

Spatially Aware Expectation Maximization (SpAEM): Application to Prostate TRUS Segmentation

Mahdi Orooji^{a,†}, Rachel Sparks^b, B. Nicolas Bloch^c, Ernest Feleppa^d, Dean Barratt^b, Anant Madabhushi^{a,‡}

[†]mahdi.orooji@case.edu, [‡]anant.madabhushi@case.edu

^aCenter for Computational Imaging and Personalized Diagnostics, Case Western Reserve University,

^bCentre for Medical Image Computing, University College of London

^cDepartment of Radiology, Boston Medical Center & Boston University,

^dLizzi Center for Biomedical Engineering, Riverside Research

ABSTRACT

In this paper we introduce Spatially Aware Expectation Maximization (SpAEM), a new parameter estimation method which incorporates information pertaining to spatial prior probability into the traditional expectation-maximization framework. For estimating the parameters of a given class, the spatial prior probability allows us to weight the contribution of any pixel based on the probability of that pixel belonging to the class of interest. In this paper we evaluate SpAEM for the problem of prostate capsule segmentation in transrectal ultrasound (TRUS) images. In cohort of 6 patients, SpAEM qualitatively and quantitatively outperforms traditional EM in distinguishing the foreground (prostate) from background (non-prostate) regions by around 45% in terms of the Sorensen Dice overlap measure, when compared against expert annotations. The variance of the estimated parameters measured via Cramer-Rao Lower Bound suggests that SpAEM yields unbiased estimates. Finally, on a synthetic TRUS image, the Cramer-Von Mises (CVM) criteria shows that SpAEM improves the estimation accuracy by around 51% and 88% for prostate and background, respectively, as compared to traditional EM.

1. INTRODUCTION

Image clustering or partitioning schemes have attempted to model an image as comprising of a finite number of distinct mixture distributions.¹ Each mixture component typically corresponds to a distinct class e.g. cancer versus non-cancer or prostate versus background. Each classes has (usually) unknown statistical parameters (such as mean and variance) which may be estimated via a dedicated training set. The objective of this work is estimating the parameters of different image classes based on the set of extracted features and via the introduction of a new spatial prior probability term. Parameters estimated within the prostate and background regions can be used to separate out the individual distributions from the mixture model.

Parameter estimation plays an important role in a wide number of image processing and computer vision tasks such as kernel optimization,² classification,³ segmentation,⁴ and object detection.⁵ Approaches for estimating class parameters encompass a wide spectrum of techniques and have been widely considered in literature.^{1,5-7} Assuming distinct classes represent different segments in an image, parameter estimation can be used for image segmentation. The estimation can be performed either prior to image segmentation, such as Expectation-Maximization (EM) or during the segmentation, such as K-Nearest Neighbors.⁸ In these segmentation approaches, the accuracy of estimation directly affects the accuracy of segmentation.

Maximum-likelihood (ML) has been widely used to formalize the parameter estimation problem. EM has been widely applied to iteratively solve ML problem.⁸ The performance of EM is good provided the probability of observing features' values of a pixel are independent of its spatial information, which is not true for most applications.⁹ Carson *et al.* invoked spatial information by including Cartesian information of a pixel via the observation matrix.⁹ Markov Random Field (MRF)-EM approaches^{10,11} consider neighborhood constraints by incorporating local spatial constraints into the EM frameworks. However, MRF-EM needs precise estimation of the Markov chain transient matrix which is a challenging task.¹²

In this paper, we introduce new spatial information, namely *spatial prior probability*. Most current parameter estimation paradigms assume that all pixels have the same contribution in estimating the parameters associated with the unknown class distributions. However, if a pixel more likely belongs to a specific class (e.g. prostate in TRUS image), then the approximate spatial location of the pixel can be used to enhance the accuracy of

the parameter estimation process. For instance, a transrectal ultrasound image of the prostate will more than likely involve the prostate in the center of the image with the regions corresponding to the borders representing the background. For a specific known category of images where domain information is available, the use of the spatial prior probability can be a very important complement to parameter estimation techniques.

