Accurate Prostate Volume Estimation Using Multifeature Active Shape Models on T2-weighted MRI

Robert Toth, MS, B. Nicolas Bloch, MD, Elizabeth M. Genega, MD, Neil M. Rofsky, MD, Robert E. Lenkinski, PhD, Mark A. Rosen, MD, PhD, Arjun Kalyanpur, MD, Sona Pungavkar, MD, Anant Madabhushi, PhD

Rationale and Objectives: Accurate prostate volume estimation is useful for calculating prostate-specific antigen density and in evaluating posttreatment response. In the clinic, prostate volume estimation involves modeling the prostate as an ellipsoid or a spheroid from transrectal ultrasound, or T2-weighted magnetic resonance imaging (MRI). However, this requires some degree of manual intervention, and may not always yield accurate estimates. In this article, we present a multifeature active shape model (MFA) based segmentation scheme for estimating prostate volume from in vivo T2-weighted MRI.

Materials and Methods: We aim to automatically determine the location of the prostate boundary on in vivo T2-weighted MRI, and subsequently determine the area of the prostate on each slice. The resulting planimetric areas are aggregated to yield the volume of the prostate for a given patient. Using a set of training images, the MFA learns the most discriminating statistical texture descriptors of the prostate boundary via a forward feature selection algorithm. After identification of the optimal image features, the MFA is deformed to accurately fit the prostate border. An expert radiologist segmented the prostate boundary on each slice and the planimetric aggregation of the enclosed areas yielded the ground truth prostate volume estimate. The volume estimation obtained via the MFA was then compared against volume estimations obtained via the ellipsoidal, Myschetzky, and prolated spheroids models.

Results: We evaluated our MFA volume estimation method on a total 45 T2-weighted in vivo MRI studies, corresponding to both 1.5 Tesla and 3.0 Tesla field strengths. The results revealed that the ellipsoidal, Myschetzky, and prolate spheroid models overestimated prostate volumes, with volume fractions of 1.14, 1.53, and 1.96, respectively. By comparison, the MFA yielded a mean volume fraction of 1.05, evaluated using a fivefold cross-validation scheme. A correlation with the ground truth volume estimations showed that the MFA had an r^2 value of 0.82, whereas the clinical volume estimation schemes had a maximum value of 0.70.

Conclusions: Our MFA scheme involves minimal user intervention, is computationally efficient and results in volume estimations more accurate than state of the art clinical models.

Key Words: Prostate volume; active shape models; prostate cancer; MRI; texture; image processing.

Prostate volume has been shown to be a strong predictor of treatment outcome for patients with prostate cancer (1,2), especially when combined with a baseline prostate-specific antigen (PSA) level (3). Prostate volume has also been shown to be useful in determining PSA density (4). The most common method for estimating the prostate volume involves modeling the prostate as a simple geometric shape based on manually estimated measurements of the anteroposterior, transverse, and craniocaudal lengths of the prostate.

The most common models for approximating the prostate shape are the ellipsoid model (4,5,6–14) and the prolate spheroid model (4,6,9). It is important to note that the ellipsoidal model has been a clinical standard for comparisons from at least 1991 (7) to the present day (12,14). Some researchers have reported that in several cases the ellipsoid model underestimated the prostate volume (6,8,11,15). Tewari et al (11) and Eri et al (6) both found that the ellipsoid model underestimated the prostate volume by about 10%. Matthews et al (8) found that the ellipsoid model from transrectal ultrasound (TRUS) imagery underestimated the volume for large prostates (>50 mL), but overestimated the volume for small prostates (<30 mL). Myschetzky et al overcame this underestimation by proposing a new formula in which the ellipsoid volume estimation is multiplied by a factor of 1.34 (15). Additionally, methods involving manual intervention are typically subject to inter- and intraobserver variability (16,17) and these volume estimations are not highly reproducible.

Acad Radiol 2011; 18:745–754

From the Department of Biomedical Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ 08854 (R.T., A.M.); Beth Israel Deaconess Medical Center, Boston, MA (B.N.B., E.M.G., N.M.R., R.E.L.); University of Pennsylvania, Department of Radiology, Philadelphia, PA (M.A.R.); Teleradiology Solutions, Bangalore, India (A.K.); Dr. Balabhai Nanavate Hospital, Mumbai, India (S.P.). Received July 22, 2010; accepted January 11, 2011. **Address correspondence to:** A.M. e-mail: anantm@cri.rutgers.edu

doi:10.1016/j.acra.2011.01.016

Although most prostate volume estimations are done using TRUS imagery, a strong correlation ($r^2 = 0.925$) has been shown between the volume estimations obtained using TRUS and from magnetic resonance imaging (MRI) (5). In addition, the ellipsoidal model was found to yield accurate volume estimations for T2-weighted MRI of the prostate, even when an endorectal coil was used (12). In previous work (13), it was found that the ellipsoidal volume estimations were more accurate than a planimetry-based approach (aggregating a series of measurements from each slice) when using a surface coil; in contrast to other work (12) in which planimetry estimates were found to yield more accurate volume estimations compared to the ellipsoidal model estimates when using an endorectal coil. In previous work (5), a planimetry based volume estimation was performed by measuring the areas from manual two-dimensional (2D) segmentations of the prostate on each slice.

Our prostate volume estimation method is related to the technique used by Hoffelt et al (5), in which the gland areas obtained by manual segmentation of the capsule were aggregated across multiple 2D sections. However, although Hoffelt et al (5) obtained the prostate areas manually, we aim to perform the capsule segmentations automatically via the use of a multifeature active shape model (MFA) (18).

Although active shape models (ASMs) are a popular segmentation technique, they sometimes fail to converge to the desired object boundary in the case of weak image gradients (19). ASMs essentially model the shape of an object a statistical variations in a set of anatomical landmarks the appearance of an object as a Gaussian distribution of intensities near each anatomical landmark. The appearance model typically uses the intensities of the image to learn a statistical appearance model. However, relying solely on the intensity information may not be sufficient for accurately detecting the correct boundary, especially if different regions of the image, or different regions within the desired object, have similar intensity values. This is particularly true of MRI in which strong bias field inhomogeneity artifacts can significantly obfuscate object boundaries (20).

