Fully Automatic Segmentation of the Prostate using Active Appearance Models

Graham Vincent, Gwenael Guillard, and Mike Bowes

Imorphics Ltd., Kilburn House, Manchester Science Park, Manchester, M15 6SE, UK.

Abstract. We present a fully automatic model based system for segmenting the prostate in magnetic resonance (MR) images. The segmentation method is based on Active Appearance Models (AAM) built from manually segmented examples provided by the MICCAI 2012 Promise12 team. High quality correspondences for the model are generated using a Minimum Description Length (MDL) Groupwise Image Registration method. A multi start optimisation scheme is used to robustly match the model to new images.
The model has been cross validated on the training data to a good degree of accuracy, and successfully segmented all the test data.

1 Introduction

Prostate diseases such as prostate cancer, prostatitis and benign prostate hyperplasia (BPH) are common in afflictions in men. Prostate cancer is the third most common cancer worldwide, and (due to a huge increase in screening) is now the most commonly diagnosed cancer in men in the US [4].

Several diagnostic and therapeutic procedures could benefit from accurate estimates of the volume and boundary of the prostate. In particular accurate prostate volume measurement combined with prostate-specific antigen level is thought to be a good indicator of treatment outcome [9], and robust measurements of the prostate boundary is an important component of radiotherapy planning [5]

It is well known that prostate segmentation in MR is challenging, as prostates show a wide range of shape variation, and the prostate and neighbouring blood vessels, bladder, urethra, rectum and seminal vessels vary considerably in image response. and have complex intensity distributions. Furthermore MR suffers from distortions due to field inhomogeneity. Prior information on shape and intensity is essential.

Here we have used a statistical modelling pipeline (developed for the segmentation of the knee [12]) to build a model of the total prostate. The model fitting is based on the Active Appearance Models (AAMs) of Cootes *et al*, in which the statistics of shape and image information, and the correlations between them, are calculated from a training set of images. Variants of AAMs have been extensively developed (see [3] for a review).

The model is built using 50 transverse T2-weighted MR images and voxel-based segmentations provided by the Promise12 organisers. The model was evaluated using a Leave-One-Out Cross Validation (LOOCV) and by comparing the automated segmentation to the reference segmentation. The organisers also provided 30 test cases without reference segmentations. The model was used to segment these test cases and the results visually validated.

Several authors have developed segmentation methods for total prostate which use statistical shape or appearance models as part of the process.

Toth *et al* [11] use an extension to Active Shape Models in which the models are trained on multiple features for each voxel based on combinations of Gaussian feature responses, the features being selected during training for their discrimination between prostate border and background.

Martin *et al* [7] uses a probabilistic atlas registered to each example followed by a 3D deformable model driven by image features and the probabilistic priors from the registered atlas,

Other authors have used classification of voxel based descriptors. Moschidis and Graham [8] report on the differential segmentation of peripheral, transitional and fibro-muscular zones, as well as the total prostate. Moschidis uses random forest classification on a feature space consisting of intensities, spatial position, Haralick and Laws features, regularised by a graph cut in which the regional terms are provided by the classifiers and the boundary term based on edge strength.

2 Methods

2.1 Converting voxel-based segmentations to surfaces

Our statistical models are surface based, so a marching-cube algorithm ([6]) was used to generate surfaces from the voxel-based segmentations which have 1 inside the prostate, 0 outside.

2.2 Generating surface correspondences

Statistical appearance models rely on a large set of anatomically equivalent landmarks (also known as *correspondences*) across the region of interest. Generating good quality correspondences is key to developing generalisable yet specific models.

To obtain the anatomical correspondences on the prostate surfaces we used a variant of the Minimum Description Length approach to Groupwise Image Registration (MDL-GIR) of Cootes *et al* [2]. The MDL-GIR method finds the set of deformations which register all the images together as efficiently as possible. This idea is made concrete by the use of Information Theory to define the amount of information required to encode a model using a particular set of deformations. The method is an optimisation to find the set of deformations requiring the least amount of information to encode. The output is a reference mean image and a set of deformations which map the mean image to each example image.

We apply the MDL-GIR method to the signed distance images derived from the segmented surfaces. The output reference mean image is, like the input images, a signed distance image and can be straight forwardly segmented using the zero valued iso-surface. The mean surface is then propagated by the appropriate deformation field into the frame of each example. For each example the propagated surface lies close to the segmented surface and is projected onto it to generate correspondence points which are guaranteed to lie on the segmented surface.

The number of correspondence points output from this process is approximately 7000.

2.3 Active appearance models

An appearance model is a statistical model of the shape of a structure and associated imaging information. It is useful to process the imaging information further to obtain feature response images such as gradients, corners and other points of interest [10]. We refer to all such imaging information and their derivatives as texture.

An appearance model has a set of parameters which control both the shape and the texture, and are *generative* i.e. a specific parameterisation can generate a realistic looking example of the shape and texture.

An AAM can match its appearance model an image from a rough initial estimate, by optimising the model parameters to generate an example which matches the image as closely as possible (using the least squares sum of residuals). This can be made efficient by pre-computing the Jacobian describing the average change in residuals with respect to changes in model parameters on a training set.

