Prostate MR image segmentation using 3D Active Appearance Models

Bianca Maan and Ferdi van der Heijden

MIRA Institute for Biomedical Technology and Technical Medicine, University of Twente, the Netherlands;

Abstract. This paper presents a method for automatic segmentation of the prostate from transversal T2-weighted images based on 3D Active Appearance Models (AAM). The algorithm consist of two stages. Firstly, Shape Context based non-rigid surface registration of the manual segmented images is used to obtain the point correspondence between the given training cases. Subsequently, an AAM is used to segment the prostate on 50 training cases. The method is evaluated using a 5-fold cross validation over 5 repetitions. The mean Dice similarity coefficient and 95% Hausdorff distance are 0.78 and 7.32 mm respectively.

1 Background

Prostate segmentation is essential for calculating prostate volume, image fusion, creating patient-specific prostate anatomical models, and as a pre-processing step for many computer aided diagnosis algorithm. Furthermore, information about the size, volume, shape and location of the prostate relative to adjacent organs is an essential part of planning for minimally invasive therapies and biopsies.

Because manual segmentation of the prostate is time-consuming and highly subjective, (semi)-automatic segmentation methods are preferable. However, segmenting the prostate in MR images is challenging due to the large variations of prostate shape between subjects, the lack of clear prostate boundaries and the similar intensity profiles of the prostate and surrounding tissues.

The 2012 MICCAI challenge: "Prostate MR Image Segmentation" involves segmentation of the prostate on transversal T2-weighted images. The goal of the challenge is to evaluate segmentation algorithms on images from multiple centers and multiple MRI device vendors.

Only a few prostate segmentation methods for T2-weighted MR images currently exist. Klein et al. [1] proposed a method based on non-rigid registration of a set of pre-labeled atlas images, against the target patients image, using mutual information. Subsequently, the segmentation is obtained as the average of the best matched registered atlas sets. Multiple modifications are published on this atlas based prostate segmentation method [2–4].

The methods presented by Toth et al. [5] and Ghose et al. [6, 7] are based on statistical shape models. Toth et al. used a levelset-based statistical shape

model – which is landmark free and fully 3D – in combination with an statistical appearance model. In order to segment the prostate on a new image, a Bayesian classifier is first employed to pre-classify the image voxels as belonging to the prostate or the surroundings. Subsequently the shape model is fitted to the segmentation followed by an adaptive update of the statistical appearance model until convergence. Ghose et al. presented several segmentation methods based on 2D active appearance models (AAM). Active appearance models include both shape and appearance of the prostate. In [6], they presented a texture enhanced 2D AAM. In [7] they applied 2D AAM with a 3D shape restriction imposed by rigidly registering the obtained volume to a 3D average model of the prostate [7].

It has been demonstrated that the combination of both shape and intensity prior information improves the segmentation accuracy. We present a novel framework for automated prostate segmentation based on 3D active appearance models (AAM) in combination with Shape Context based non-rigid surface registration to obtain a point distribution model.

2 Methodology

Our proposed method is built on the work presented by Kroon et al. [8]. They used a Shape Context based non-rigid surface registration in combination with Active Shape Models to segment knee cartilage. We adapted the framework to use the Shape Context registration with AAM to include the texture of the prostate.

An overview of the proposed system is shown in Figure 1 and consists of a training and a testing part. Active Appearance Models (AAM) contain a statistical model of the shape and grey-level appearance of the organ which can generalise to almost any valid example. AAM learn what are valid shape and intensity variations from their training images. During the testing part, the model is fitted to a new image whereafter the results for the segmentation are compared with the provided ground truth segmentation.

2.1 Shape context registration

The first step in AAM training is describing the prostate surface in each training case by a set of n landmark points. Every landmark in a training case must have a corresponding landmark in all other training case. These point correspondences allow Principle Component Analysis (PCA) to extract the principal modes of shape variations.