In this work we present a new ASM that we call the MFA. We calculate the gray level statistics of each image by convolving a set of kernels with the intensity image. These include the Kirsch (21) and Sobel (22) kernels to better quantify the edges of the prostate border. Although traditional ASMs use neighboring intensity information, they are dependent on the normal to the shape at any given landmark point. By contrast, the Gaussian and mean kernels take neighboring information into account and yet do not depend on the normal of the shape. Additionally, the Cartesian x and y coordinates of each landmark point are included as additional "features." Further, because texture features of the prostate boundary are not always optimally modeled as a Gaussian, we describe the distributions as sums of multiple Gaussians (GMM) (23), allowing us to better characterize the feature distributions at each landmark on the prostate boundary. A forward feature selection scheme is employed to determine the best textural features in terms of discriminability between the prostate border and background. Only these features are then employed in conjunction with the MFA.

The MFA is employed to estimate the gland area on each slice, which is multiplied by the slice interval (distance between center of adjacent slices) to yield an estimation of the prostate volume. This estimation is compared to the ellipsoid (4), Myschetzky (15), and prolate spheroid (4) volume estimation techniques. All four methods were evaluated in terms of accuracy with respect to a ground truth estimate of the prostate volume obtained via expert radiologist derived segmentations of the prostate on individual 2D slices.

MATERIALS AND METHODS

Data Description and Notation

The datasets considered in this study comprised 19 1.5 Tesla (T) MRI studies obtained from the American College Of Radiology Imaging Network trial (24) and 26 3T T2-weighted MRI studies from the Beth Israel Medical Center in Boston, henceforth denoted as D_1 and D_2 respectively. A complete description of the 45 MRI datasets considered in this study is provided in Table 1. The volume estimation for the ellipsoid method is denoted as V_{Ell}, the Myschetzky method as V_{Mys}, and the prolate spheroid method as V_{Sph}. The MFA-based segmentation method yields a volume estimation V_{MFA} and the expertly determined volume estimation is referred to as V_{Ex}.

Ground Truth Estimations of Prostate Volume

The ground truth volume (V_{Ex}) for the prostate in each of the 45 studies was determined as follows. For each study C, an expert radiologist provided a manual segmentation of the prostate for all slices in which the prostate was visible. The set comprising the area estimates of the prostate from all M slices within a single three-dimensional (3D) study C, is denoted as $S_{Ex} = \{A_m, \mid m \in \{1, \ldots, M\}\}$ where A_m denotes the segmented area of 2D slice m. The estimated prostate areas (region contained within the manual delineations of the capsule) on all slices are integrated and multiplied by the slice interval T. This is similar to the approach presented elsewhere (5), in which planimetry area estimates were aggregated to estimate the prostate volume. The ground truth prostate volume (V_{Ex}) in C is then calculated as

$$V_{Ex} = T \cdot \sum_{m=1}^{M} A_m. \quad (1)$$

Clinically Employed Prostate Volume Estimation Models

For the ellipsoid, Myschetzky, and prolate spheroid models, an expert manually determined the transverse (D_1), craniocaudal

TABLE 1. Data Description

Dataset	Field Strength	Total Studies	Slice (M) per Study	X-Y Dimensions		T (mm)
				Pixels	mm	
D_1	1.5 Tesla	19	$10 \leq M \leq 17$	256×256	140×140	3.0
D_2	3.0 Tesla	26	$8 \leq M \leq 20$	512×512	140×140	2.2

(D_2), and anteroposterior (D_3) lengths of the prostate for each of the 45 studies. The formulae for the clinically employed volume estimation techniques are given in Table 2.

MFA-based Prostate Volume Estimations (V_{MFA})

The MFA is a novel extension of the traditional ASM (18), but uses multiple texture features to characterize the prostate border. The MFA contribution comprises the 5 main steps, which are summarized in Figure 1. A statistical shape model comprising 100 landmark points is generated, and a statistical appearance model comprised of 5 texture features drives the segmentation towards the prostate border. The final set of landmarks on the prostate border is used to estimate the area A of each slice.

Generating a statistical shape model. For each slice from each training image, 100 landmarks are manually placed along the prostate border. $\mathbf{X}$ is used to represent a series of 100 x and y Cartesian coordinates, so that $\mathbf{X} \in \mathbb{R}^{200}$. Principal component analysis is performed on all $\boldsymbol{X}$ (18), so that the shape of the prostate can be characterized by 10 parameters, in turn explaining 98% of the variation seen in the prostate shape. This 10-dimensional vector of parameters $\mathbf{b} \in \mathbb{R}^{10}$ can now be employed to describe a specific shape $\mathbf{X_b}$. The details of this system are shown in the Appendix, which can be accessed online. Figure 1 shows several different prostate shapes obtained by varying values of b in the respective module.

Feature extraction. A series of 50 kernels is convolved over each training image to generate a series of texture features. The 50 kernels comprised Gaussian filter responses with varying strengths and localized standard deviations with varying neighborhood sizes. Following the application of these feature operators, each voxel c is now associated with a 50-dimensional feature vector, $\mathbf{G}(c) \in \mathbf{R}^{50}$.

Generate an appearance model. The traditional ASM methodology used neighboring intensities around each of the 100 landmark points to describe the appearance of the prostate border. In this extension, the distribution of texture features $\mathbf{G}$ (instead of intensities) at each landmark is modeled using a sum of multiple Gaussian distributions (23).

Forward feature selection. A feed forward feature selection is employed for identification of only the most discriminatory textural attributes that are to be used in conjunction with the MFA. Thus while 50 texture features are initially generated, only a few (ie, 5) discriminating ones are selected for use in conjunction with the final appearance model.