In order to evaluate the newly introduced SpAEM scheme, we employed it to distinguish the foreground (prostate) from the background in Transrectal Ultrasound (TRUS) imagery. SpAEM is used to estimate the parameters of background and prostate, following which the estimated parameters are employed to separate out the two classes. Prostate segmentation plays a key role in different stages of clinical decision making process. Prostate volume can be directly determined from prostate gland segmentation and can be used for diagnosis of benign prostate hyperplasia.¹³ In addition, prostate gland segmentation facilitates multimodal image fusion for tumor localization in biopsy and radiation therapy.¹⁴ Due to the importance in delineating the prostate on TRUS, semi and fully automated prostate segmentation methods for TRUS imagery have been developed. For a review of some of these techniques we refer the reader to¹⁵ and the references therein.

Despite recent advances in transducer design, resulting in improved spatial and temporal resolution, TRUS image segmentation is still heavily dependent on the quality of data.^{16,17} Additionally, TRUS segmentation is vulnerable to a variety of artifacts, such as, different levels of signal attenuation, shadowing artifacts, and speckle. Low contrast between areas of interest is another difficult problem in prostate segmentation on TRUS.¹⁸

Previous prostate segmentation algorithms for TRUS images have focused on the use of prior knowledge such as prostate shape¹⁹ or prior spatial probability.²⁰ Recently, 2D and 3D boundary extraction methods have been developed based on probabilistic data association filters.¹⁷ However, these approaches assume prior information for the prostate shape (such as concavity) and/or need manually placed initial seed locations inside the prostate.²¹ Recently, active contour-based approaches have been used for prostate segmentation. These methods assumed prior spatial properties,¹⁰ or need some initial points on²² or near²³ the prostate boundary. These methods have also been employed to fuse multi-modal images of the prostate.²⁴ Our SpAEM scheme is a fully automated segmentation algorithm that uses only the spatial prior probability for parameter estimation. The estimated parameters are then used to distinguish the prostate from the background.

In this work, first we calculate spatial prior probability from a dedicated training set. Then we modify the traditional EM framework to invoke a spatial prior probability associated with every location in the image. To do so, a new E-step and M-step are introduced. The new framework is applied to TRUS images to estimate parameters corresponding to the prostate as well as background. Estimated parameters are used to separate out prostate from the background. In addition, to demonstrate the accuracy of SpAEM, we employ Cramer-Rao lower bound (CRLB)²⁵ and Cramer-Von Mises (CVM) criterion.²⁶ CLRB expresses the lower bound on the variance of estimated parameters. Experimental results show that the new method converges to CRLB. CVM is a criterion to measure the goodness of fit of estimated values to their corresponding theoretical estimates. Experimental results also demonstrate that the newly introduced method outperforms traditional EM.

The rest of the paper is organized as follows. In Section 2, we present our assumptions and introduce relevant notation. We introduce our new parameter estimation framework in Section 3. Section 4 describes our experimental design for evaluating our system and Section 5 presents our experimental results; We provide concluding remarks in Section 6.

2. FRAMEWORK AND NOTATION

2.1 Problem Statement

An image I is modeled as a mixture of finitely many distributions, each with a different set of parameters. Those parameters are usually unknown and may be estimated. Each distribution represents a distinct class, for example prostate versus background. In this work we estimate the parameters of K different image classes based on L extracted features and the available spatial prior information.

