Region-Specific Hierarchical Segmentation of MR Prostate Using Discriminative Learning

Neil Birkbeck, Jingdan Zhang, and S. Kevin Zhou

Siemens Corporate Research

Abstract. We demonstrate the effectiveness of learning-based methods and hierarchical boundary deformation for efficient, accurate segmentation of prostate in T2 weighted MRI data. After normalizing intra- and inter-image intensity variation, Marginal Space Learning (MSL) is used to align a statistical mesh model on the image. This mesh is then hierarchically refined to the image boundary using spatially varying surface classifiers. Using 10-fold cross validation on 50 cases, we achieve accurate delineations (0.89 dice coeff., 1.91mm surface error) in under 3s.

1 Background

The improved contrast of MR abdominal imaging over other modalities affords more accurate and consistent manual contouring of prostate [1,2,3]. Consequently, MR imaging now plays an increasing role in prostate radiotherapy planning, implying automated algorithms for segmenting abdominal structures in MR are necessary to improve the clinical workflow.

However, automated segmentation of the prostate in MR is challenging due not only to global inter-scan variability and intra-scan intensity variation caused by endorectal coil, but also due to the similar appearance of prostate and seminal vesicles at the apex, geometric variability due to disease and adjacent structures such as rectum and bladder.

Many existing automatic approaches are registration-based: one or more template images with segmentations are registered to the target image and the aligned segmentations are fused [4,5,6]. Variation in intensity and appearance is overcome through the use of the appropriate energy (e.g., mutual information) and the use of multiple templates [5]. Although these registration methods perform well when one of the template images is similar to the target image, they are often expensive, requiring several minutes to return a segmentation [6].

In this work, we leverage the successes of learning-based methods on other structures (e.g., liver [7], heart [8]) in other imaging modalities to efficiently segment the prostate in MR. Instead of using an explicit registration of images, a training set of images is compactly represented with discriminative classifiers that are used to align a statistical mesh model onto the image. Rather than relying on edge detectors or hand-designed features [9], we use classifiers that aggregate and choose the best image features from a large feature pool. We apply this approach to T2-weighted axial scans abdominal scans, and demonstrate accurate prostate segmentations (dice coefficient of 0.89) in less than $3s$.

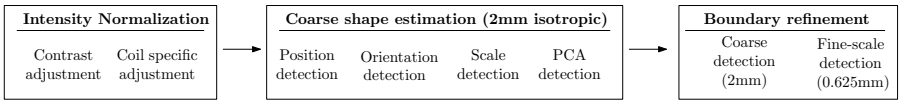

Fig. 1. Our prostate segmentation pipeline first normalizes the image data, then detects the rough shape, and then uses a coarse-to-fine boundary refinement.

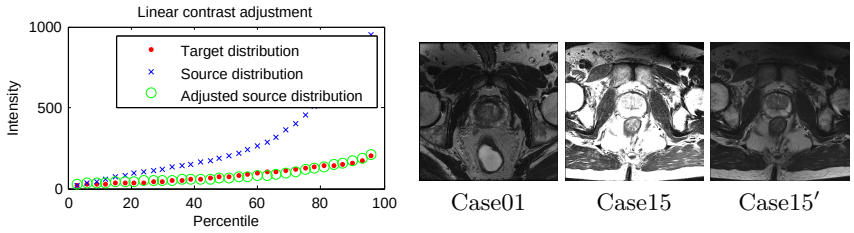

Fig. 2. An illustration of sorted intensity values before and after linear alignment of Case15 to Case01 gives a Case15' with similar global intensity as Case01.

2 Methodology

Our segmentation pipeline addresses the challenges of MR prostate segmentation through the use of robust machine learning techniques (Fig. 1). To overcome variability in intensity, we start by performing an intensity normalization to adjust for global contrast changes (§2.1). Images with endorectal coil (ERC) are then further enhanced by flattening the intensity profile on the bright regions near the coil.