TABLE 2. Enumeration of Prostate Volume Estimation Techniques Employed in this Article with Corresponding Formulae

Experiment	Description	Model	V
ϵ_1	Ellipsoid	*Ell*	$D_1 \cdot D_2 \cdot D_3 \cdot \pi/6$
ϵ_2	Myschetzky	*Mys*	$D_1 \cdot D_2 \cdot D_3 \cdot 0.7$
ϵ_3	Prolate spheroid	*Sph*	$(D_1)^2 \cdot D_2 \cdot \pi/6$
ϵ_4	Multifeature ASM	*MFA*	$T \cdot \sum_{m=1}^{M} A_m$
	Expert	*Ex*	$T \cdot \sum_{m=1}^{M} A_m$

Segmentation using the MFA. Using the trained appearance model, the most probable locations of the prostate border on a new image are determined. The probability of voxel c belonging to landmark point n is denoted by $P_n(c)$. The locations $\widehat{\mathbf{X}}$ that had the highest probability of corresponding to a landmark point were thus selected. Finally, $\mathbf{b}$ was modified to optimally fit $\widehat{\mathbf{X}}$ as per the ASM fitting technique in (18), yielding a final set of landmarks $\mathbf{X_b}$. The landmarks $\mathbf{X_b}$ are connected via linear interpolation, and the segmented area can then be determined. The area inside the segmentation for each slice m is given as A_m. Equation 1 is then used to calculate V_{MFA} from the segmentations of all slices in a given 3D volumetric acquisition.

Experiments Performed

Table 2 summarizes the experiments and formulations for all of the volume estimation techniques employed in this article. Four experiments (ϵ_1–ϵ_4) were performed. Experiments ϵ_1–ϵ_3 consist of estimating the prostate volume using the current clinical models (ellipsoid, Myschetzky, spheroid), and the volume estimated from the MFA is denoted as ϵ_4. For ϵ_4 all volumes were translated to a common coordinate frame by clicking on the center of the prostate in a single midgland slice for each volume. This took approximately 1 second per study, and constituted the initialization (and the only place involving user interaction) of the MFA. All Cartesian coordinates (for both the landmarks $\mathbf{X}$ and for x_c and y_c in $\mathbf{G}$) were in reference to this centroid. To run the ASM, the mean shape was placed on top of a new image, and a 50×50 region was searched for the prostate border around each landmark.

For ϵ_4, a fivefold cross validation across patients was performed. To perform the cross-validation, 4/5 of the studies were randomly selected, and used to train an MFA. Then,

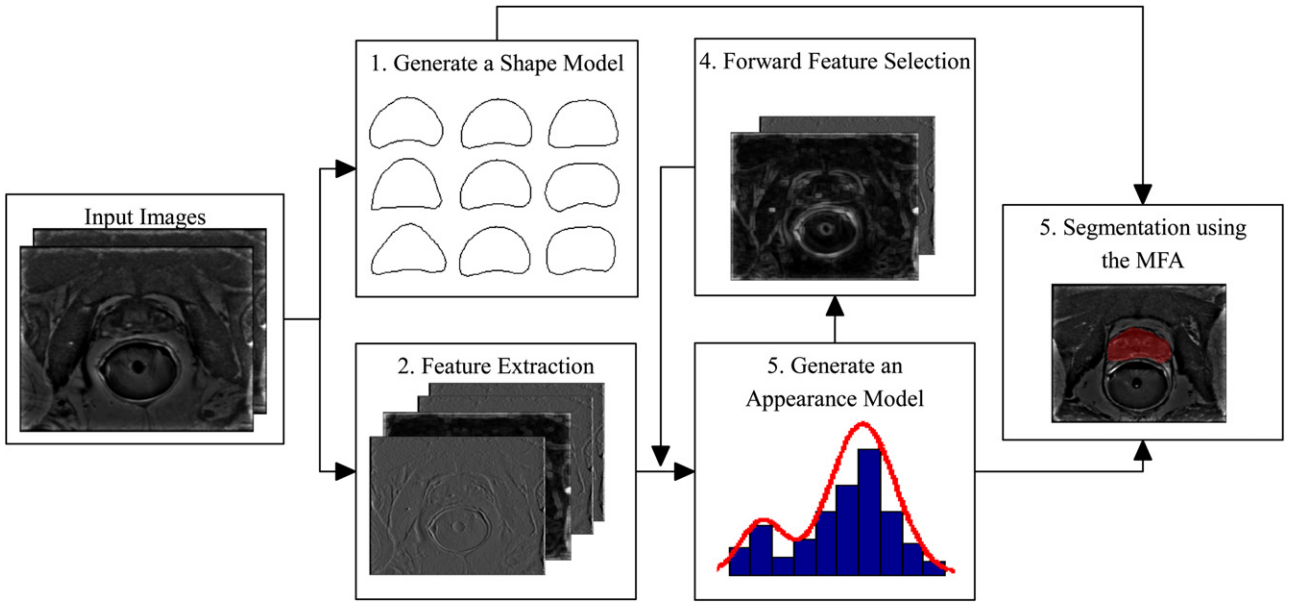

Figure 1. Modules and pathways comprising our multifeature active shape model (ASM) segmentation scheme for prostate volume estimation on endorectal magnetic resonance imaging.

the remaining 1/5 of the studies (which were not used to train) were segmented using the trained MFA. This was repeated until all studies had been tested. Thus the same studies were never used to train and test simultaneously. It should be noted that due to the extreme differences in image quality and structural information, separate cross validations were performed for D_1 and D_2, respectively.

RESULTS

Pearson's Correlation Coefficient

We first compared V_{MFA} with the clinical models V_{Ell}, V_{Mys}, and V_{Sph} for the 45 volumetric studies. This was done by calculating the Pearson correlation coefficient (25) (the r^2 value) between each of V_{MFA}, V_{Ell}, V_{Mys}, and V_{Sph} with V_{Ex} over all 45 testing studies. The hypothesis for these experiments was that V_{MFA} should have at least as high an r^2 value as V_{Ell}, V_{Mys}, and V_{Sph} with respect to V_{Ex}. The results of these experiments are shown in Figure 2 and Table 3. The highest r^2 value (0.82) was obtained between V_{MFA} and V_{Ex}.