AAMs require an initial estimate of the model parameters including position, rotation and scale. We provide multiple initial estimates at a grid of starting points across the image. The grid of starting points are typically 20mm apart in all directions. This is done at a low image and model resolution with a small number of measured residuals to make it reasonably fast. The results of these searches are ranked according to the sum of squares of the residual, and a proportion (typically 75%) removed from consideration. The remaining search results are used to initialise models at a higher resolution, and so on. Finally, the single best result at the highest resolution gives the segmentation result.

2.4 Segmentation pipeline

In summary, the segmentation proceeds according to the following pipeline:

- For each image:
  - Run N (typically 10s-100s) AAMs of the prostate from a grid of starting positions across the image at low resolution.
    - Run the 25% best results at increased resolution
    - Repeat until at highest resolution
  - Choose best result

2.5 Experiments

The model is built from 50 transverse T2-weighted MR images and voxel based segmentations provided by the Promise12 organisers. These images are multi-centre, multi-vendor and multi-protocol and as a consequence are challenging: there is variation in voxel size, dynamic range, position, field of view and anatomic appearance.

To test the performance of the algorithm we use the standard Leave-One-Out Cross Validation (LOOCV) technique where each image i in turn is segmented using a model built from the training set with image i removed and then compared against the reference segmentation.

A variety of statistics are reported in the literature: the most common are the Dice Similarity Coefficient (DSC), Mean Absolute Distance (MAD) and the 95th% Hausdorff Distance (95%HD) [1]. The DSC measures the amount of overlap between the reference segmentation and the automated segmentation. DSC can range from zero to one, where zero represents no overlap and one corresponds to identical segmentations. The directed Hausdorff Distance (HD) identifies the point on the reference segmentation that is the farthest from any point on the model segmentation and measures the distance from this point to the nearest point on the model segmentation. The 95%HD is less sensitive to outliers than HD since it considers the point representing the 95th percentile of the distances instead of the farthest. Although these measures are quite standard, they can vary slightly in their definition and the lack of an exact definition or of a reference in a paper make any comparisons difficult. This is especially true for HD since a directed and a symmetrical versions of it exist and authors rarely specify which one they use.

3 Results and Discussion

In this section we present the results of the LOOCV on the training datasets provided for the Promise12 Grand Challenge and made a comparison with recent papers in the literature (see Table 1). We note that this is not a comprehensive review, and that comparisons should always be considered tentative, as datasets can vary in consistency - which is of course part of the motivation for the Grand Challenges.

Table 1. A comparison between our method and recent papers in the literature, where possible.

	mean DSC	median DSC	MAD (mm)	95th%HD (mm)
Our fully automated AAM	0.88 (0.03)	0.89	1.44 (0.48)	4.17 (1.35)
Martin *et al* [7]		0.87	2.41	
Toth *et al* [11]	0.85 (0.05)			
Moschidis *et al* [8]			1.0 (0.1)	

As the DSC is an overall measure that does not give any spatial information, we also generate a colour-coded error surface. One benefit of our method is that the training 3D pointsets and the automatic segmentation result are anatomically corresponded thus providing thousands of registered 3D landmarks. We are then able to generate statistics on a per-point basis, in this case, the distance error. For each 3D point and for each segmentation, the Euclidean distance of the point to the nearest point in the reference segmentation pointset is computed and represents the error. Then the mean error across the dataset is evaluated for this corresponded point. To display the results, we compute a mean prostate corresponded pointset from the shape component of the AAM and we colour each point according to its mean error.

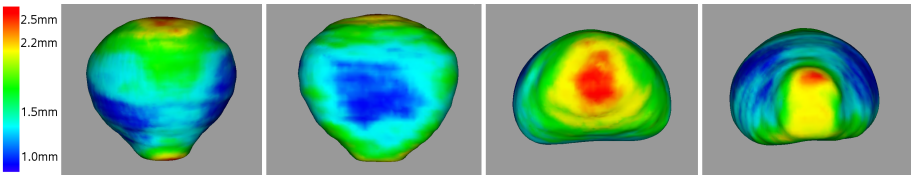

Fig. 1. *Map of the mean distance error projected. From left to right: anterior, posterior, apex, base.*

The spatial distribution of the errors (Fig. 1) shows that our model produces slightly less good results at the apex and the base of the prostate. The lack of accuracy at the top and the bottom can be explained by the fact that these areas present a quite continuous change of tissue without strong edge or texture. We are confident that including more structures such as the bladder in the model would significantly improve the results. The anterior apex is surrounded by tissues that have texture with very variable edges (Fig. 2), parallel to the prostate boundary and that are often even stronger that the boundary itself. Incorporating structured texture detectors into our model may help improve the accuracy in these regions.

It was extremely encouraging that the prostate model worked well "out of the box". The model segmented every test case completely automatically with DSC scores similar or better than those reported in the literature (see Table 4).

The model was also used on test images provided by the MICCAI 2012 Promise12 team. These images present the same characteristics than the training data. Since no segmentation is available, we only performed a visual check and considered that the results were very similar to the LOOCV experiment.