Notation	Description	Notation	Description
I	Image scene.	$\Theta = \{\mathcal{M}, \Lambda\}$	Set of actual parameters.
N	Total number of pixels.	$\hat{\Theta} = \{\hat{\mathcal{M}}, \hat{\Lambda}\}$	Set of estimated parameters.
L	Total number of extracted features.	$\bar{\mu}_k = [\mu_{k,l}]_{1 \times L}$	Vector of actual means of k th class.
K	Total number of classes.	$\hat{\mu}_k = [\hat{\mu}_{k,l}]_{1 \times L}$	Vector of estimated means of k th class.
$D_{N \times L}$	Matrix of observations.	$\Sigma_k^2 = [\sigma_{k,i,j}^2]_{L \times L}$	Set of actual covariance matrix of k th class.
\bar{d}_n	Features vector of n th pixel.	$\hat{\Sigma}_k^2 = [\hat{\sigma}_{k,i,j}^2]_{L \times L}$	Set of estimated covariance matrix of k th class.
$\Xi = [\xi_{n,k}]_{N \times K}$	Spatial prior probability matrix.	$I_f(\theta)$	Fisher Information of parameter θ .

Table 1. Description of notation used throughout this paper.

2.2 Notation

Suppose N denotes the total number of pixels of I . Observation matrix $D_{N \times L} = [\bar{d}_1, \bar{d}_2, \dots, \bar{d}_N]^T$ comprises all image features such that a vector $\bar{d}_n = \{d_{n,1}, d_{n,2}, \dots, d_{n,L}\}$ for $n \in \{1, 2, \dots, N\}$ is a feature vector corresponding to n th pixel. I is desired to be segmented into K distinct classes of, S_1, S_2, \dots, S_K such that the union of all classes cover the entire image, i.e., $\bigcup_{k=1}^K S_k = I$ and $\bigcap_{k=1}^K S_k = \emptyset$ and each pixel belongs to only one class.

The set of parameters of all classes (which is unknown) is denoted by $\Theta = \{\mathcal{M}, \Lambda\}$. $\mathcal{M}_{K \times L} = [\bar{\mu}_1, \bar{\mu}_2, \dots, \bar{\mu}_K]^T$ is the set of means of all classes. k th member of \mathcal{M} is $\bar{\mu}_k = \{\mu_{k,1}, \mu_{k,2}, \dots, \mu_{k,L}\}$ which is the mean of the k th class. Similarly, $\Lambda^2 = [\Sigma_{1L \times L}^2, \Sigma_{2L \times L}^2, \dots, \Sigma_{KL \times L}^2]$ is the set of covariance matrices of all classes and $\Sigma_k^2 = [\sigma_{k,i,j}^2]$ is the covariance matrix of k th class. Table 1 lists the notation used throughout this paper.

2.3 Spatial Prior Probability

In this work we introduce *Spatial Prior Probability* matrix. For an image, there is a corresponding row with K entities in the spatial prior probability matrix. In this matrix, k th elements of n th row expresses the prior probability that n th pixel belongs to k th class. In conventional parameter estimation approaches, all pixels have the same weight in estimating the parameters. This assumption regarding an equal contribution from all pixels in parameter estimation is questionable. It certainly would make sense to additionally weight the contributions of those pixels which have a higher likelihood of belonging to the class of interest. By including spatial prior probability matrix in parameter estimation, we are weighting the contribution of pixels based on their prior probabilities.

Let us $\Xi_{N \times K} = [\xi_{n,k}]$ denotes spatial prior probability. Note that $\sum_{k=1}^K \xi_{n,k} = 1$, for any given n . Ξ is calculated from a set of J distinct training studies. Let us set $g_{n,j}^{(k)} = 1$ if n th pixel of j th training image belongs to the k th class. So, each element of the spatial prior probability matrix is calculated by averaging over all training images as:

$$\xi_{n,k} = \frac{1}{J} \sum_{j=1}^J g_{n,j}^{(k)}, \quad \text{for } k \in \{1, 2, \dots, K\}. \quad (1)$$

As previously discussed in Section 1, for a number of different settings it is quite likely that one would know where specific target classes (e.g prostate on a TRUS image) are most likely to occur. Additionally this information could be learned from training data. Figure 1 demonstrates an example of spatial prior probability for the prostate, learned from a set of 43 pre-segmented TRUS images.