In the next phase of the pipeline, a statistical model of shape variation (2.2) is aligned to these normalized images using marginal space learning (2.3), which efficiently searches for the pose initialization using discriminative classifiers. This shape initialization is then refined by a coarse-to-fine boundary refinement that uses surface varying classifiers to discriminate the boundary of the prostate from adjacent soft-tissue (2.4). To ensure a feasible shape, this final refinement constraints the resulting shape by the statistical shape model.

2.1 Intensity Normalization

In the first phase, we apply a brightness and contrast adjustment to the intensities of the images (Fig. 2). We select one image as a target image, $\hat{I}$, and find a least squares solution for the linear intensity transformation for each image, I_i:

$$a, b = \text{argmin}_{a,b} \sum_{j=3}^{98} (\text{prctile}(\hat{I}, j) - (\text{prctile}(I_i, j)a + b))^2, \tag{1}$$

$$I_i' = I_i a + b, \tag{2}$$

where $\text{prctile}(\hat{I}, j)$ is the j^{th} percentile of the intensities of the target image.

Input Mask Result

Fig. 3. An example of the intensity normalization applied to the ERC images. The input image is thresholded to obtain a mask, which is then used to define the domain for the Poisson editing. The result retains gradient features within the bright region, but the overall intensity within that region is reduced.

For images acquired with an endorectal coil we found that the sharp spikes in intensity near the coil to have a negative impact on our segmentation results. Our next normalization seeks to reduce the overall brightness of these regions while retaining the local image structure.

This is done through an automatic application of Poisson image editing [10]. The bright region, $\Omega \subset \mathbf{R}^2$, is extracted from the image as the non-zero elements of a mask image $M = ((I > \tau_1) \oplus B) \wedge (I > \tau_2)$, where the intensity thresholds, τ_1 and τ_2, are chosen such that $\tau_1 > \tau_2$ and $\oplus B$ is a dilation with a circular ball of radius 7. We then seek to find the adjusted image intensities, $f : \Omega \mapsto \mathbf{R}$, such that the boundary of Ω matches the surrounding image region and so that the gradient within Ω is similar to high pass version of the image. Letting $g(x) = (I - G_\sigma * I)(x)$, be the high pass filtered image, with Gaussian G_σ, the problem of recovering the intensities is posed as a minimization:

$$E(f) = \min \int_\Omega |\nabla f - \nabla g|^2 dx \quad \text{where} \quad f = I \quad \text{on} \quad \delta\Omega. \tag{3}$$

The minimizer of (3) a solution to the Poisson equation:

$$\nabla^2 f = \nabla^2 g. \tag{4}$$

We apply the normalization per slice and empirically found $\sigma = 4$ to be a good value for the the filtering. Figure gives an example of the Poisson editing on slice 19 of Case04 from the PROMISE12 training set.

2.2 Shape Modeling

In order to build the statistical model, the binary input training segmentation masks are first converted to a mesh representation using marching cubes. The resulting meshes are then evenly spherically sampled and brought into alignment using iterated closest points. The result is a single set of triangles, $\mathcal{T}$, and the M vertex positions $\mathcal{V}_i = \{\mathbf{v}_{ij}\}_{j=1}^M$, for each training image, i.

Using these corresponding shapes, orientation and scale variation are removed using Procrustes analysis, which gives a mean shape, $\bar{\mathcal{V}} = \{\bar{\mathbf{v}}_j\}_{j=1}^M$, and

the orientation, $\mathbf{r}_i$, scale, $\mathbf{s}_i$, and translation, $\mathbf{p}_i$ that best map the mean shape to the input meshes, $\mathcal{V}_i$. The remaining variability in the shape is represented with a point distribution model, and the strongest shape modes, $\mathbf{U}_i = \mathbf{u}_{ik}$ are extracted through PCA. A shape can be represented as a linear combination of the modes, blended with the shape coefficients, λ_k:

$$\mathbf{v}_j = \bar{\mathbf{v}}_j + \sum_k \mathbf{u}_{kj}\lambda_k. \tag{5}$$

For efficiency, two levels of a mesh hierarchy are used, the finer resolution has 1127 vertices & 2250 triangles, and the coarser resolution has 565 vertices & 1126 triangles. The coarser resolution is obtained by downsampling the mean shape and extracting the corresponding vertices.

