# Segmentation of the Prostate in MR Images by Atlas Matching using Localised Mutual Information

S Klein*, U van der Heide, M Staring, A Kotte, B Raaymakers, J Pluim
University Medical Center Utrecht, Utrecht, NL

In this paper, an automatic method for delineating the prostate in MR scans is presented. The method is based on nonrigid registration of a set of prelabelled atlas images. Each atlas image is nonrigidly registered with the target patient image. After that, atlas images that match well to the patient image are selected and the segmentation is obtained by a majority voting rule. Two registration methods are investigated. The first one uses the common mutual information as a similarity measure. The second one uses a localised version of mutual information. Experiments are performed on 38 MR images using a leave-one-out approach. The automatic segmentations are evaluated with manual segmentations by computing their overlap. The localised mutual information measure outperforms the commonly used global version and achieves a median Dice similarity coefficient of 0.82. The spatial distribution of the segmentation errors is visualised using a spherical coordinate mapping of the prostate boundary.

# 1 INTRODUCTION

Prostate cancer treatment by radiation therapy requires an accurate localisation of the prostate: neighbouring tissue (rectum and bladder) should be spared, while the tumour should receive a maximum dose. For the treatment planning, computed tomography (CT) images are primarily used, but increasingly magnetic resonance (MR) images are added, because of their soft-tissue contrast. In current practice at our hospital a manual delineation of the prostate is made, based on the CT and MR scans, which is a labour-intensive task.

In this paper, we investigate an automatic segmentation method based on atlas matching. The atlas consists of a set of manually labelled MR images. Multiple atlas images are used, instead of a single image, to account for the large anatomical variability between subjects and for the differences in bladder and rectum filling. Using a nonrigid registration algorithm, all atlas images are matched to the patient's MR image that is to be segmented. The deformed manual segmentations of the atlas images are combined into a single segmentation of the patient's image.

The nonrigid registration method is based on maximisation of a similarity measure. A commonly used similarity measure is the mutual information [8]. The mutual information of two images is derived from their joint intensity probability density function. It is assumed that this probability density function does not vary over the images' spatial domains. This assumption is violated in the MR scans we use, due to the presence of magnetic field inhomogeneities. Therefore, a localised version of the mutual information, as for example proposed in [4], might be more appropriate. Localised mutual information is defined as a sum of multiple mutual information measures, computed on different, possibly overlapping, image regions. For our application, we compare the performance of the localised mutual information measure with the performance of the common, 'global' mutual information.

A disadvantage of the localised mutual information is its large computation time compared to the global mutual information. For registration of 3D images, a computation time of 30 minutes on a cluster of 24 processors is reported in [4]. To reduce this, we propose a stochastic optimisation procedure, which results in a runtime of about 15 minutes on a single processor.

# 2 METHOD

The segmentation method consists of two stages: 1) registration of all atlas images to the patient image, and 2) selection and combination of the deformed atlas label images. A set of $M$ accurately labelled images, which serve as an atlas, are assumed to be available. The $i$th image in this atlas set is referred to as $A_i(\boldsymbol{x})$. The corresponding label image is called $L_i(\boldsymbol{x})$, a binary image, where 'ones' represent prostate tissue and 'zeros' everything else. The patient's image to be segmented is denoted by $P(\boldsymbol{x})$. The goal of the automatic segmentation method is to produce a label image $\hat{L}_P(\boldsymbol{x})$ that accurately defines the prostate of the patient. Ideally, this label image should be equal to a manual segmentation $L_P(\boldsymbol{x})$ created by a radiation oncologist.

## 2.1 Registration

In the registration stage, each atlas image $A_i$ is matched to the patient image $P$. A coordinate transformation $T_i(\boldsymbol{x})$ is estimated that maximises the similarity of $P$ and the deformed atlas $A_i \circ T_i$. The registration is performed in two steps. Firstly, global pose differences are compensated for by an affine registration. After that a nonrigid registration is performed, using a coordinate transformation that is parameterised by cubic B-splines [7]. The parameters that describe the transformation (so, either the affine matrix elements, or the B-spline coefficients), are represented by the vector $\boldsymbol{\mu}$.