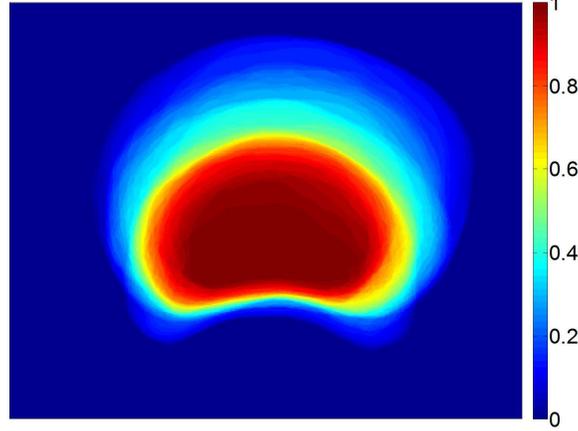


Figure 1. Spatial prior probability of prostate segment in TRUS image, calculated by averaging over 43 pre-segmented prostate masks using equation (1).

2.4 Feature Set Probability

Without loss of generality, we assume the distribution of extracted features may be accurately modeled as a multivariate Gaussian distribution.²⁰

Matrix of Classes: To make the feature set parameters estimation problem tractable, we define an auxiliary (latent) variable $Z_{N \times K} = [z_{n,k}]$, such that any n th row of Z has only one element of 1 at k th column which implies n th pixel of I belongs to S_k , and the rest of the elements of the n th row are zero.

Considering the assumptions and defined variables, the conditional probability of observation matrix D can be calculated by marginalizing over random variable Z as follows,

$$P(D|\Theta, \Xi) = \sum_{\forall Z} P(D|Z, \Theta, \Xi)P(Z|\Theta, \Xi), \quad (2)$$

where the conditional distribution of the observation matrix given Θ and Z is given by:

$$P(D|Z, \Theta, \Xi) = \prod_{n=1}^N P(\bar{d}_n|Z, \Theta) = \prod_{n=1}^N \prod_{k=1}^K \left[\frac{1}{(2\pi)^{L/2} |\Sigma_k|} \exp \left((\bar{d}_n - \bar{\mu}_k)^T \Sigma_k^{-2} (\bar{d}_n - \bar{\mu}_k) \right) \right]^{z_{n,k}}. \quad (3)$$

The matrix of classes is independent of the distribution parameters and its probability value is a function of spatial prior probability matrix as:

$$P(Z|\Theta, \Xi) = P(Z|\Xi) = \prod_{n=1}^N \prod_{k=1}^K [\xi_{n,k}]^{z_{n,k}}. \quad (4)$$

One should note that for other feature distributions rather than Gaussian (such as the distribution of the intensity in TRUS images which is Rayleigh distribution²⁷), equation (3) needs to be substituted by a proper corresponding distribution function.

3. PARAMETER ESTIMATION USING SPATIALLY AWARE EXPECTATION MAXIMIZATION (SpAEM)

SpAEM is expressed by,

$$\hat{\Theta} = \arg \max_{\Theta} P(D|\Theta, \Xi) = \arg \max_{\Theta} \log \left(\sum_{\forall Z} P(D|Z, \Theta)P(Z|\Xi) \right). \quad (5)$$

Spatial prior probability is expected to improve the parameter estimation for k th class by, first excluding pixels for which $\xi_{n,k} = 0$, and then increasing the contribution of pixels with large $\xi_{n,k}$.

By substituting values from (3) and (4) into (5) one can show that the ML problem does not have an analytical solution. Besides, the order of the computational complexity for the numerical solution is $O(K^N)$ which not tractable. For example, parameter estimation of two classes in a 256×256 image needs calculating the likelihood value of approximately $10^{20,000}$ possible instances of Z .

