

Incremental Shape Statistics Learning for Prostate Tracking in TRUS

Pingkun Yan and Jochen Kruecker

Philips Research North America
345 Scarborough Road, Briarcliff Manor, NY 10510

Abstract. Automatic delineation of the prostate boundary in transrectal ultrasound (TRUS) can play a key role in image-guided prostate intervention. However, it is a very challenging task for several reasons, especially due to the large variation of the prostate shape from the base to the apex. To deal with the problem, a new method for incrementally learning the patient-specific local shape statistics is proposed in this paper to help achieve robust and accurate boundary delineation over the entire prostate gland. The proposed method is fast and memory efficient in that new shapes can be merged into the shape statistics without re-computing using all the training shapes, which makes it suitable for use in real-time interventional applications. In our work, the learned shape statistics is incorporated into a modified sequential inference model for tracking the prostate boundary. Experimental results show that the proposed method is more robust and accurate than the active shape model using global population-based shape statistics in delineating the prostate boundary in TRUS.

1 Introduction

Transrectal ultrasound (TRUS) is currently the most commonly used imaging modality for image-guided biopsy and therapy of prostate cancer due to its real-time nature, cost effectiveness, and simplicity. Accurate delineation of the prostate boundary in TRUS can play a key role in image-guided prostate interventions. However, extracting the prostate boundary in TRUS is a challenging task due to the low signal-to-noise ratio of ultrasound imaging, the large intensity variation both inside and outside of the prostate, and especially the large shape variations through the whole gland from the base to the apex.

Methods for segmenting the prostate in static TRUS images have been reported in the past. Abolmaesumi and Sirouspour [1] traced the prostate boundary by moving from one estimated boundary point to the next following a trajectory described by a constant velocity dynamic model. Ladak et al. [2] employed a discrete dynamic contour guided by shape statistics to segment the prostate in TRUS. Shen et al. [3] proposed a segmentation method by first classifying a TRUS image using support vector machine with Gabor filter bank and then fitting a statistical prostate shape model to the output image. These static image segmentation methods have been successful in the mid-gland, because the

boundary shapes in that area are quite regular with small variation. However, due to the large prostate shape variation in the base and the apex areas, poor segmentation performance in those areas was indicated [3, 4]. Recently, Yan et al. [5] addressed the problem using a dynamic shape modeling approach. The prostate shape statistics is continuously recomputed from several of the most recently segmented prostate shapes. The method obtained considerable success by incorporating the adaptive local shape statistics into the segmentation process. However, due to the continuous model recalculation using the obtained shapes, this method is not computationally efficient and also requires storing all the training shapes.

In this paper, we propose a new method for incrementally learning the patient-specific local shape statistics to help achieve accurate boundary delineation over the whole prostate gland in TRUS. New shapes that become available during the delineation process can be merged into the shape statistics using an incremental subspace learning method [6, 7] without explicitly recomputing using all the training shapes. Therefore, the proposed learning method is computationally efficient and removes the need of storing all the training shapes, which makes it suitable for use in interventional applications. In addition, the incremental learning method updates the shape statistics in a smoother way compared to the work in [5]. Instead of segmenting each frame separately, the shape statistics is incorporated into a modified sequential inference model for contour based tracking to exploit the shape change dynamics from frame to frame. The prostate boundary tracker starts with the global population-based shape statistics. As more and more prostate boundary shapes from a particular patient are available along with the tracking, the learned shape statistics will become patient-specific and can quickly adapt to the prostate boundaries in a local neighborhood by weighing more on the most recent shapes. Finally, since the proposed method does not employ any numerical optimization, it is much faster compared to the previous methods on prostate boundary delineation [2, 3, 5], which is a highly desired feature for use in real-time interventional applications. The performance of the method is demonstrated by tracking the prostate boundaries in a series of TRUS sequences.