We want to use the vertices of the prostate surfaces to train the AAM. Therefore all surfaces must contain an equal number of vertices, and every vertex must have a corresponding vertex at relatively the same location in all training cases. To obtain the corresponding points we use Shape Context based non-rigid registration of the binary segmentation surfaces. The method is described in detail in [8] and [9], we will describe it briefly.
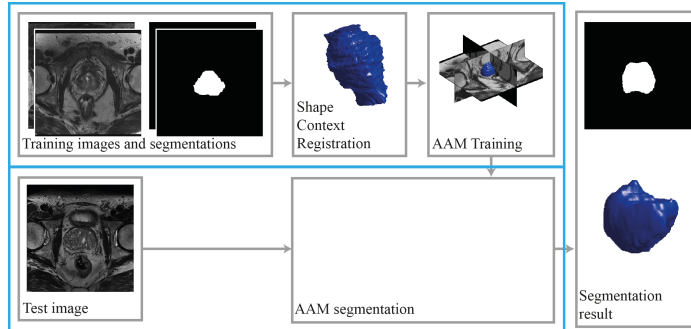
Fig. 1. Schematic overview of our prostate segmentation approach. The upper blue box shows the training of the AAM. The lower blue box shows the Test part of the AAM.

The method consists of four steps: the first step is extracting the surface of the segmented prostate. The second step is constructing a descriptor, the *shape context*, for each point in both data sets. The Shape Context describes every point on the surface by the vectors originating from that point to all other points on the surface. For every point on the surface, $n-1$ vectors are obtained. The full Shape Context representation is too detailed. Therefore, for every point p_i on the surface, a coarse histogram of the relative coordinates of the remaining $n-1$ points is defined to be the shape context.

The third step is calculating the matrix $C_{i,j} = C(p_i, q_j)$ which contains the costs between each pair of points. Considering points p_i and point q_j with histograms h_i and h_j respectively, the costs can be calculated as:

$$C(p_i, q_j) = \frac{1}{2} \sum_{k=1}^{K} \frac{[h_i(k) - h_j(k)]^2}{[h_i(k) + h_j(k)]} \tag{1}$$

The fourth step is matching the points by minimizing the total matching cost.

The total Shape Context based non-rigid registration works as follows [8]: One prostate surface is selected as reference surface. This surface is registered to a surface of another training case. Subsequently the Shape Context is calculated for all points in both shapes (reference and training). 3D Shape Context is used [8,10], which is an extension to the original 2D Shape Context [9]. For every point on the reference surface, we search for a point in the training surface with the minimum cost, and also from the training surface to the reference surface. The obtained point correspondences are then used to construct a B-spline grid which warps the reference surface to the shape of the training surface. We repeat this process –using a coarse to fine grid– until the reference surface is transformed into the shape of the training surface. In some regions, the transformed surface points can still deviate from the real boundary by one pixel. Therefore, the algorithm uses a modified iterative closest point (ICP) algorithm to calculate

the vectors from the vertices to the closest surface. Using these vectors the final B-spline grid is updated.

2.2   3D Active Appearance Model

After obtaining the corresponding points, all point sets are aligned into a common reference frame. Let each point set be represented by a vector $\mathbf{x}$. Then principle component analysis (PCA) is applied to determine the principal modes of the shape variations.

$$
\begin{aligned}
\mathbf{x} &= [x_1, x_2, ..., x_n, y_1, y_2, ..., y_n, z_1, z_2, ..., z_n]^T \\
\tilde{\mathbf{x}} &= \bar{\mathbf{x}} + \mathbf{\Phi_s b_s}
\end{aligned}
\tag{2}
$$

with $\tilde{\mathbf{x}}$ an estimate of the shape, $\bar{\mathbf{x}}$ the mean shape, $\mathbf{\Phi_s}$ the t eigenvectors corresponding to the largest eigenvalues and $\mathbf{b_s}$ a set of shape deformation parameters [11].

The appearance model can be described in a similar way. First we warp each training image –using a piece-wise affine warp– so that its points correspond to the mean shape points. Then we sample the grey-level information of the region covered by the mean shape. After applying PCA to the normalised appearance data we obtain the model

$$
\begin{aligned}
\mathbf{g} &= [g_1, g_2, ..., g_m]^T \\
\tilde{\mathbf{g}} &= \bar{\mathbf{g}} + \mathbf{\Phi_g b_g}
\end{aligned}
\tag{3}
$$

with $\mathbf{g}$ representing the appearance vector of m voxels, $\tilde{\mathbf{g}}$ an estimation of the grey-level appearance, $\bar{\mathbf{g}}$ the mean grey-level appearance, $\mathbf{\Phi}_g$ the t eigenvectors and $\mathbf{b_g}$ a set of appearance variation parameters.