EM is a powerful iterative algorithm to estimate the parameters of the mixture models when the associated log-likelihood maximization problem is too complicated to solve analytically. To estimate the parameters, we introduce Spatially Aware Expectation-Maximization (SpAEM) to iteratively solve (5). We modified the two steps of Expectation and Maximization as follows:

Expectation Step:

Current and revised estimation of Θ are denoted by Θ^{OLD} and Θ^{NEW} , respectively. The conditional expectation of $\log P(D, Z|\Theta)$ given D and the current estimation of Θ is given by,

$$\begin{aligned} Q(\Theta; \Theta^{\text{OLD}}) &= E_Z [\log P(D, Z|\Theta)] \\ &= \sum_{n=1}^N \sum_{k=1}^K E_Z [z_{n,k} | D, \Theta^{\text{OLD}}] \left(\log \xi_{n,k} - \frac{L}{2} \log 2\pi - \log |\Sigma| + (\bar{d}_n - \bar{\mu}_k)^T \Sigma_k^{-2} (\bar{d}_n - \bar{\mu}_k) \right) \end{aligned} \tag{6}$$

where $E_Z[.|x]$ is the conditional expectation given x , respect to z . After some manipulation one can show that:

$$\gamma_{n,k} \triangleq E_Z [z_{n,k} | D, \Theta^{\text{OLD}}] = P(z_{n,k} = 1 | \bar{d}_n, \Theta^{\text{OLD}}) = \frac{\xi_{n,k} P(\bar{d}_n | z_{n,k}, \Theta^{\text{OLD}})}{\sum_{k=1}^K \xi_{n,k} P(\bar{d}_n | z_{n,k}, \Theta^{\text{OLD}})} \tag{7}$$

Maximization Step:

It has been shown in⁸ that the log-likelihood function is the monotonic increasing function of the number of EM iteration steps. So, regardless of the initial value of Θ , EM converges. Θ^{NEW} refers to the solution of:

$$\Theta^{\text{NEW}} = \arg \max_{\Theta} Q(\Theta, \Theta^{\text{OLD}}) \tag{8}$$

By calculating the partial derivative of (6) and (7) respect to $\bar{\mu}_k$ and Σ_k , one can show that

$$\hat{\mu}_{k,l} = \frac{\sum_{n=1}^N \gamma_{n,k} d_{n,l}}{\sum_{n=1}^N \gamma_{n,k}}, \tag{9}$$

$$\hat{\sigma}_{k,l}^2 = \frac{\sum_{n=1}^N \gamma_{n,k} (d_{n,l} - \mu_{k,l})^2}{\sum_{n=1}^N \gamma_{n,k}} \tag{10}$$

4. EXPERIMENTAL DESIGN

To evaluate the proficiency of SpAEM approach, first, we applied SpAEM for prostate segmentation in TRUS imagery, and compare the results with traditional EM. Second, we evaluate the accuracy of the SpAEM parameter estimation in compare to traditional EM.

4.1 Dataset

The presented approach was applied on TRUS images for six patients. TRUS images were acquired using a bi-planer side-firing transrectal probe. Prostate boundary on TRUS was manually delineated by an expert radiologist. Pixels closer to the probe appear brighter in ultrasound images. To correct this attenuation artifact, similar to²⁰, we employed a smoothing Gaussian kernel filter and divided the intensity values of the image by the output of the filter. The output of this filter transform is an intensity-corrected image.

4.2 Image Segmentation

We assumed there are two classes in a TRUS image, one is the prostate and the other one is background. We employed SpAEM to distinguish the two segments. Prostate segmentation in TRUS is challenging task due to low contrast, low SNR, high speckle, micro-calcification, and high imaging artifacts shadowing.¹⁷

Prostate segmentation consists of three steps.

Step-1: For a TRUS image, I , extract L features to create observation matrix D .

Step-2: Using SpAEM to estimate mean and variance of two classes corresponding to prostate ($\hat{\mu}_1, \hat{\Sigma}_1^2$) and background ($\hat{\mu}_2, \hat{\Sigma}_2^2$).