2 Incremental Shape Statistics Learning

Let a prostate contour S_t in the frame acquired at time t represented by a series of k contour points, i.e. $S_t = \{\mathbf{v}_t^i | i = 1, \dots, k\}$. Each contour point \mathbf{v}_t^i is a 2D vector $\{x_t^i, y_t^i\}$. The contour points are equally spaced (Euclidean distance based spacing) and sampled from the shapes [5, 9]. The active shape model (ASM) introduced by Cootes et al. [8] is used in our work for shape modeling, which provides a compact statistical shape representation by modeling contour point distribution. Suppose that the shape statistics has been computed from previous n prostate boundary observations at time t_0 . The shapes can be denoted by $A = \{S_i - \bar{S}_A | i = 1, \dots, n\}$, where $\bar{S}_A = \frac{1}{n} \sum_{i=1}^n S_i$ is the mean shape. The observed shapes can be decomposed by using the singular value decomposition

(SVD) as $A = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where each column of \mathbf{U} is an eigenvector of A and $\mathbf{\Sigma}$ is a diagonal matrix containing the corresponding eigenvalues. With the computed shape statistics, a new shape S can be represented by using the mean shape and the eigenvectors linearly combined by a coefficient vector $\mathbf{b} = \mathbf{U}^T(S - \bar{S}_A)$. The approximation of the shape S constrained by the shape statistics can be obtained by $\hat{S} = \bar{S}_A + \mathbf{U}\mathbf{b}$.

As the TRUS probe moves to image different areas of the prostate gland, it is desirable to use the newly available shapes to update the previous shape statistics to have the patient-specific local prostate shape knowledge. Let $B = \{S_i - \bar{S}_B | i = 1, \dots, m\}$ represent the differences between the recent m shapes acquired after t_0 and their mean shape $\bar{S}_B = \frac{1}{m} \sum_{i=1}^m S_i$. The objective is to merge the new shape set B into the shape statistics computed from the set A at time t_0 . To achieve this goal, the incremental subspace learning algorithm [6,7], which was originally used for learning the appearance of a rectangular area in video frames, is modified to incrementally learn the shape statistics of deformable contours. Thus, the shape statistics can be updated by merging the new observations without recomputing using all the training shapes. Let C denote the concatenation of shape sets A and B . The mean shape in the new shape statistics is computed as

$$\bar{S}_C = \frac{nf}{nf+m} \bar{S}_A + \frac{m}{nf+m} \bar{S}_B, \quad (1)$$

where f is a forgetting factor $f \in \mathcal{R}$ and $0 \leq f \leq 1$. The smaller the value of f , the faster the information from the old shape statistics is discarded, and vice versa. To assist the effective SVD computation, let \tilde{B} be the component of B orthogonal to \mathbf{U} and $\hat{B} = [B_1 \ \dots \ B_m \ \sqrt{\frac{nm}{n+m}}(\bar{S}_B - \bar{S}_A)]$. The concatenated matrix C can be expressed as $C = [A \ B] = [\mathbf{U} \ \tilde{B}]\mathbf{R} \begin{bmatrix} \mathbf{V}^T & 0 \\ 0 & \mathbf{I} \end{bmatrix}$, where $\mathbf{R} = \begin{bmatrix} f\mathbf{\Sigma} & \mathbf{U}^T \hat{B} \\ 0 & \tilde{B}(\hat{B} - \mathbf{U}\mathbf{U}^T \hat{B}) \end{bmatrix}$. Finally, the incrementally updated shape statistics of C can be obtained as $\mathbf{U}' = [\mathbf{U} \ \tilde{B}]\tilde{\mathbf{U}}$ and $\mathbf{\Sigma}' = \tilde{\mathbf{\Sigma}}$, where $\tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^T$ is the SVD of \mathbf{R} .

The proposed incremental learning algorithm involves computation of QR decomposition of a $k \times (m+1)$ matrix and SVD of a $(e+m+1) \times (e+m+1)$ matrix for each update, where e is the number of eigenshapes at t_0 . Compared to the simplistic approach of applying SVD to a $k \times (n+m)$ matrix for recomputing shape statistics at each step, our proposed method learns the shape statistics in a more efficient way, especially when the number of training shapes n gets large. There is also no need to store the $k \times n$ contour points of the old training shapes. The resulted reduction in computational time and memory usage makes the algorithm well-suited for real-time image-guidance tasks.