Shape and appearance are often correlated, therefore we can combine both models and apply a third PCA:

$$
\begin{aligned}
\mathbf{b} &= \begin{pmatrix} W_\mathbf{s} \mathbf{b_s} \\ \mathbf{b_g} \end{pmatrix} = \begin{pmatrix} W_\mathbf{s} \mathbf{\Phi_s}^T (\mathbf{x} - \bar{\mathbf{x}}) \\ \mathbf{\Phi_g}^T (\mathbf{g} - \bar{\mathbf{g}}) \end{pmatrix} \\
\mathbf{b_c} &= \mathbf{Qc}
\end{aligned}
\tag{4}
$$

where W_s is a diagonal matrix of weights for each shape parameter, allowing for the difference in units (distance vs. intensity) between the shape and grey models. $\mathbf{Q}$ are the eigenvectors and $\mathbf{c}$ is a vector of appearance parameters controlling both shape and grey-level appearance. Now the shape and grey-levels can be expressed as function of $\mathbf{c}$ using

$$
\begin{aligned}
\mathbf{x} &= \bar{\mathbf{x}} + \mathbf{\Phi_s} W_\mathbf{s}^{-1} \mathbf{Q_s c} \\
\mathbf{g} &= \bar{\mathbf{g}} + \mathbf{\Phi_g Q_g c}.
\end{aligned}
\tag{5}
$$

The AAM is optimized minimizing the difference between the test image and the synthesized images [11]. Since the appearance models can have many parameters, the fitting to a new image appears to be a difficult optimisation

problem. However, every optimisation problem is in fact a similar optimisation problem. Therefore, prior knowledge about how to adjust the model parameters during image search is acquired by perturbing the combined model and the pose parameters with some known offsets, and recording the corresponding changes in the appearance.

The mean model is initialized by manual selecting the center of the prostate based on visual inspection. Subsequently, the AAM is performed using a two resolutions with each 15 iterations.

3 Experimental Design

3.1 MR Images

The training data comprise 50 anonymized patient cases which include transversal T2-weighted MR images of the prostate. The training data is multi-center and multi-vendor, and has different acquisition protocols. This means that there is a difference in slice thickness, slice orientation, images with/without endorectal coil etcetera.

Furthermore, a reference segmentation of the prostate is provided for each case in the training data. The images are downloaded in .mhd format from the MICCAI challenge website: http://promise12.grand-challenge.org/

Pre-processing The scans are not uniformly sampled, therefore we re-sample the scans and segmentations to a voxel resolution of 0.5 x 0.5 x 0.5 mm, using cubic interpolation for the MRI data and nearest neighbor interpolation for the label data. The orientation of the slices differ per case in the training data. Therefore the images are rotated to obtain images with equal orientations.

3.2 Experiments

For evaluation of our proposed method we perform a 5-fold cross-validation over 5 repetitions. During one 5-fold cross-validation experiment, the training data is randomly partitioned into five sub sets. Of the five sub sets, four sub sets are used as training data, while the remaining sub set is retained as testing data for the model. The process is then repeated five times, with each of the sub sets used exactly once as test data.

Furthermore, we splitted the training cases in two groups; the first group containing the cases with endorectal coil, the second group containing the cases without endorectal coil. To evaluate our algorithm using two seperate models, we performed a 3-fold cross validation over 5 repetitions on these two groups.

The segmentations are evaluated using the Dice coefficient (DSC) and the 95% Haussdorf distance. Moreover, slices of MR Images with their segmentation are shown providing a qualitative evaluation.

