for use in the clinical environment.

There is still some work to be done to achieve more accurate segmentation. Patch learning could also be introduced to make more convincing weighting of each voxel during the fusion process. We are also investigating methods to improve registration in scans where there is only small overlap with the region of interest.

Acknowledgments

This research was funded by Cancer Research UK under project A8087/C24250. We are grateful to the radiology and theatre staff at the Imperial College Healthcare NHS Trust for their help and cooperation throughout this project. The work has ethical approval from the London - Dulwich research ethics committee.
Bibliography