Step-3: Classify n th pixel as either prostate ($\hat{k} = 1$) or background ($\hat{k} = 2$) from following equation.

$$\hat{k} = \arg \max_k \xi_{n,k}^\alpha P^{1-\alpha}(\bar{d}_n | \hat{\mu}_k, \hat{\Sigma}_k^2). \quad (11)$$

In (11) a degree of freedom, α , is a weighting factor to adjust the effect of spatial prior probability relative to probability of observed vector. Note that $P(\bar{d}_n | \hat{\mu}_k, \hat{\Sigma}_k^2)$ is multiplication of L distribution functions, so, $\alpha = \frac{L}{L+1}$ implies the weight of spatial prior probability is identical to each distribution function. Algorithm. 1 demonstrates SpAEM algorithm for prostate segmentation. Note that, different stopping criteria have been previously proposed,²⁸ mainly by comparing the difference between estimated parameters of two consecutive iterations with some threshold. In this work, we compare the difference between two consecutive estimations of $\hat{\mu}_{1,1}$ with the value of 10^{-3} .

Data: Observation matrix $D_{N \times L}$, and Spatial Prior Probability Matrix $\Xi_{N \times K}$

Result: Estimating the parameters $\hat{\Theta}$, and estimating the class of each pixel \hat{k}

Estimate the parameters, $\hat{\Theta}$, using EM-Algorithm:

begin

 Assume an initial value for Θ^{old} ;

while *Stopping criterion is not satisfied*, **do**

 E Step: Find $\gamma_{n,k}$ using (7);

 M Step: Estimate $\Theta^{\text{new}} = [(\hat{\mu}_{k,l}, \hat{\sigma}_{k,l}^2)]$ for $1 \leq k \leq K$ and $1 \leq l \leq L$ using (9);

end

end

$\hat{\Theta} = \Theta^{\text{new}}$;

Prostate segmentation using $\hat{\Theta}$:

begin

 For n th pixel, $1 \leq n \leq N$, estimate \hat{k} using (11)

end

Algorithm 1: Prostate segmentation using SpAEM

4.2.1 Employed Image Features

A set of intensity and texture features are used for prostate segmentation. In this paper, we employed 7 features described in Table 2, in which i_n and \mathcal{N}_n denote the intensity value and the neighborhood of the n th pixel, respectively. The proficiency of these features in distinguishing prostate tissue from non-prostate tissue has already been investigated in.²⁰

Feature	Description	Formulation
Intensity	Intensity value	i_n for $1 \leq n \leq N$
Mean	Average intensity value within \mathcal{N}_n .	$\bar{i}_n = \frac{1}{ \mathcal{N}_n } \sum_{m \in \mathcal{N}_n} i_m$
Median	Median intensity value within \mathcal{N}_n .	$\text{median}(i_m)_{m \in \mathcal{N}_n}$
Range	Range of intensity values within \mathcal{N}_n .	$\max_{m \in \mathcal{N}_n} (i_m) - \min_{m \in \mathcal{N}_n} (i_m)$
Variance	Variance of a Gaussian distribution of intensity values within \mathcal{N}_n .	$\sqrt{\frac{1}{ \mathcal{N}_n } \sum_{m \in \mathcal{N}_n} (i_m - \bar{i}_n)^2}$
Rayleigh Variance ²⁹	Variance of the Rayleigh distribution calculated within \mathcal{N}_n .	$\sqrt{\frac{1}{2 \mathcal{N}_n } \sum_{m \in \mathcal{N}_n} i_m^2}$
m -parameter ³⁰	The m -parameter of the Nakagami distribution, which controls the shape of the distribution, calculated within \mathcal{N}_n .	We use the method of Greenwood and Durand ³¹ to estimate m .

Table 2. Description of intensity and texture features used for prostate segmentation in TRUS.