4 Results and Discussion

Table 4 shows the mean and median DSC and 95% Hausdorff distance evaluated using a 5 fold cross validation over 5 repetitions. It is generally accepted that a DSC value > 0.7 represents a good agreement. The mean DSC overlap for our segmentation was 0.78 (sd = 0.12, calculated over all cases and all iterations), with a median DSC of 0.82, which confirms overall satisfactory results of the segmentation. However, the algorithm was unable to segment the prostate in case 23 which contains a very large prostate that exceeds the boundaries of the MRI image. Discarding this outlier result, a mean DSC of 0.80 (sd=0.08) is obtained.

The presence of an endorectal coil has influence on the prostate shape and image intensities. Therefore we decided to train two separate models for the cases with and without coil. The results in Table 4 show that the mean DSC is increased to 0.81 (sd = 0.12), with a median DSC of 0.83. From this results we can conclude that training the AAM on images per institute, and per imaging protocol, will slightly improve the segmentation results. However, a generally applicable segmentation method is preferable. Therefore we segment the test cases with one AAM instead of two separate models.

Although our DSC values are slightly higher than the values reported by Dowling et al. [3] (mean: 0.73) and Gubern-Merida et al. [4] (median: 0.79), they are lower compared to several other articles [1, 5–7, 12]. However, it is to be noted that they mostly use training and testing data from only one institute, containing less variation.

The mean 95% HD value for all datasets is 7.32 mm (sd = 4.91) with a median value of 6.18 mm. As with the DSC values, splitting the model in two parts (with and without coil) improves the HD values to 6.43 mm (sd = 4.63) with median 5.19 mm. Our HD values are lower than the values reported by Gubern-Merida et al. [4].

Table 1. Table with mean evaluation values over all training cases and five iterations.

	One shape model	Two shape models
mean DSC	0.78 ± 0.12	0.81 ± 0.12
median DSC	0.82	0.83
mean 95% HD [mm]	7.32 ± 4.91	6.43 ± 4.63
median 95% HD [mm]	6.18	5.19

Figure 2 shows a surface mesh generated from the segmentation result of case 19 (DSC = 0.84) with a colormap indicating the vertex distance to the ground truth. Visual inspection of these meshes for all cases indicate that the region with the highest segmentation error is not consistent. Figure 3 shows 4 slices from different prostates (with DSC > 0.7) with the ground truth contour in red and the segmentation result in green. The automatic segmentation contours show good agreement with the manual ground truth segmentations.
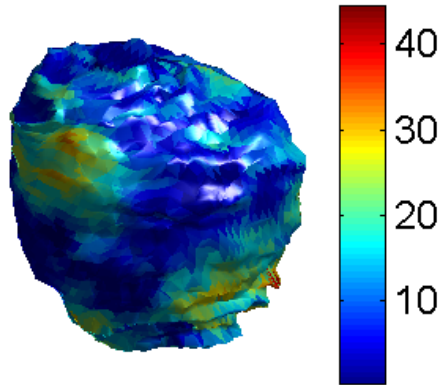
Fig. 2. Surface mesh from case 30. The colormap indicates the vertex distance (in mm) between the segmentation result and the ground truth.
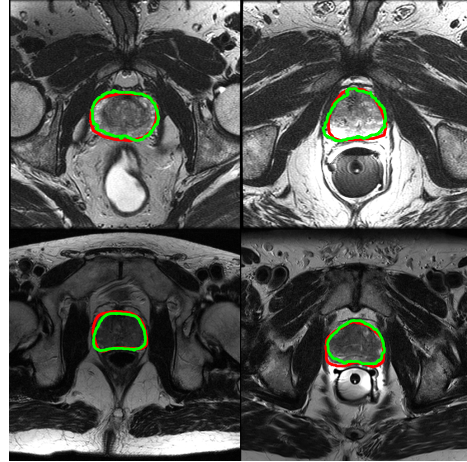
Fig. 3. Segmentation result for case 01, 11 (top), 27 and 46 (bottom). The red line represents the ground truth outline while the green line represents the segmentation result.

The automatic segmentation required approximately 3 minutes on a standard desktop PC (Intel Core(TM) i7 CPU 870 @ 2.93GHz). See Table 2 for more information about the efficiency of our algorithm.