Two similarity measures are compared: mutual information (*MI*) and localised mutual information (*LMI*). The mutual information of two $d$-dimensional images $I(\boldsymbol{x}), J(\boldsymbol{x}) : \Omega \subset \mathbb{R}^d \mapsto \mathbb{R}$ is defined as follows:

$$MI(I, J; \Omega) = \sum_k \sum_m p_{IJ}(k, m) \log \frac{p_{IJ}(k, m)}{p_I(k) p_J(m)}, \tag{1}$$

where $p_I$ and $p_J$ denote the discrete marginal intensity probabilities of $I$ and $J$, respectively, and $p_{IJ}$ represents the discrete joint intensity probability. The intensity probabilities are estimated from a discrete set of intensity pairs $(I(\boldsymbol{x}_i), J(\boldsymbol{x}_i))$, where the coordinates $\boldsymbol{x}_i$ are sampled from the continuous image domain $\Omega$. A common choice is to use all voxel locations, or a uniformly sampled subset of those. It is assumed that the true intensity probabilities do not vary over $\Omega$. When this is assumption is violated, it may be better to evaluate the mutual information on multiple subregions, each having a more stationary intensity distribution. Adding the resulting mutual information values of all subregions gives us the localised mutual information *LMI*:

$$LMI(I, J) = \frac{1}{N} \sum_{\boldsymbol{x}_j} MI(I, J; \mathcal{N}(\boldsymbol{x}_j)). \tag{2}$$

In this equation $\mathcal{N}(\boldsymbol{x}_j) \subseteq \Omega$ represents a spatial neighbourhood centred on $\boldsymbol{x}_j$. The number of neighbourhoods is denoted by $N$. The neighbourhood centre coordinates $\boldsymbol{x}_j$ are samples from $\Omega$. We may choose them to be all voxel locations, or some subset of those. The neighbourhoods $\mathcal{N}(\boldsymbol{x}_j)$ must be chosen large enough to allow for a reliable estimation of the intensity probabilities, but small enough to ensure that the influence of the MR-induced inhomogeneities is negligible. A cubic region of $50 \times 50 \times 50$ mm gave good results in our application. It should be noted that our formulation of localised mutual information is slightly different from that in [4], but the underlying idea is the same. In [4], the contribution of an intensity pair $(I(\boldsymbol{x}_i), J(\boldsymbol{x}_i))$ to $MI(I, J; \mathcal{N}(\boldsymbol{x}_j))$ is weighted by the distance $|\boldsymbol{x}_i - \boldsymbol{x}_j|$, whereas we employ an identical weighting for each pair.

For maximisation of the similarity measure we employ an iterative optimisation routine. The parameters $\boldsymbol{\mu}$ that describe the transformation are updated in each iteration $k$ by taking a step in the direction of the derivative of the similarity measure with respect to $\boldsymbol{\mu}$. In [5] it is demonstrated for *MI*, that, when the derivative is computed using only a very small amount of randomly sampled intensity pairs, convergence to the solution is still achieved. Two important conditions for this are that new samples are selected in every iteration and that the step size is a slowly decaying function of the iteration number $k$. For *LMI* we can use the same strategy and even extend it by using a small set of neighbourhoods, randomly selected in every iteration. In this work we use $N = 1$, which results in equal computational costs per iteration for *MI* and *LMI*. The resulting algorithm is an example of an incremental gradient optimisation method [1].

The mutual information *MI* is implemented according to [8], using a joint histogram size of $16 \times 16$. The number of samples selected in each iteration is set to 1000. A four-level multiresolution scheme is employed in both the affine and the nonrigid registration step. A Gaussian image pyramid is used for the image data. The nonrigid registrations are performed using a B-spline control point spacing of 64, 32, 16, and 8 mm in all directions, for each resolution respectively. Per resolution, 2000 iterations are performed.

### 2.2 Label image combination

A selection is made of the registered atlas images that are most similar to the patient's image. Similarity is measured by the mutual information, computed on a region of interest around the prostate. The region of interest is defined on the atlas image by a 20 times repeated dilation of the atlas label image $L_i$ using a 3×3×3 kernel approximating a sphere. An atlas $A_i$ is selected if it satisfies the following criterion:

$$\frac{MI(P, A_i \circ T_i)}{\max_j MI(P, A_j \circ T_j)} \geq \alpha, \tag{3}$$

where *MI* denotes the mutual information and $0 \leq \alpha \leq 1$ is a tunable parameter. A value of 0 means that all atlas scans are included in the selection. A value of 1 implies that only the atlas scan with the highest similarity measure is used.