4.2.2 Evaluation of Segmentation Performance

As a qualitative metric, we employed Sorensen-Dice coefficient³² to compare prostate segmentation results. The set of pixels segmented as being part of the prostate is denoted by \mathcal{E} . Segmentation results are compared to manual delineations performed by an expert which serves as ground truth for segmentation evaluation. For each image, the set of pixels lying within the manual delineation of prostate is denoted by \mathcal{G} . Sorensen-Dice coefficient is given by,

$$\frac{2|\mathcal{E} \cap \mathcal{G}|}{|\mathcal{E}| + |\mathcal{G}|} \quad (12)$$

where $|\cdot|$ denotes the cardinality of a set.

4.3 Qualitative Metrics to Assess the Accuracy of Parameter Estimation

4.3.1 Cramer-Rao Lower Bound (CRLB)

CRLB expresses the lower bound on the variance of estimated parameters. provided the estimator is unbiased.²⁵ Suppose $\hat{\theta}$ denotes the unbiased estimation of parameter θ , the minimum achievable variance of $\hat{\theta}$ given θ is

$$\text{var}\{\hat{\theta} \mid \theta\} \geq 1/I_f(\theta), \quad (13)$$

where

$$I_f(\theta) = E \left[\frac{\partial^2}{\partial \theta^2} \log p(D|\theta) \right] \quad (14)$$

is the Fisher Information.³³ From (13) and (14) and after some algebraic manipulation one can show that the CRLB of the mean and the variance of k th class is given by:

$$\text{var}\{\hat{\mu}_{k,l} \mid \mu_{k,l}\} \geq \sigma_{k,l}^2 / \mathcal{N}_k \quad (15)$$

$$\text{var}\{\hat{\sigma}_{k,l}^2 \mid \sigma_{k,l}^2\} \geq 2\sigma_{k,l}^4 / \mathcal{N}_k \quad (16)$$

where \mathcal{N}_k is the number of pixels belonging to k th class.

4.3.2 Cramer-Von Mises (CVM) Criterion

CVM is a criterion to measure the goodness of fit of theoretical cumulative distribution function (CDF) compared to empirically estimated CDF.²⁶ It is given by:

$$\mathcal{D}_f = \frac{1}{N} \sum_{n=0}^N [\mathfrak{F}_\theta(n) - F_\theta(n)]^2 \mathfrak{P}_\theta(n) \quad (17)$$

where F_θ , \mathfrak{F}_θ , and \mathfrak{P}_θ are theoretical CDF, empirical CDF and empirical PDF, respectively.

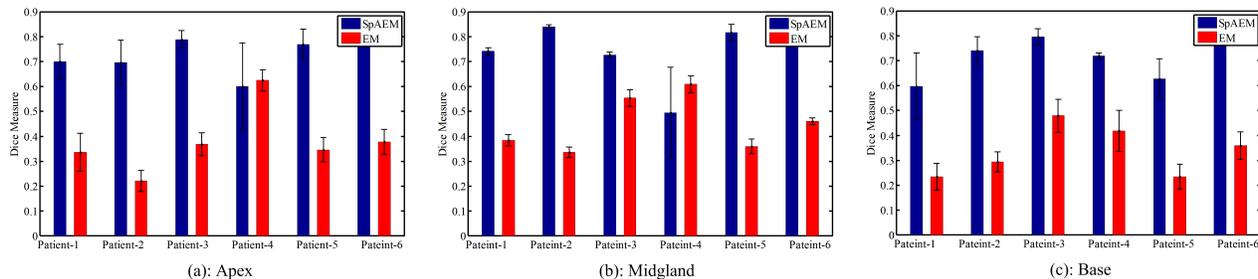


Figure 2. Dice coefficients for prostate segmentation in apex, midland, and base of prostate, respectively, using SpAEM and EM.