Comparison of Volume Fractions

The volume fraction between V and V_{Ex} was calculated for each of the 45 studies in which a value of 1.00 indicates that the estimated volume is exactly equal to the ground truth volume. The results from these calculations are shown in Table 4. The ellipsoid, Myschetzky, and prolate spheroid methodologies had mean volume fractions of 1.14, 1.53, and 1.96. respectively, with standard deviations of 0.25, 0.34, and 0.59, respectively over 45 studies.

The MFA (ϵ_4) had a volume fraction of 1.05 with a standard deviation of 0.21, and is shown in Figures 3d-h. The capsule segmentations obtained via a traditional intensity-driven ASM (18) are shown in Figures 3b and 3f. In Figure 3, the aggregation of segmentations from all slices in a study from D_1 (a-d) and a study from D_2 (e-h) are shown in green, with the aggregation of ground truth segmentations shown in red. The traditional intensity-driven ASM consistently detected edges within the prostate (being affected by noise and local minima), and thus severely undersegmented the gland. However, the MFA as compared to the traditional intensity driven ASM was able to, in most cases, accurately determine the correct prostate boundary and hence prostate volume. This is attributable to the use of multiple, optimally identified textural features in the MFA which are robust to local noise and minima, unlike image intensities.

Statistical Significance Between Volume Fractions

The MFA had a mean volume fraction V_{MFA}/V_{Ex} closest to 1.00, and a paired Student t-test was performed to determine if this was statistically better than each of the other volume fractions (V_{Ell}/V_{Ex}, V_{Mys}/V_{Ex}, V_{Sph}/V_{Ex}). The null hypothesis, therefore, was that the mean volume fractions of the MFA (V_{MFA}/V_{Ex}) and the other methodologies were equal. The results shown in Table 5 illustrate that V_{MFA}/V_{Ex} was statistically significantly different from V_{Ell}/V_{Ex} to a $P < .05$ level of significance, and from V_{Mys}/V_{Ex} and V_{Sph}/V_{Ex} to a $P < .01$ level of significance.

DISCUSSION

An automatic and reproducible method for estimating the volume of the prostate from in vivo T2-weighted MRI data using MFA. The MFA incorporated multiple statistical texture features including the Kirsch, Sobel, Gaussian, and

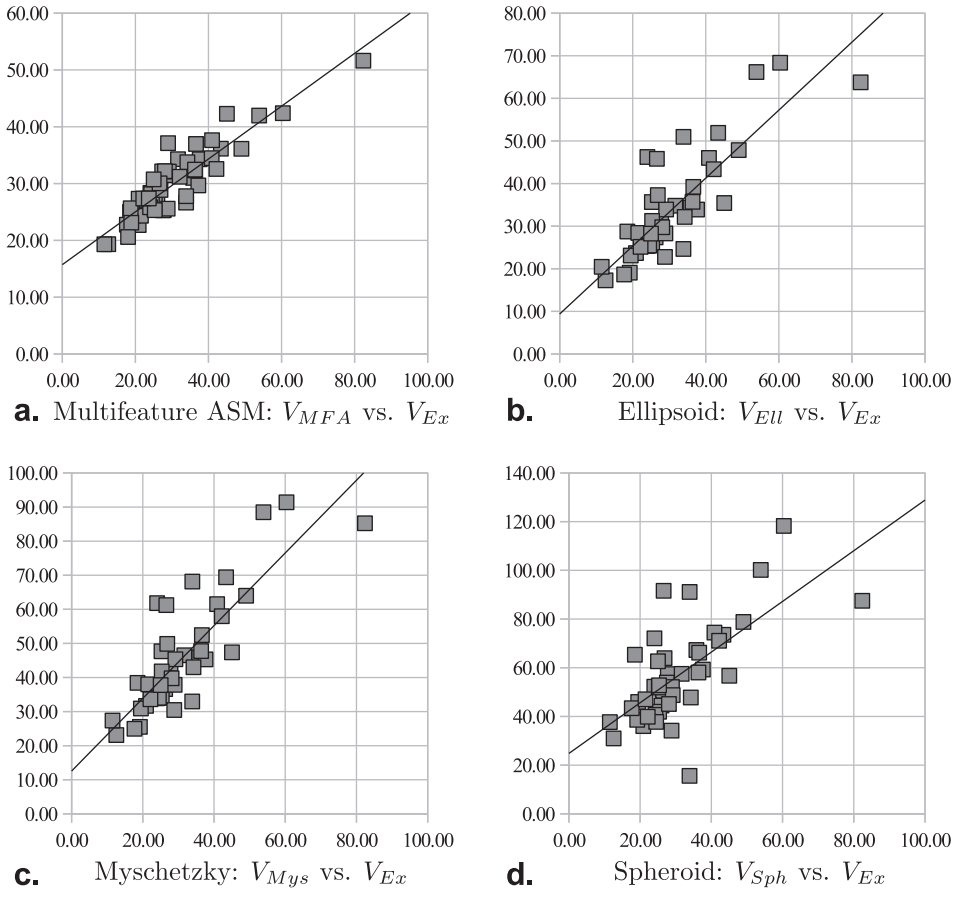

Figure 2. **(a–d)** The correlation plots of the volume estimates V_{MFA}, V_{Ell}, V_{Mys}, and V_{Sph}, respectively, versus the expertly determined ground truth volume estimation V_{Ex} (X-axis), where each data point on each graph represents a single patient study. The axes are represented in milliliters.

TABLE 3. Pearson's Correlation Coefficient (r^2) between V and V_{Ex} over 45 Studies

Model	V_{Ell}	V_{Mys}	V_{Sph}	V_{MFA}
r^2	0.700	0.700	0.454	0.823

TABLE 4. Comparison of V/V_{Ex} in Terms of Mean, Standard Deviation (SD), and Standard Error (STE) over 45 Studies

Experiment	V	Mean	SD	STE
ϵ_1	V_{Ell}	1.143	0.252	0.0376
ϵ_2	V_{Mys}	1.528	0.337	0.0502
ϵ_3	V_{Sph}	1.958	0.587	0.0875
ϵ_4	V_{MFA}	1.053	1.207	0.0277

mean intensity kernels to better characterize the prostate border. In addition, GMMs were used to model the distribution of texture features instead of a traditional single Gaussian, and a forward feature selection algorithm only retained the optimal features for prostate segmentation.