The deformed segmentations of the selected atlas images are combined into a single segmentation $\hat{L}_P(\boldsymbol{x})$ of the patient image. This is done by averaging the deformed label images and thresholding the result at a value of 0.5, which is equivalent to a majority voting rule.

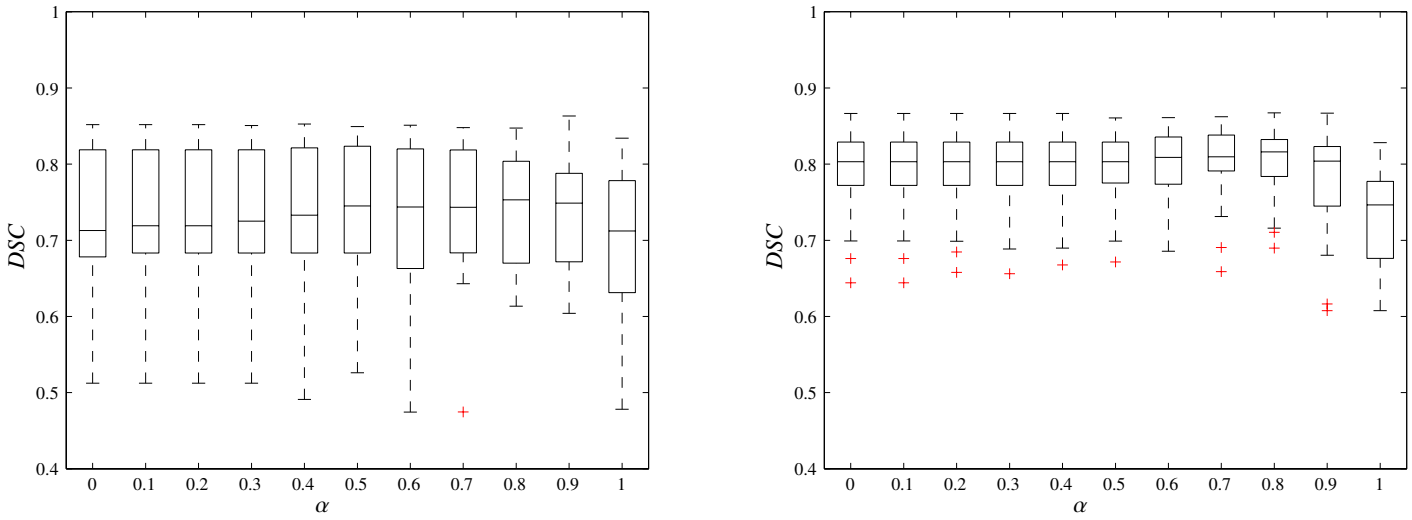
## 3 EXPERIMENTS

The proposed automatic segmentation method is evaluated using a set of 38 MR scans. The images were acquired with a Philips 3T scanner. The MR protocol, a balanced FFE gradient echo sequence, was optimised for visibility of the prostate and rectum. The scans originate from eight volunteers and were made in the context of another study. Seven volunteers were scanned five times, one volunteer was scanned three times. The time between two scans was at least one day, and the volunteers were asked to try to vary the content of their rectum and bladder, to get as much variety between the scans as possible. The scans have a dimension of 512×512×90 voxels of size 0.49×0.49×1.0 mm. Manual segmentations, made by an experienced observer and approved by a radiation oncologist, are available for each scan. The seminal vesicles were marked separately, which allows us to evaluate the segmentation method on the whole prostate or on the prostate excluding the seminal vesicles. Unless mentioned otherwise, results are presented for segmentation of the prostate including the seminal vesicles.

To reduce computation time, the scans are manually cropped to a rectangular region roughly encompassing the prostate, bladder, and rectum. This is the only step in the algorithm that requires user intervention. A histogram equalisation is applied to the images to ensure that not all image intensities fall within only a few histogram bins during computation of the mutual information.

Leave-one-out experiments are performed to evaluate the segmentation algorithm. Atlas images that originate from the same person as the test image are excluded from the atlas set, which results in an atlas composed of 33 images, different for each test image.