3 Sequential Inference Model for Tracking

Instead of separately segmenting each TRUS frame, the incremental shape statistics learning algorithm is incorporated into a visual tracking framework for

prostate boundary delineation to exploit the gradual shape transition along with the movement of the TRUS probe. Given a set of observed TRUS frames $\mathcal{I}_t = \{I_i | i = 1, \dots, t\}$, the aim is to delineate the prostate shape S_t in the current frame t , which can be formulated as to find S_t that maximizes the probability $p(S_t | \mathcal{I}_t)$. By using Bayes' theorem, we have

$$p(S_t | \mathcal{I}_t) \propto p(I_t | S_t) \int p(S_t | S_{t-1}) p(S_{t-1} | \mathcal{I}_{t-1}) dS_{t-1}. \quad (2)$$

The probability $p(S_{t-1} | \mathcal{I}_{t-1})$ denotes the tracking result obtained in the previous frame. The probability $p(S_t | S_{t-1})$ describes the dynamic process of propagating the prostate shapes from frame $t-1$ to frame t . The probability $p(I_t | S_t)$ measures how well the propagated contour S_t can be fit into the frame t . The details of the probability terms on the right side of (2) are presented as follows.

For computational efficiency and robustness, prostate shapes are projected into the subspace of the incrementally learned shape statistics by using an affine transformation \mathbf{T}_t and the coefficient vector \mathbf{b}_t as in Section 2

$$S_t = \mathbf{T}_t (\bar{S}_t + \mathbf{U}_t \mathbf{b}_t). \quad (3)$$

Let $\mathbf{x}_t = [x_t, y_t, \theta_t, s_t, \alpha_t, \phi_t]^T$ be the translations along x and y axes, rotation angle, scale, aspect ratio, and skew direction in the affine transformation. The state vector \mathbf{X}_t can be composed by concatenating the transformation parameters and shape representation coefficients as $\mathbf{X}_t = [\mathbf{x}_t; \mathbf{b}_t]$. The prostate shape transition between the TRUS frames can be described as

$$\mathbf{X}_t = \mathbf{X}_{t-1} + \Delta \mathbf{X} = [(\mathbf{x}_t + \Delta \mathbf{x}); (\mathbf{b}_t + \Delta \mathbf{b})], \quad (4)$$

where the difference $\Delta \mathbf{X}$ follows normal distribution $\mathcal{N}(\Delta \mathbf{X}; 0, \Psi)$ and Ψ is a diagonal covariance matrix whose elements are the corresponding variances of the elements in \mathbf{X}_t . Therefore, we have

$$p(S_t | S_{t-1}) = p(\mathbf{X}_t | \mathbf{X}_{t-1}) = \mathcal{N}(\mathbf{X}_t; \mathbf{X}_{t-1}, \Psi). \quad (5)$$

The probability $p(I_t | S_t)$ of observing TRUS frame I_t given a shape S_t can be measured by the distance between S_t and the detected prostate boundary. The dark-to-bright transition has been shown to be an effective prostate boundary indication feature [9], which can be computed by the pixel intensity contrast along the normal vector of the contours. Let $\{F_{max}^i | i = 1, \dots, k\}$ denote the points with the maximal feature values corresponding to each contour point, the probability $p(I_t | S_t)$ can be measured

$$p(I_t | S_t) \propto \exp \left(-\frac{1}{2r^2k} \sum_{i=1}^k \|F_{max}^i - x_t^i\|^2 \right), \quad (6)$$

where r is the spacing between the points on the normal vector profile.