Table 2. Details about the algorithm.

	Parameter	Value
Algorithm	*Language:*	Matlab 7.13.0.564
	Libraries/Packages:	Image Analysis Toolkit; 2 packages from Mathworks File exchange [10,13]
	GPU Optimizations:	-
	Multi-Threaded:	Yes
	User Interaction:	(1) Selection of the middle of the prostate
Machine	*CPU Clock Speed:*	2.93 GHz
	Machine CPU Count:	4
	Machine Memory:	16 GB
	Memory Used During Segmentation:	± 2 GB
Time	*Preprocessing time:*	± 2 minutes (per case)
	Shape Context registration:	± 3.5 minutes (per case)
	Training Time:	± 3 hours (for 50 training cases)
	Segmentation Time:	± 1.8 minutes (per case)
	User Interaction Time:	± 50 seconds (per case)

5 Concluding Remarks

Evaluation of our segmentation method has been performed using 50 transversal T2-weighted MR images. A 5-fold cross validation over 5 repetitions is performed giving a mean DSC of 0.78 with median 0.82. The 95% HD and visual inspection of the results also show good agreement of the prostate segmentation.

The method could be further improved. For example by adding texture information to the AAM [6], adding additional preprocessing steps such as removing the bias field, and using several initial pose parameters during the initialisation of the model.

References

1. Klein, S., van der Heide, U.a., Lips, I.M., van Vulpen, M., Staring, M., Pluim, J.P.W.: Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. Medical Physics **35**(4) (2008) 1407
2. Martin, S., Daanen, V., Troccaz, J.: Atlas-based prostate segmentation using an hybrid registration. International Journal of Computer Assisted Radiology and Surgery **3**(6) (September 2008) 485–492
3. Dowling, J., Fripp, J., Greer, P., Ourselin, S.: Automatic atlas-based segmentation of the prostate : A MICCAI 2009 Prostate Segmentation Challenge entry (2009)
4. Gubern-merida, A., Marti, R.: Atlas Based Segmentation of the prostate in MR images (2009)
5. Toth, R., Bulman, J., Patel, A.D., Bloch, B.N., Genega, E.M., Rofsky, N.M., Lenkinski, R.E., Madabhushi, A.: Integrating an adaptive region-based appearance model with a landmark-free statistical shape model: application to prostate MRI segmentation. Proceedings of SPIE Medical Imaging: Image processing **7962**(May) (2011) 79622V–79622V–12
6. Ghose, S., Oliver, a., Marti, R., Llado, X., Freixenet, J., Vilanova, J.C., Meriaudeau, F.: Prostate Segmentation with Texture Enhanced Active Appearance Model. 2010 Sixth International Conference on Signal-Image Technology and Internet Based Systems (December 2010) 18–22
7. Ghose, S., Oliver, A., Marti, R., Llado, X., Freixenet, J., Mitra, J., Vilanova, J.C., Meriaudeau, F.: A hybrid framework of multiple active appearance models and global registration for 3D prostate segmentation in MRI. In: SPIE Medical Imaging; Image processing. Volume 8314. (2012) 83140S–1
8. Kroon, D.J., Kowalski, P., Tekieli, W., Reeuwijk, E., Saris, D., Slump, C.H.: MRI based knee cartilage assessment. Proceedings of SPIE Medical Imaging: Computer-Aided Diagnosis **8315**(0) (2012) 83151V–83151V–10
9. Belongie, S., Malik, J.: Shape matching and object recognition using shape contexts. Pattern Analysis and Machine **24**(24) (2002) 509–522
10. Kroon, D.J.: Shape Context Based Corresponding Point Models
11. Cootes, T., Edwards, G.: Active appearance models. Proceedings of European Conference on Computer Vision **2** (1998) 484–498
12. Makni, N., Puech, P., Lopes, R., Dewalle, a.S., Colot, O., Betrouni, N.: Combining a deformable model and a probabilistic framework for an automatic 3D segmentation of prostate on MRI. International journal of computer assisted radiology and surgery **4**(2) (March 2009) 181–8
13. Kroon: Active Shape Model (ASM) and Active Appearance Model (AAM)