The results are evaluated by comparing the automatically generated segmentations with the manual segmentations. A well-known measure of segmentation overlap is the Dice similarity coefficient (*DSC*) [2]. In [9], the intraobserver reproducibility of




(a) *MI* (b) *LMI*

Figure 1: A box-and-whisker plot of the Dice similarity coefficients for different values of $\alpha$, using (a) *MI* and (b) *LMI*. The box shows the first quartile, the median and the third quartile of the data. The whiskers display the total range of the data. The pluses represent outliers.

manually delineating the prostate's peripheral zone on T2-weighted MR images is investigated. A mean *DSC* of 0.883 is reported for 1.5T preoperative scans, and 0.838 for 0.5T scans. Interobserver reproducibility for delineations of the prostate in CT images is measured in [3], where a median *DSC* of 0.82 is reported.

The *DSC* does not provide insight in the spatial distribution of the segmentation errors. The corpus (prostate *without* seminal vesicles) has an approximately spherical shape, which allows us to use a spherical coordinate mapping for visualisation of the segmentation accuracy [3, 6]. The position on the boundary of the corpus is parameterised by two angles, $\theta$ and $\phi$, where the centre of mass of the manual segmentation serves as the centre of rotation. The shortest Euclidian distance between the manual and automatic segmentation boundaries is plotted against $\theta$ and $\phi$.

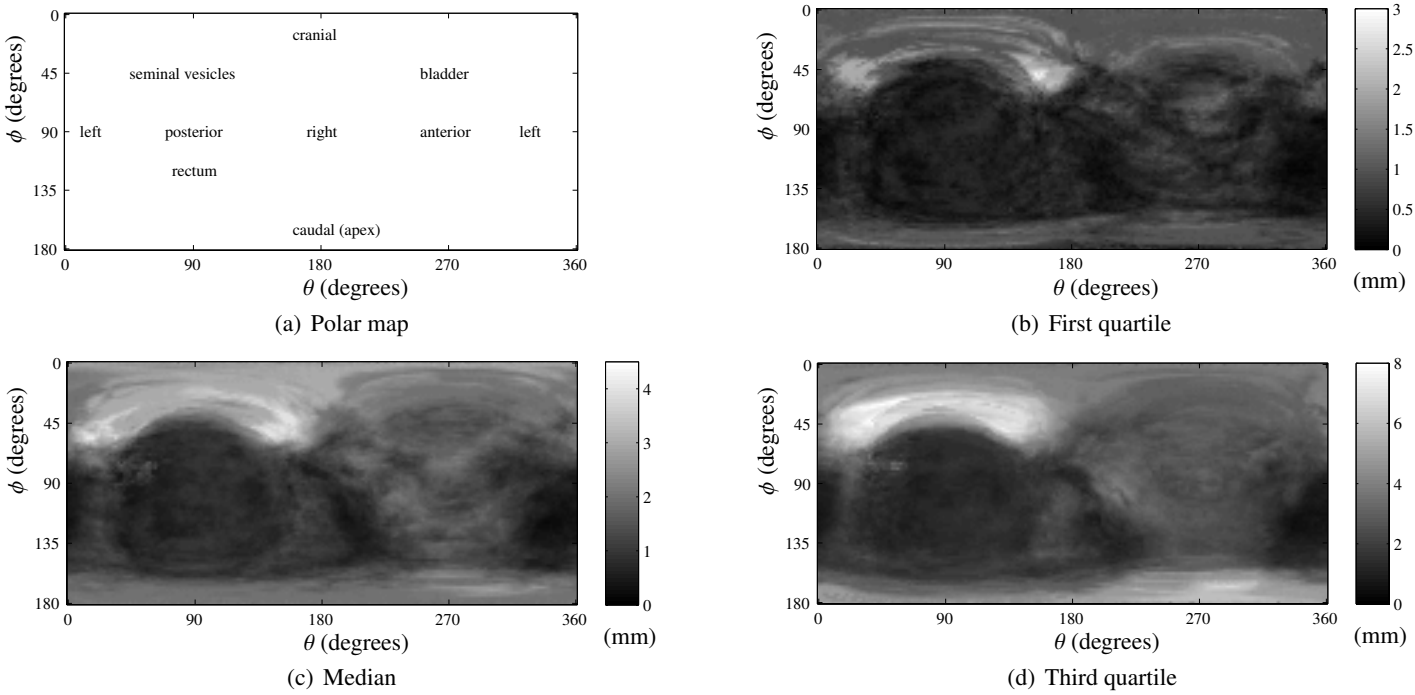
# 4 RESULTS

The leave-one-out experiments were performed using *MI* and *LMI* and for different values of $\alpha$. The *DSC* values were computed for all 38 test scans. Figure 1 summarises the results in two box-and-whisker plots, one for *MI* and one for *LMI*. Each box-and-whisker visualises the distribution of *DSC* values for a specific value of $\alpha$. The results show clearly that *LMI* outperforms *MI* in this application. Both for *MI* and *LMI* the optimal value of $\alpha$ is in the range 0.5-0.8.