The particle filter based CONDENSATION algorithm introduced by Isard and Blake [10] is modified for solving this deformable contour tracking problem formulated in (2). The workflow of the tracking method is shown in Fig. 1. Since the

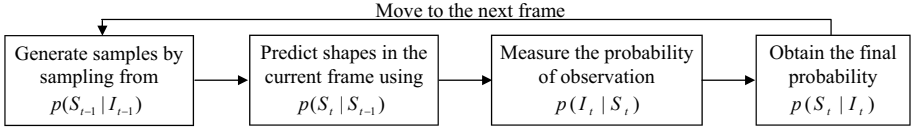


Fig. 1. Workflow of the particle filter based tracking method using the sequential inference model

proposed method does not employ any numerical optimization, it runs in linear time $O(N)$, where N is the number of samples. The final shape can be estimated by computing the expectation of all the samples $\hat{S}_t = \sum_{i=1}^N p(S_t^i | \mathcal{I}_t) S_t^i$.

4 Experimental Results

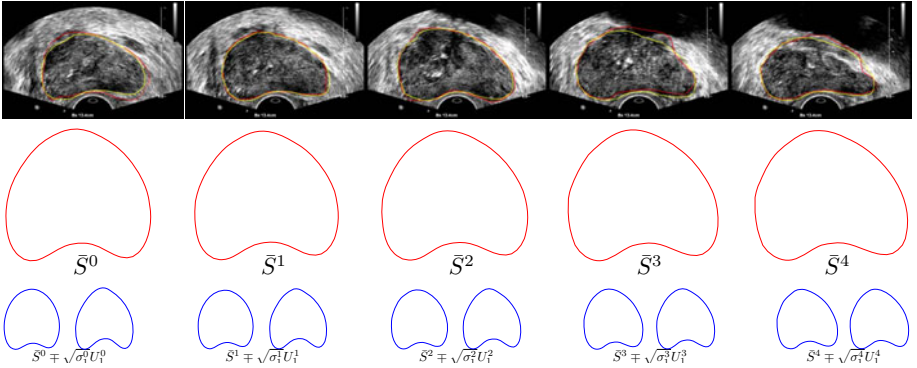
In our experiments, the proposed method was retrospectively tested on sequences collected during TRUS guided prostate cancer biopsy procedures using an iU22 ultrasound system (Philips Healthcare, Andover, MA). The TRUS frames were digitized by a video card at 30 frames per second (fps). Each grabbed TRUS frame has 640×480 pixels and the pixel sizes were 0.1493mm and 0.1798mm in 4cm and 5cm depth settings, respectively. In total, 3215 video frames were grabbed from 20 different patients. An experienced radiologist manually delineated the prostate boundary in one out of every ten frames of the TRUS sequences, in total 316 frames, to provide the ground truth for validation.

In our experiments, the tracked prostate boundary was represented by a contour with $k = 64$ points. The initial prostate shape statistics was computed using 62 prostate shapes previously obtained from a different data set. To quickly incorporate the newly tracked shapes into the shape statistics, the incremental learning was performed in every frame. The forgetting factor f in (1) was empirically set to 0.95 to weigh the learning more on the recently tracked contours, while maintaining the smoothness of the incrementally learned shape statistics. The number of samples N used in the particle filter is a trade-off between the efficiency and accuracy. A larger number of samples may give a bit more accurate tracking results, however, will also make the algorithm slower. The results demonstrated in this paper were obtained by using $N=1000$ samples.

The proposed method starts tracking from a given initial contour, which can be obtained by either manual delineation or using an automatic 2D TRUS image segmentation algorithm [3, 5]. The initial values of the affine transformation parameters and the shape representation coefficients can be computed by using the initial contour. The variance of the normal distribution in the dynamical model (5) was empirically set to be $[2.0, 2.0, 0.001, 0.05, 0.05, 0.001]$ for the transformation part and $\sqrt{\Sigma}/10$ for the shape representation part. Only the first 3 most significant eigenvectors in the shape statistics were used in our work. The proposed method was developed in Matlab. Once initialized, it tracks the prostate boundary in TRUS fully automatically. On a Core2 1.86 GHz PC, we achieved