2.3 Marginal Space Learning

During testing on an unseen image, I, an initial segmentation is obtained by aligning the shape model to the image data. We use Marginal Space Learning (MSL) [8] to recover these unknown pose parameters and the first 3 shape coefficients, $\lambda_{1:3}$, for the test image by approximating the posterior:

$$\theta = (\mathbf{p}, \mathbf{r}, \mathbf{s}, \lambda_{1:3}) = \text{argmax}_{\mathbf{p},\mathbf{r},\mathbf{s},\lambda_{1:3}} Pr(\mathbf{p}, \mathbf{r}, \mathbf{s}, \lambda_{1:3}|I). \tag{6}$$

Instead of searching all parameters simultaneously, MSL decomposes the search space into subsequent estimates of 3D searches over position, orientation, and scale. The position is first isolated, and the posterior, $Pr(\mathbf{p}|I)$, is approximated using a discriminative classifier,

$$Pr(\mathbf{p}|I) = Pr(y = +1|I, \mathbf{p}), \tag{7}$$

where the binary random variable, y, takes a value of $+1$ if the prostate is at position $\mathbf{p}$ in image I. In other words, a set of likely candidate positions can be obtained by scanning a classifier over the image.

Given these position estimates, orientation can be estimated in a similar manner,

$$Pr(\mathbf{r}|I, \mathbf{p}) = Pr(y = +1|I, \mathbf{p}, \mathbf{r}), \tag{8}$$

where the search augments each of the position candidates with the plausible set of orientations learned from training data. Scale and the first three PCA components are estimated analogously.

Training For each of the discriminative classifiers (position, orientation, scale, and PCA), a probabilistic boosting tree (PBT) [11] is trained using the known poses and image data in the training set. For the pose estimation, in training and testing, the volume data is resampled to be 2mm isotropic. Position detection uses 3D Haar features, and, to train the binary classifiers, positive instances are sampled within a 7mm range of the true position and negatives are sampled randomly from regions greater than 15mm away. Orientation, scale, and PCA estimates use steerable features sampled within a bounding box of the oriented cube. See ref. [8] for more details.

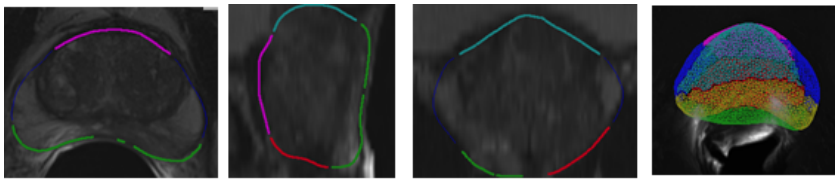

Fig. 4. The surface of the prostate is partitioned to allow for region specific classifiers.

2.4 Boundary Deformation

The pose estimated using MSL aligns the shape model to the image, giving a good initial segmentation. The mesh surface is then iteratively refined using a non-rigid, hierarchical boundary deformation [7]. Each mesh vertex is displaced $\mathbf{v}_i \leftarrow \mathbf{v}_i + t_i\mathbf{n}_i$ along the surface normal, $\mathbf{n}_i$, by finding the most likely displacement:

$$t_i = \text{argmax}_{-\tau \leq t \leq \tau} Pr(\mathbf{v}_i + t\mathbf{n}_i|I), \tag{9}$$

where $Pr(\mathbf{v}_i + t\mathbf{n}_i|I)$ is again modeled with a discriminative classifier, and the search for the best displacement is limited to the range τ. After displacing each point independently, regularity is imposed by projecting the resulting shape onto the linear shape space.

The boundary refinement takes place first on the low resolution mesh and 2mm isotropic volumes. In a subsequent phase, the mesh is upsampled and a finer scale refinement is performed. We perform 10 iterations at each level of the mesh hierarchy, reducing τ by a factor of 0.8 each iteration.