For further evaluation, we use *LMI* as the similarity measure and $\alpha = 0.8$. Figure 2 shows the spatial distribution of the segmentation errors. A map of the surface is given in Figure 2(a). For each test scan, the distance between the manually and automatically determined corpus boundaries was calculated as a function of the spherical coordinates $\theta$ and $\phi$. The results for the 38 test scans are summarised in Figures 2(b)-(d) by the first quartile, median, and third quartile. From the figures it is evident that the largest errors occur at the border between the corpus and the vesicles. In Figure 2(b) it can be seen that in 75% of the test scans the error at the vesicles is larger than 2 mm approximately. At the boundary between prostate and rectum smaller errors are made. Figure 2(c) and 2(d) show that in 50% of the test cases the distance remains below 1 mm, and in 75% of the cases below 1.5 mm approximately. The edge between prostate and bladder is determined with slightly worse accuracy (median of 2 mm, third quartile of 3 mm). The tip of the apex ($\phi$ close to 180°) is also determined less accurately.

# 5 CONCLUSION

An automatic prostate segmentation method for pelvic MR images has been investigated in this work. The method is based on matching of manually segmented atlas images. To account for the large variability in shape, multiple atlas images are combined. A computationally efficient localised mutual information similarity measure is used in the matching stage. It is shown to greatly improve the results compared to the common, global mutual information metric.




(a) Polar map (b) First quartile

(c) Median (d) Third quartile

Figure 2: A polar map of the spatial error distribution of the prostate segmentation, excluding the seminal vesicles.

Evaluation has been performed on volunteer data. The method was tested on 38 test scans using leave-one-out experiments. The segmentations obtained by the automatic segmentation method are compared to manual segmentations by means of the Dice similarity coefficient and by plots of the spatial error distribution. A median Dice similarity coefficient of 0.82 is achieved. In the spatial error distribution plots, it can be seen that on a large part of the prostate surface the segmentation errors remain below 2-3 mm for 75% of the test scans. The most serious segmentation errors are made in the seminal vesicles, which is confirmed by visual inspection. Imaging artefacts caused by air in the rectum also influence the results negatively in some cases.

Future work includes evaluation on patient data and comparison with interobserver variability in delineating the prostate.

## 6 ACKNOWLEDGMENTS

The authors thank Ellen Kerkhof for providing the manual segmentations.

## References

[1] D. P. Bertsekas and J. N. Tsitsiklis. Gradient convergence in gradient methods with errors. *SIAM Journal on Optimization*, 10(3):627–642, 2000.

[2] L. R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[3] M. Foskey et al. Large deformation three-dimensional image registration in image-guided radiation therapy. *Physics in Medicine and Biology*, 50:5869–5892, 2005.

[4] G. Hermosillo. *Variational Methods for Multimodal Image Matching*. PhD thesis, INRIA, 2002.

[5] S. Klein et al. A comparison of acceleration techniques for nonrigid medical image registration. In J.P.W. Pluim, B. Likar, and F.A. Gerritsen, editors, *International Workshop on Biomedical Image Registration*, volume 4057 of *Lecture Notes in Computer Science*, pages 151–159. Springer-Verlag Berlin Heidelberg, 2006.

[6] C. Rasch et al. Definition of the prostate in CT and MRI: a multi-observer study. *Int. J. Radiation Oncology Biol. Phys.*, 43:57–66, 1999.

[7] D. Rueckert et al. Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.

[8] P. Thévenaz and M. Unser. Optimization of mutual information for multiresolution image registration. *IEEE Transactions on Image Processing*, 9(12):2083–2099, 2000.

[9] K. H. Zou et al. Statistical validation of image segmentation quality based on a spatial overlap index. *Academic Radiology*, 11:178–189, 2004.