Table 1. MAD errors (mm) of the ASM and the proposed tracking method on delineating the prostate boundary in the TRUS sequences

Method	seq-1	seq-2	seq-3	seq-4	seq-5	seq-6	seq-7	seq-8	seq-9	seq-10	
ASM	3.28	2.54	2.24	1.66	3.09	2.54	1.73	2.67	2.34	2.80	
Tracking	0.95	1.92	1.45	1.62	2.16	0.83	0.63	0.76	1.54	1.36	
	seq-11	seq-12	seq-13	seq-14	seq-15	seq-16	seq-17	seq-18	seq-19	seq-20	Average
ASM	1.66	1.20	3.01	1.63	3.02	4.73	1.92	2.97	2.53	2.61	2.51±0.78
Tracking	1.52	0.96	1.89	1.46	2.26	1.22	1.49	1.98	1.42	1.39	1.44±0.46

**Fig. 2.** The 1st row shows the delineation of prostate boundary in TRUS with the manually delineated contour in red and the automatic tracking result in yellow; the 2nd row shows the mean shapes in the incrementally learned shape statistics at each stage corresponding to the video frames; the 3rd row shows the shapes generated by varying the mean shapes with the largest mode of variation

12 fps with the above parameter setting. On the same computer, the ASM segmented the sequences with 3 fps.

Fig. 2 shows the tracking results of a TRUS sequence and the learned shape statistics corresponding to the displayed frames. It can be observed that the mean shapes and the major shape components gradually change towards the tracked shapes through incremental learning. Fig. 3 demonstrates the tracking results of the proposed method in different regions of the prostate gland. The initialization contours are included to show that the proposed method can successfully track the prostate boundary when the prostate boundary shape becomes significantly different from the initial shape. In general, the tracking results are close to the manually delineated contours.

Besides the visual inspection, we also quantitatively evaluated the performance of the proposed method. The ASM [8], which has been widely used for segmenting the prostate in TRUS [2, 3, 5, 9], was applied to the same data set by detecting the prostate boundary using the same boundary feature and initial contours for comparison. For TRUS sequence segmentation, the result of ASM

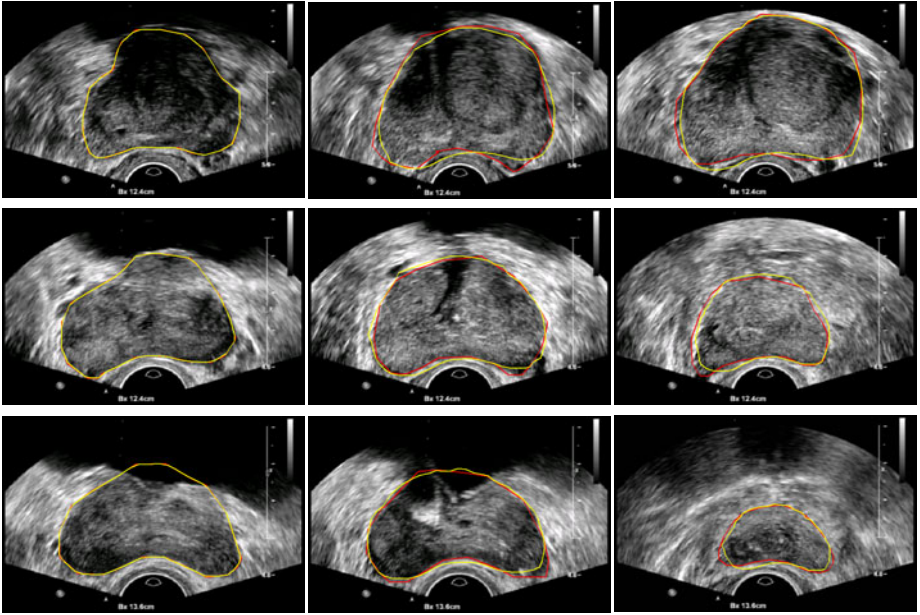


Fig. 3. The first column shows the initialized contours to start the tracking of each sequence. The rest two columns show the tracking results in different part of the prostate gland, where the manually delineated contours are shown in red and the tracked contours are shown in yellow.