For the higher resolution mesh, in order to account for the varying surface appearance on different regions of the boundary, separate boundary classifiers are utilized [7]. Specifically, the prostate surface is partitioned into 6 regions near the bladder, rectum, peripheral zone (Figure 4).

Training The discriminative model in (9) also uses a PBT classifier and steerable features. Positive samples for the classifier are taken to be ground truth mesh points, and negatives are chosen within a predefined distance from the mesh. The sampling pattern of the steerable features and the negative range are optimized using a testing set.

3 Experimental Design

To train our algorithm we used 10-fold cross validation on the training set available from the PROMISE12 data set[1]. The data consists of 24 cases with endorectal coil and 26 cases with body coil, with a mean resolution of $0.49 \times 0.49 \times 3.27$mm. For the final submission, we trained on the entire set of 50 volumes.

[1] http://promise12.grand-challenge.org

Table 1. Segmentation accuracy using 10-fold cross-validation (CV) with and without image normalization, and the testing on training results for the entire data set (All).

Data	Image Norm.	PCA Detection MSE Mean	PCA Detection MSE Med.	PCA Detection Dice Mean	PCA Detection Dice Med.	Low-res boundary MSE Mean	Low-res boundary MSE Med.	Low-res boundary Dice Mean	Low-res boundary Dice Med.	Hi-res boundary MSE Mean	Hi-res boundary MSE Med.	Hi-res boundary Dice Mean	Hi-res boundary Dice Med.
CV	No	4.02	3.46	0.71	0.74	2.43	1.87	0.83	0.85	2.11	1.51	0.85	0.88
CV	Yes	4.20	3.65	0.70	0.71	2.34	1.94	0.83	0.84	1.91	1.49	0.86	0.89
All	Yes	2.45	2.22	0.82	0.83	1.71	1.64	0.87	0.88	1.42	1.35	0.89	0.90

For evaluation, we compute the symmetric mean surface-to-surface distance (MSE) between two surfaces X and Y as:

$$d_{\text{mse}}(X,Y) = \frac{1}{2}(d(X,Y) + d(Y,X)) \tag{10}$$

$$\text{where} \quad d(X,Y) = \frac{1}{|X|}\sum_{x \in X} d(x,Y), \tag{11}$$

$$\text{and} \quad d(x,Y) = \min_{y \in Y} |x - y|. \tag{12}$$

When computing d_{mse}, the ground truth segmentation masks are first converted to a mesh representation using marching cubes.

And we also report the dice coefficient computed on the segmented masks, $\mathbf{X}$ and $\mathbf{Y}$:

$$d_{\text{dice}}(\mathbf{X},\mathbf{Y}) = 2\frac{|\mathbf{X} \bigcap \mathbf{Y}|}{|\mathbf{X}| + |\mathbf{X}|}. \tag{13}$$

Since our segmentation is mesh-based, we first convert to isotropic masks of 0.5mm resolution to measure the dice coefficient.

4 Results and Discussion

In this section we present quantitative and qualitative results of our system on the training data from the PROMISE12 website.

4.1 Quantitative & Qualitative Test Results

Table 1 illustrates the quantitative results for different phases of the algorithm for the cross-validation (CV) cases with and without intensity normalization and for testing on training data (All). Notice that each phase, from pose initialization to fine-scale boundary refinement, provides an improvement in accuracy.

We see that the image normalization appears to be detrimental in the initial pose estimation, but gives improved results for the final segmentation. The 10-fold cross validation with normalization results in a mean dice coefficient of 0.86 and a median dice coefficient of 0.89. In the CV configurations, Case23 fails to be detected, causing an increase in mean error. The reason for this failure is that

Test Case09 Test Case21 Test Case28

Fig. 5. Qualitative results on axial slices of 3 unseen test cases from the PROMISE12 challenge.

Case23 has an abnormally large volume of 315ml vs the median volume of 43ml. Ensuring that the training data has large enough variation resolves this issue (as seen in the test on training case, row *All*). In our experience an increased amount of training data is important for robust pose detection but less so for improvement in the boundary detection.

Figure 5 illustrates some of the results on the unseen test cases.