The MFA-based segmentation scheme had a higher correlation with the ground truth ($r^2 = 0.82$) compared to such traditional schemes as the ellipsoid ($r^2 = 0.70$), Myschetzky ($r^2 = 0.70$), and prolate spheroid ($r^2 = 0.45$) models. It was to be expected that the ellipsoidal and Myschetzky have the same r^2 value because they are simply scaled variants of each other. In addition, the prolate spheroid expectedly performed the worst of the clinical estimation techniques, as it only used two axes in its volume estimation while the ellipsoidal and Myschetzky used measurements from three axes.

Qualitative results also revealed that our MFA was able to easily out-perform the traditional, intensity-driven ASM. Reasons for this include the use of textural features such as image gradients, which are not prone to intensity artifacts such as bias field. In addition, the distribution of features is not necessarily best modeled as a Gaussian, which the traditional ASM assumes. In the MFA model, we use GMMs to model our distribution of features, which can capture non-Gaussian shapes of distributions. Finally, we only retain the optimal features in our appearance model, thus automatically discounting noisy and nondiscriminatory features.

That the ellipsoidal estimation performed better than the Myschetzky estimation in terms of volume fractions was not surprising because the Myschetzky correction aims to increase the ellipsoidal model's volume estimation. This would only be useful if the ellipsoidal volume estimation happened to underestimate the capsule's volume. However, unlike has been previously reported in the literature for TRUS imagery (6,8,11,15), in our study the ellipsoidal formula did not underestimate the prostate volume, and hence no Myschetzky correction was

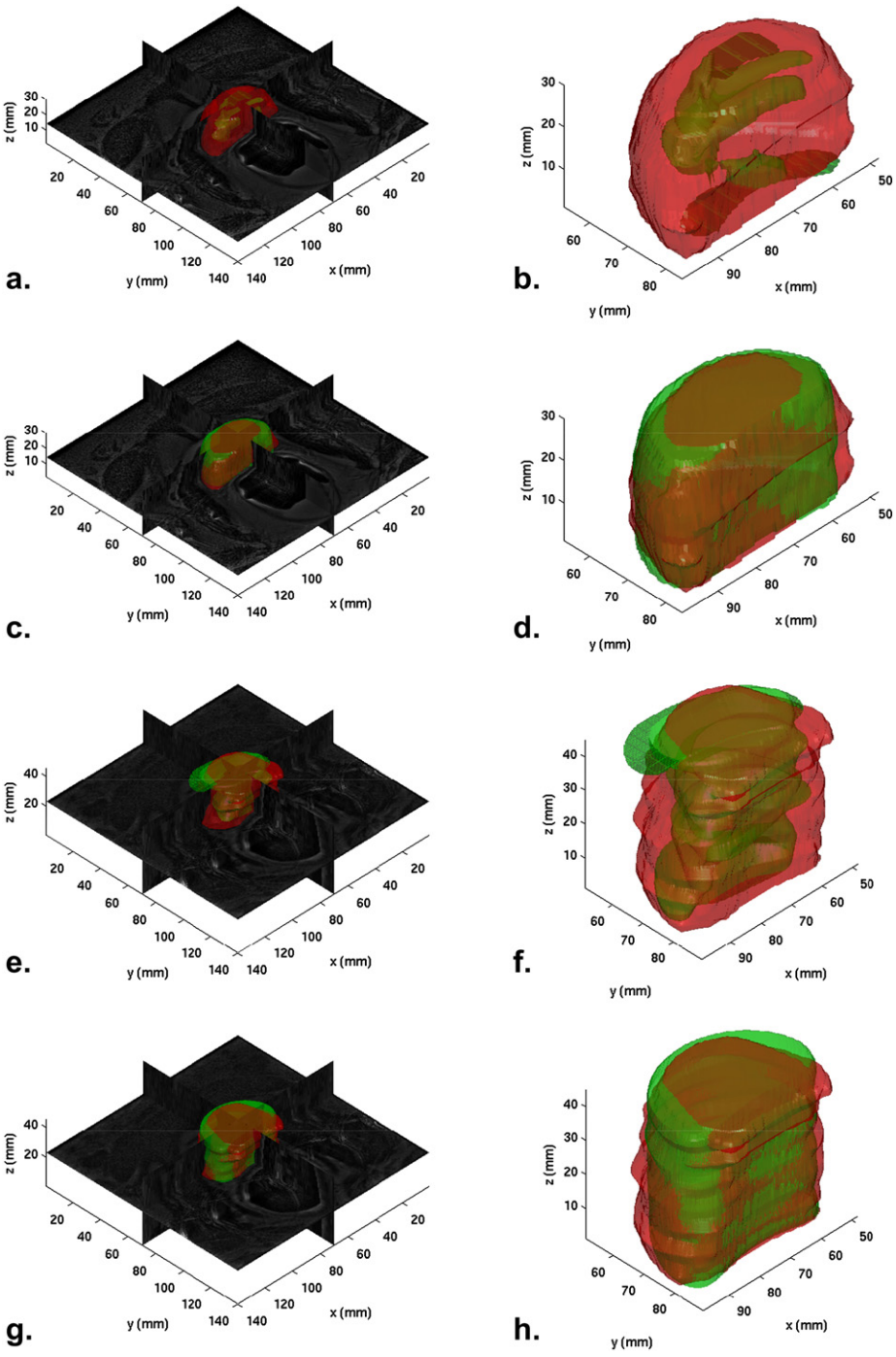

Figure 3. Images of the resulting volume from a study from D_1 **(a–d)** and a study from D_2 **(e–h)**. In each image, the *red* represents the ground truth and the *green* represents the aggregation of segmentations for each slice. For each result, we show the volumes overlaid on the T2-weighted magnetic resonance image (*column 1*), and just the volumes themselves (*column 2*). To compare our results with the traditional ASM, **(a,b,e,f)** show a traditional intensity-driven active shape model (ASM), whereas **(c,d,g,h)** show V_{MFA}.

necessary. This is perhaps because MRI was used instead of the noisier and lower quality TRUS imagery (11). The MFA volume fractions were found to be statistically different from the volume fractions estimated by the geometric models and the intensity driven ASM (not shown).