Table 2. MAD errors (mm) of the ASM and the proposed tracking method in the base, mid-gland, and apex areas over the 20 TRUS sequences, respectively

Method	base	mid-gland	apex
ASM	2.53 ± 1.17	1.48 ± 0.59	3.52 ± 1.63
Tracking	1.39 ± 0.67	1.20 ± 0.42	1.73 ± 0.76

from one frame was propagated to the next as the initialization. The mean absolute distance (MAD) error was employed for quantitative measurement. As shown in Table 1, the accuracy of the proposed method was significantly better than the ASM, which uses only the global population-based shape statistics. In our work, the statistical significance was evaluated using paired t test ($p < 0.05$).

It is also interesting to see the performance of the methods in each of the base, mid-gland, and apex regions. For quantitative evaluation, the ground truth segmentations of each patient were sequentially divided into three groups with equal number of frames in each group. Since the sequence was obtained by scanning through the prostate from the base to apex, the three groups were labeled as base, mid-gland, and apex in the specified order. The evaluation results are shown in Table 2. Compared to the ASM, prostate boundary delineation

accuracy was significantly improved in all the regions by using the proposed tracking approach with incrementally learning the shape statistics.

5 Conclusions

In this paper, we presented a new algorithm for incremental shape statistics learning, which is well-suited for use in image-guidance applications due to its computation and memory efficiency. The learning algorithm was successfully incorporated into a modified sequential inference model to help achieve more robust and accurate delineation of the prostate boundary in TRUS over the entire prostate gland. Experimental results show that the contour delineation error of the proposed method was 42.6% less than that of the ASM. Significant improvement was made in the base and apex areas with 45.1% and 50.1% less delineation errors, respectively. To the best of our knowledge, the proposed tracking algorithm is the first method that aims to extract the prostate boundary from TRUS in real-time during image-guided prostate interventions. In our future work, we will investigate more efficient software implementation of the algorithm and also explore hardware acceleration to get the program run in real-time.

References

1. Abolmaesumi, P., Sirouspour, M.R.: An interacting multiple model probabilistic data association filter for cavity boundary extraction from ultrasound images. *IEEE Trans. Med. Imaging* 23(6), 772–784 (2004)
2. Ladak, et al.: Prostate boundary segmentation from 2D ultrasound images. *Med. Phys.* 27(8), 1777–1788 (2000)
3. Shen, et al.: Segmentation of prostate boundaries from ultrasound images using statistical shape model. *IEEE Trans. Med. Imaging* 22, 539–551 (2003)
4. Yan, et al.: Optimal search guided by partial active shape model for prostate segmentation in TRUS images. In: *Medical Imaging 2009*, vol. 7261. SPIE (2009)
5. Yan, et al.: Segmenting TRUS video sequences using local shape statistics. In: *Medical Imaging 2010*, vol. 7625. SPIE (2010)
6. Levy, A., Lindenbaum, M.: Sequential karhunen-loeve basis extraction and its application to images. *IEEE Transactions on Image Processing* 9(8), 1371–1374 (2000)
7. Ross, D.A., Lim, J., Lin, R.S., Yang, M.H.: Incremental learning for robust visual tracking. *International Journal of Computer Vision* 77, 125–141 (2008)
8. Cootes, et al.: Active shape models – their training and application. *Comput. Vision Image Understand.* 61(1), 38–59 (1995)
9. Hodge, et al.: Prostate boundary segmentation from ultrasound images using 2D active shape models. *Comp. Methods Prog. Biomed.* 84, 99–113 (2006)
10. Isard, M., Blake, A.: CONDENSATION – conditional density propagation for visual tracking. *International Journal of Computer Vision* 29(1), 5–28 (1998)