4.2 Implementation & Runtime Details

All of our experiments have been run on a 2 processor Intel(R) Xeon(R) 2.7 GHz CPU with 72.0GB of RAM and running Windows Server 2008. In total, the system has 12 cores. With hyper-threading a maximum of 24 threads are used.

Implementation Details Our segmentation algorithm is implemented within a modular C++ software library for detection and segmentation of medical imaging [12]. Our implementation links to VXL and boost and also uses OpenMP for fine-grained parallelization of both the training and the testing. While it is also possible to use the GPU for acceleration, we have not done so in this paper. The intensity normalization component is not yet optimized and is implemented as a preprocess in Matlab.

Runtime and Resource Usage Using the system detailed above, training all the detectors takes 2 hours on the entire training set. Intensity normalization takes about 1.2s/volume (on average) in Matlab [2]. Running the full detection and segmentation up to the fine-scale boundary takes a total of 1.1s per volume. (segmentation up to the coarse level boundary detection only takes 0.6s). The peak memory usage while running all of the 30 unseen test cases was 86 MB.

[2] Assuming only the ERC images need the Poisson normalization

5 Concluding Remarks

We demonstrated the effectiveness of learning-based methods for fast and accurate segmentation of prostate in MR scans. Overall, our system behaved well on the PROMISE12 data, and we have shown that our segmentation can serve as a replacement (or initialization) for slower registration-based methods.

References

1. Rasch, C., Barillot, I., Remeijer, P., Touw, A., van Herk, M., Lebesque, J.V.: Definition of the prostate in CT and MRI: a multi-observer study (January 1999)
2. Villeirs, G.M., Van Vaerenbergh, K., Vakaet, L., Bral, S., Claus, F., De Neve, W.J., Verstraete, K.L., De Meerleer, G.O.: Interobserver delineation variation using CT versus combined CT + MRI in intensity modulated radiotherapy for prostate cancer. Strahlentherapie und Onkologie **181** (2005) 424–430 10.1007/s00066-005-1383-x.
3. Lee, Y., Bollet, M., Charles-Edwards, G., Flower, M., Leach, M., McNair, H., Moore, E., Rowbottom, C., Webb, S.: Radiotherapy treatment planning of prostate cancer using magnetic resonance imaging alone. Radiotherapy and oncology **66**(2) (2003) 203–216
4. Martin, S., Daanen, V., Troccaz, J.: Atlas-based prostate segmentation using an hybrid registration. International Journal of Computer Assisted Radiology and Surgery **3**(6) (2008) 485–492
5. Klein, S., van der Heide, U., Lips, I., van Vulpen, M., Staring, M., Pluim, J.: Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. Medical Physics **35** (2008) 1407
6. Sebastien Martin, Vincent Daanen, J.T.: Automated segmentation of the prostate in 3D MR images using a probabilistic atlas and a spatially constrained deformable model. Medical Physics **1** (2010)
7. Ling, H., Zhou, S.K., Zheng, Y., Georgescu, B., Suehling, M., Comaniciu, D.: Hierarchical, learning-based automatic liver segmentation. In: CVPR, Los Alamitos, CA, USA, IEEE Computer Society (2008) 1–8
8. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. IEEE Trans. Med. Imag. **27**(11) (November 2008) 1668–1681
9. Vikal, S., Haker, S., Tempany, C., Fichtinger, G.: Prostate contouring in MRI guided biopsy. In: Proceedings of SPIE. Volume 7259., NIH Public Access (2009) 72594A
10. Pérez, P., Gangnet, M., Blake, A.: Poisson image editing. ACM Trans. Graph. **22**(3) (July 2003) 313–318
11. Tu, Z.: Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering. In: IEEE Int. Conf. on Comp. Vis. and Patt. Recog. (2005) 1589 –1596 Vol. 2
12. Sofka, M., Ralovich, K., Birkbeck, N., Zhang, J., Zhou, S.: Integrated detection network (IDN) for pose and boundary estimation in medical images. In: Proceedings of the 8th International Symposium on Biomedical Imaging (ISBI 2011), Chicago, IL (30 Mar – 2 Apr 2011)