In addition, Table 4 shows that all of the methods overestimated the volume. The most probable explanation for the clinical methods overestimating the volume is that the craniocaudal axis D_2 in the clinical models encompassed more slices than were manually segmented by the expert to estimate V_{Ex}. However, it should be noted that the MFA's mean volume fraction V_{MFA}/V_{Ex} was just barely over 1.0 (1.05). In fact a value of 1.0 would fall well within half of a standard deviation of 1.05, suggesting that the MFA yielded a very accurate segmentation.

Our 2D MFA was used to generate a segmentation of the visible gland on each slice of a 3D dataset. Because the MFA models the object border using a multidimensional distribution, a large number of training images are required for accurate model generation. A 2D MFA is employed on account of the limited number of 3D studies would prevent accurate statistical models from being generated in 3D (26). However, although we only had access to a limited number

TABLE 5. *P* Values from Each Set of Paired Student *t*-tests Between V_{MFA}/V_{Ex} and V_{Ell}/V_{Ex}, V_{Mys}/V_{Ex}, V_{Sph}/V_{Ex} over 45 Studies

	V_{Ell}/V_{Ex}	V_{Mys}/V_{Ex}	V_{Sph}/V_{Ex}
V_{Mys}/V_{Ex}	4.33×10^{-2}*	2.01×10^{-12}**	7.85×10^{-16}**

*$P < .05$.
**$P < .01$.

of 3D studies, they constituted a total of 690 2D image slices, which was more than sufficient to generate accurate statistical models in 2D (26).

That the geometric based models V_{Ell}, V_{Mys}, V_{Sph} performed significantly worse compared to the MFA based volume estimations V_{MFA} was most likely because of two factors. The first potential cause is that the very tip of the base and the very tip of the apex of the prostate were used to estimate the geometric models (D_2 includes this), whereas the expert segmentations may not have necessarily included the extreme tips of the prostate. This would have yielded much higher volume estimations from the geometric models compared to the surrogate ground truth volume estimations. The second possible reason is that the geometric models are inherently convex, whereas the prostate may have distinctly concave regions. Although the ASM would be able to model these non-convex regions (Fig 3d-h), the geometric models would naturally overestimate the prostate volume in the vicinity of these concavities. Note that almost any segmentation scheme (not just ASMs) would find it difficult to correctly identify the prostate boundary on the extreme base and apical sections (27). Other researchers have also acknowledged difficulties associated with prostate segmentation at gland base and apex.

We also evaluated our MFA scheme in terms of segmentation accuracy by calculating the Dice similarity coefficient (28). The MFA achieved a mean Dice similarity coefficient of 0.8483 with a standard deviation of 0.0448 (standard error 0.0060) over 45 studies. This compared favorably against the prostate MRI segmentation models in (29,30–32). In addition, none of these schemes explicitly computed volume or compared their results against clinical models. Furthermore, our MFA scheme can run in real time, requiring an average of 55 seconds per volume (approximately $512 \times 512 \times 20$ voxels) on a 2.66 GHz, 4GB PC running Matlab under Linux.

Limitations of this study include the fact that expert segmentations were used as a surrogate of the ground truth volume. An alternative would have been to use the volume of the excised prostatectomy specimen as the gold standard, but these were not available for this study. A secondary, minor limitation was the limited number of 3D studies (preventing the use of a full 3D ASM). In summary, the MFA volume estimation method can save valuable time for clinicians and can yield a consistently accurate, near realtime prostate volume estimation that is extremely useful for evaluating post-treatment response.

ACKNOWLEDGMENTS

This work was made possible via grants from the Wallace H. Coulter Foundation, Aresty Foundation for Undergraduate Research, New Jersey Commission on Cancer Research, National Cancer Institute (Grant Nos. R01CA136535-01, ARRA-NCl-3 R21CA127186, R21CA127186, R03CA128081-01, and R03CA143991-01), and The Cancer Institute of New Jersey.