For a practical application it might be acceptable to allow some user interaction (e.g. an initial localisation) to improve speed and accuracy. However our focus here was to achieve genuine automation. This is not necessarily just a matter of more compute power: very specific models are needed to avoid hallu-

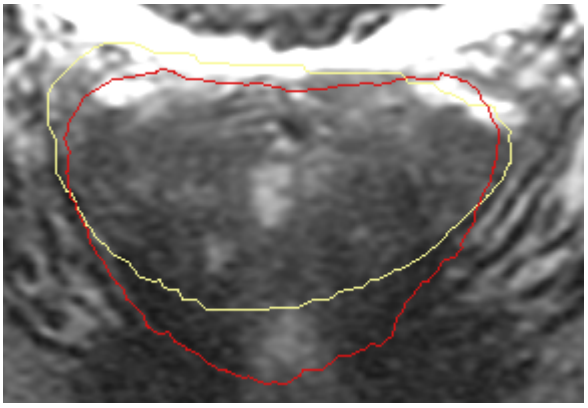

Fig. 2. *Representative slice of the top front of the prostate with highly textured surrounding tissues. In yellow, reference segmentation ; in red, automated segmentation.*

cinating in other areas of the image whilst ensuring the models are generalisable enough to match to new data.

4 System details

Language	C++
External Libraries	VXL
GPU Optimisation	None
Multi-threaded?	No
User interaction?	None - fully automatic
CPU clock speed	2.83 GHz
Machine CPU count	4
Machine memory	8GB
Memory used during segmentation	300MB
Training time	2 hours
Segmentation time per image	8 minutes
User interaction time	0 seconds

5 Conclusions

In this paper we have presented a fully automatic AAM based model to segment the prostate from T2 weighted MR images, built and cross validated on the Promise12 training data, and which successfully and automatically segmented the Promise12 test data.

References

1. Neculai Archip, Olivier Clatz, Stephen Whalen, Dan Kacher, Andriy Fedorov, Andriy Kot, Nikos Chrisochoides, Ferenc Jolesz, Alexandra Golby, Peter Black, and Simon Warfield. Non-rigid alignment of preoperative mri, fmri, and dt-mri with intra-operative mri for enhanced visualization and navigation in image-guided neurosurgery. *NeuroImage*, 35(2):609–24, April 2007.
2. T. F. Cootes, C. V. Petrovi, R. Schestowitz, and C. Taylor. Groupwise construction of appearance models using piece-wise affine deformations. *16th British Machine Vision Conference. Volume*, 2:879–888, 2005.
3. Tobias Heimann and Hans-Peter Meinzer. Statistical shape models for 3d medical image segmentation: A review. *Medical Image Analysis*, 13(4):543–563, 2009.
4. Pete A. Humphrey. *Prostate Pathology.* American Society for Clinical Pathology., 2003.
5. Dominique P. Huyskens, Philippe Maingon, Luc Vanuytsel, Vincent Remouchamps, Tom Roques, Bernard Dubray, Benjamin Haas, Patrik Kunz, Thomas Coradi, Ren Bhlman, Robin Reddick, Ann Van Esch, and Emile Salamon. A qualitative and a quantitative analysis of an auto-segmentation module for prostate cancer. *Radiotherapy and Oncology*, 90(3):337 – 345, 2009.
6. William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. *SIGGRAPH Comput. Graph.*, 21(4):163–169, August 1987.
7. Sébastien Martin, Vincent Daanen, and Jocelyne Troccaz. Automated Segmentation of the Prostate in 3D MR Images Using a Probabilistic Atlas and a Spatially Constrained Deformable Model. *Medical Physics*, page 1, March 2010.
8. Emmanouil Moschidis and Jim Graham. Automatic differential segmentation of the prostate in 3-d mri using random forest classification and graph-cuts optimization. In *ISBI*, pages 1472–1475. IEEE, 2012.
9. C. G. Roehrborn, P. Boyle, D. Bergner, T. Gray, M. Gittelman, T. Shown, A. Melman, R. B. Bracken, R. deVere White, A. Taylor, D. Wang, and J. Waldstreicher. Serum prostate-specific antigen and prostate volume predict long-term changes in symptoms and flow rate: results of a four-year, randomized trial comparing finasteride versus placebo. pless study group. *Urology*, 54(4):662–9, 1999.
10. I. M. Scott, T. F. Cootes, and C. J. Taylor. Improving appearance model matching using local image structure. In *In Information Processing in Medical Imaging, 18 th International Conference*, pages 258–269. Springer, 2003.
11. R. Toth, B. N. Bloch, E. Genega, N. Rofsky, R. Lenkinski, M Rosen, A Kalyanpur, S Pungavkar, and A Madabhushi. Accurate prostate volume estimation using multifeature active shape models on t2-weighted mri. *Academic Radiology*, 18(6):745 – 754, 2011-06 2011.
12. Graham Vincent, Chris Wolstenholme, Ian Scott, and Mike Bowes. Fully automatic segmentation of the knee joint using active appearance models. *Proceedings of MICCAI 2010, Grand Challenge Workshop*, 2010.