REFERENCES

1. Pierorazio P, Kinnaman M, Wosnitzer M, et al. Prostate volume and pathologic prostate cancer outcomes after radical prostatectomy. Urology 2007; 70:696–701.
2. Kaminski J, Hanlon A, Horwitz E, et al. Relationship between prostate volume, prostate-specific antigen nadir, and biochemical control. Int J Radiat Oncol Biol Phys 2002; 5:888–892.
3. Roehrborn C, Boyle P, Bergner D, et al. 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. Adult Urol 1999; 54:662–670.
4. Bangma C, Niemer A, Grobbee D, et al. Transrectal ultrasonic volumetry of the prostate: in vivo comparison of different methods. Prostate 1966; 28: 107–110.
5. Hoffelt S, Marshall L, Garzotto M, et al. A comparison of CT scan to transrectal ultrasound measured prostate volume in untreated prostate cancer. Int J Radiat Oncol Biol Phys 2003; 57:29–32.
6. Eri L, Thomassen H, Brennhovd B, et al. Accuracy and repeatability of prostate volume measurements by transrectal ultrasound. Nature 2002; 5:273–278.
7. Littrup P, Williams C, Egglin T, et al. Determination of prostate volume with transrectal us for cancer screening. Part 11. Accuracy of in vitro and in vivo techniques. Radiology 1991; 52:49–53.
8. Matthews G, Motta J, Fracchia J. The accuracy of transrectal ultrasound prostate volume estimation: clinical correlations. J Clin Ultrasound 1996; 24:501–505.
9. Nathan M, Seenivasagam K, Mei Q, et al. Transrectal ultrasonography: why are estimates of prostate volume and dimension so inaccurate? Br J Urol 1996; 77:401–407.
10. Park S, Kim J, Choi S, et al. Prostate volume measurements by TRUS using heights obtained by transaxial and midsagittal scanning: comparison with specimen volume following radical prostatectomy. Korean J Radiol 2000; 1:110–113.
11. Tewari R, Indudhara K, Shinohara E, et al. Comparison of transrectal ultrasound prostatic volume estimation with magnetic resonance imaging volume estimation and surgical specimen weight in patients with benign prostatic hyperplasia. J Clin Ultrasound 1996; 24:169–174.
12. Jeong CW, Park HK, Hong SK, et al. Comparison of prostate volume measured by transrectal ultrasonography and MRI with the actual prostate volume measured after radical prostatectomy. Urol Int 2008; 81:179–185.
13. Sosna J, Rofsky N, Gaston S, et al. Determinations of prostate volume at 3-tesla using an external phased array coil: comparison to pathologic specimens. Acad Radiol 2003; 10:846–853.
14. MacMahon P, Kennedy A, Murphy D, et al. Modified prostate volume algorithm improves transrectal us volume estimation in men presenting for prostate brachytherapy. Radiology 2009; 250:273–280.
15. Myschetzky P, Suburu R, Kelly BJ, et al. Determination of prostate gland volume by transrectal ultrasound: correlation with radical prostatectomy specimens. Scand J Urol Nephrol Suppl 1991; 137:107–111.
16. Bonilla J, Stoner E, Grino P, et al. Intra- and interobserver variability of MRI prostate volume measurements. Prostate 1997; 31:98–102.
17. Aarnink R, de la Rosette J, Debruyne F, et al. Reproducibility of prostate volume measurements from transrectal ultrasonography by an automated and a manual technique. Br J Urol 1996; 78:219–223.
18. Cootes T, Taylor C, Cooper D, et al. Active shape models - their training and application. Comput Vision Image Understanding 1995; 61:38–59.
19. Seghers D, Loeckx D, Maes F, et al. Minimal shape and intensity cost path segmentation. IEEE Trans Med Imaging 2007; 26:1115–1129.
20. Madabhushi A, Udupa J. Interplay of inhomogeneity correction and intensity standardization in MR image analysis. IEEE Trans Med Imaging 2005; 24:561–576.

21. Kirsch R. Computer determination of the constituent structure of biological images. Comp Biomed Res 1971; 4:315–328.
22. Duda R, Hart P. Pattern Classification and Scene Analysis. John Wiley and Sons, 1973.
23. Everitt B, Hand D. Finite Mixture Distributions. London, UK: Chapman and Hall, 1981.
24. American College of Radiology Imaging Network. MR imaging and MR spectroscopic imaging of prostate cancer prior to radical prostatectomy: a prospective multi-institutional clinicopathological study. Available online at: www.acrin.org/Portals/0/Protocols/6659/A6659partial_summary.pdf
25. Edwards A. An Introduction to Linear Regression and Correlation. San Francisco: W.H. Freeman and Co Ltd, 1976; 33–46.
26. Ledoita O, Wolf M. A well-conditioned estimator for large-dimensional covariance matrices. J Multivariate Anal 2004; 88:365–411.
27. Mahdavi S, Morris W, Spadinger I, et al. 3d prostate segmentation in ultrasound images based on tapered and deformed ellipsoids. In: MICCAI, vol. 12, Lecture Notes Comput Sci 960–967.
28. Dice L. Measures of the amount of ecologic association between species. Ecology 1945; 263:297–302.
29. Klein S, van der Heide U, Lips I, et al. Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. Med Phys 2008; 35:1407–1417.
30. Martin S, Daanen V, Troccaz J. Automated segmentation of the prostate in 3d MR images using a probabilistic atlas and a spatially constrained deformable model. Med Phys 2010; 37:1579–1590.
31. Pasquier D, Lacornerie T, Vermandel M, et al. Automatic segmentation of pelvic structures from magnetic resonance images for prostate cancer radiotherapy. Int J Radiat Oncol Biol Phys 2007; 68: 592–600.
32. Makni N, Puech P, Lopes R, et al. Combining a deformable model and a probabilistic framework for an automatic 3D segmentation of prostate on MRI. Int J CARS 2009; 4:181–188.
33. Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res 2003; 3:1157–1182.

APPENDIX. MULTIFEATURE ACTIVE SHAPE MODEL

Input Training Images

We first introduce some active shape model (ASM)-specific notations. Each two-dimensional (2D) section of a three-dimensional (3D) T2-weighted magnetic resonance imaging (MRI) image is defined as an image scene I = (*C*, *f*), where *C* represents a 2D grid of spatial locations (pixels) and f(c) is the MRI intensity function associated with every pixel $c \in C$. We define $\mathbf{X} = \{c_1 \ldots c_N\} \subset C$ as the set of *N* landmarks used to define the prostate shape. The mean Cartesian landmark coordinates across all training images are denoted as $\overline{\mathbf{X}} = \{\bar{c}_1, \ldots, \bar{c}_N\}$. The κ-neighborhood of pixels surrounding each $c \in C$ is denoted as $\mathrm{n}_\kappa(c)$, where for $\forall\, d \in \mathrm{n}_\kappa(c),\ \| d\text{-}c \|_2 \leq \kappa, c \notin \mathrm{n}_\kappa(c)$.

Generate a Shape Model

An ASM (29) is defined by the equation,

$$\mathbf{X} = \overline{\mathbf{X}} + \mathbf{P} \cdot \mathbf{b}, \tag{A.1}$$

where $\overline{\mathbf{X}}$ represents the mean shape, **P** is a matrix of the first few principal components (Eigenvectors) of the shape obtained via principal component analysis (PCA), and **b** is a vector defining the shape, where the individual elements of **b** can range between –3 and +3 standard deviations from the mean shape $\overline{\mathbf{X}}$. In our training stage, we have equally spaced 100 landmarks ($N = 100$) along the prostate boundary in each slice, and have aligned the landmarks by selecting the topmost landmark in each image as landmark #1.

Extract Features

In this work, each image is convolved with multiple kernels to derive statistical texture descriptors of the object boundary. The set of kernels is denoted as $S_\mathrm{k} = \{\mathrm{k}_1, \ldots, \mathrm{k}_{14}\}$, where $\mathrm{k}_1 - \mathrm{k}_4$ denote the four directional Kirsch kernels (21), $\mathrm{k}_5 - \mathrm{k}_8$ denote the four directional Sobel kernels (22), $\mathrm{k}_9 - \mathrm{k}_{11}$ denote Gaussian kernels with standard deviations $\sigma \in \{0.5, 1, 5\}$, and $\mathrm{k}_{12} - \mathrm{k}_{14}$ denote mean kernels with $\kappa \in \{5, 10, 25\}$ pixel neighborhood sizes. Examples of a Kirsch (k_1) and Sobel (k_5) kernel are,

$$\mathrm{k}_1 = \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}, \mathrm{k}_5 = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}.$$

Denoting the neighborhood surrounding *c* as $g = \{f(d) | d \in \mathrm{n}_\kappa(c)\}$ and the convolution operator as, the feature vector **G**(c) associated with each $c \in C$ is defined

$$\mathbf{G}(c) = \{\mathrm{g}(c) \otimes \mathrm{k}_1, \ldots, \mathrm{g}(c) \otimes \mathrm{k}_{14}, x_c, y_c\}. \tag{A.2}$$

In addition, we found that including the *x* and *y* Cartesian coordinates of *c* into **G** greatly improved the accuracy of our segmentations.

Generate an Appearance Model

For a given landmark point, we denote the set of feature vectors from *T* training images as $\boldsymbol{F} = \{G_1, \ldots, G_\mathrm{T}\}$. We therefore have *T* data points, each with dimensionality |G|. We model this distribution as a linear combination of *Q* Gaussians (known as a Gaussian mixture model, or GMM), by maximizing the log-likelihood of the model. If *p* denotes a normal distribution with mean $\mu \in \mathbb{R}^{|\mathrm{G}|}$ and covariance $\Sigma \in \mathbb{R}^{|\mathrm{G}| \times |\mathrm{G}|}$, then our set of Q distributions is defined by $\mu = \{\mu_1, \ldots, \mu_\mathrm{Q}\}$, $\Sigma = \{\Sigma_1, \ldots, \Sigma_\mathrm{Q}\}$ and $\mathrm{w} = \{w_1, \ldots, w_Q\}$ where *w* denotes a weight parameter. We therefore use the expectation maximization algorithm (23) to maximize the log likelihood, thereby defining our appearance model as

$$\mu, \Sigma, \mathrm{w} = \underset{\mu, \Sigma}{\operatorname{argmax}}, \mathrm{w} \sum_{t=1}^{T} \left(\ln \sum_{q=1}^{Q} w_q \cdot p(\mathbf{G}_t | \mu_q, \sum_q)\right). \tag{A.3}$$

Segmenting an Image Using the MFA

We begin with the mean shape (ie, $\mathbf{X}^0 = \overline{\mathbf{X}}$). Given a current shape X^i for iteration *i* and a set of locations $\widehat{\mathbf{X}}$ located closer to the border, the shape X^{i+1} is then calculated using Equation A.) where

$$\mathbf{b} = \mathbf{P}' \cdot \left(\widehat{\mathbf{X}}^i - \mathbf{X}^i\right) \tag{A.4}$$

Therefore, the goal is to determine locations closer to the prostate border using the appearance model. The first step is to extract the set of features from the test image. So each location *c* has a feature vector **G**(c) associated with it. Although $p(\mathbf{G}|\mu, \Sigma)$ represents the probability given a single Gaussian, we let $\mathrm{P}(\mathbf{G}|\mu, \Sigma, \mathrm{w})$ represent the probability given the entire GMM. We therefore define our locations closer to the prostate border $\widehat{\mathbf{X}}$ as,

$$\widehat{\mathbf{X}} = \left\{e_n | e_n = \underset{d \in \mathrm{n}_\kappa(c)}{\operatorname{argmax}} P(\mathbf{G}(d) | \mu, \Sigma, \mathrm{w}), n \in \{1, \ldots, N\}\right\} \tag{A.5}$$

In our system, we found that using 5 Gaussians (Q = 5) yielded accurate results without containing too many free parameters. In addition, we let $\kappa = 15$ mm so that at each iteration the shape can move up to 15 mm, which was sufficient to converge to the maximum probability in a short time. It is important to note that when including the Cartesian coordinates x_c and y_c in **G** we always search for the global maximum, as the global maximum of $P(\mathbf{G})$ is inherently limited by the Cartesian coordinates within the image. Therefore, initialization is irrelevant as the entire image can theoretically be searched. We only included a limit on κ for efficiency. In comparison, the traditional ASM essentially uses the same

objective function, but with g instead of **G** and $Q = 1$ instead of $Q = 5$.

Forward Feature Selection

Our segmentation algorithm was run through a cross-validation scheme using a forward feature selection (33) to choose which dimensions of **G** (ie, which features) to include. This was done by running a fivefold cross-validation on a subset of the training data for each feature individually. The best performing feature was then selected using the mean Dice similarity coefficient (DSC) (28).

Then, each other feature was subsequently tested in a five-fold cross-validation scheme. The feature which improved the DSC the most was then selected. This was repeated until a maximum DSC value was achieved. In the experiments we performed, the DSC value was maximized after four features were selected. For the first feature selected, $\mu \in \mathbb{R}^1$, and $\Sigma \in \mathbb{R}^2$, for the second feature, $\mu \in \mathbb{R}^2$ and $\Sigma \in \mathbb{R}^{2\times 2}$, and for our data, $\mu \in \mathbb{R}^4$ and $\Sigma \in \mathbb{R}^{4\times 4}$.