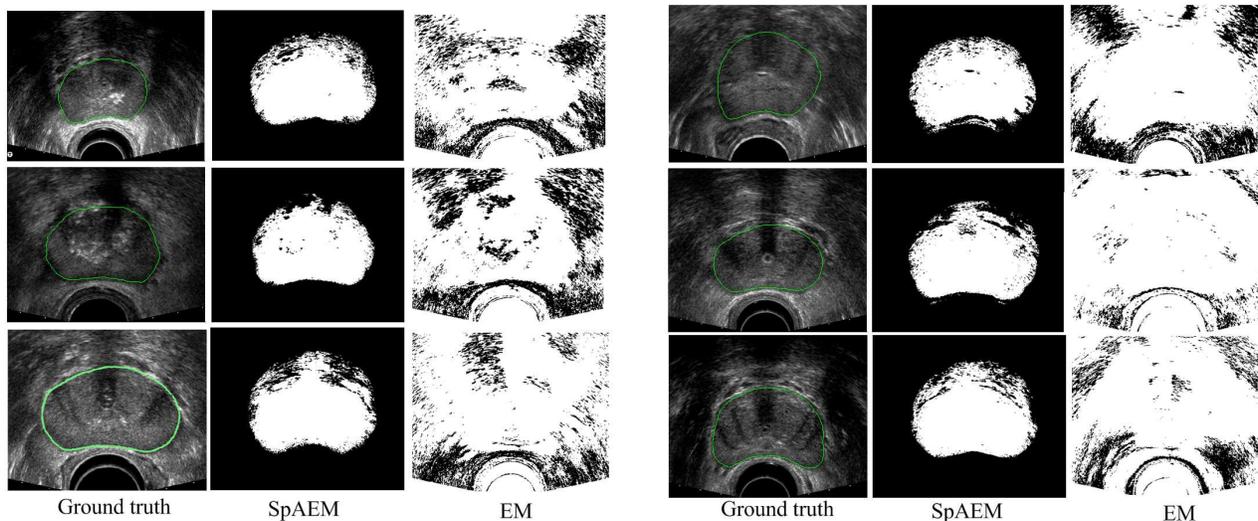


Figure 3. Quantitative results of prostate segmentation from a cohort of 6 patients, using SpAEM and EM parameter estimation.

5. RESULTS

5.1 Prostate Segmentation Results

For prostate segmentation, we employed the set of features described in Section 4.2.1. We also used Sorensen-Dice coefficient explained in Section 4.2.2 to compare prostate segmentation results from our SpAEM method and unsupervised EM Gaussian mixture model (EM-GMM) based parameter estimation. We set $\alpha = L/L + 1$, which implies that the spatial prior probability and each distribution have an identical weight.

Figure 2 shows Sorensen-Dice coefficient of prostate segmentation in (a) apex, (b) midland, and (c) base for 6 patients. As Figure 2 shows, SpAEM outperforms traditional EM approach in apex and midland for 5 out of 6 patients. The performance of SpAEM for all 6 patients is better than that of EM in base.

Figure 3 qualitatively demonstrates the performance of TRUS image segmentation approaches. The lack of prior spatial probability in EM-based segmentation method is evident in Figure 3. Note that EM algorithm is unable to reliably distinguish the background from the foreground and hence over-segments the prostate. However by imposing the spatial prior probability constraint, SpAEM is able to more accurately and specifically segment the foreground from the background.

5.1.1 Parameter Sensitivity

The performance of SpAEM is sensitive to the degree of freedom, α . Figure 4 demonstrates the effect of weighting coefficient α when the intensity value is the only feature employed for segmentation. As Figure 4 shows, the accuracy of SpAEM for prostate segmentation in midland is better than that in apex and base when the value of α is small. On the other hand, for larger values of α , SpAEM shows more accurate segmentation in apex and

base. $\alpha = \frac{L}{L+1} = 0.5$ shows more consistent performance in apex, midglad and base. One can discern from Figure 4 that if 3D information of prostate is available, the performance can be improved by adaptively changing α based on the approximate location of the ultrasound probe.

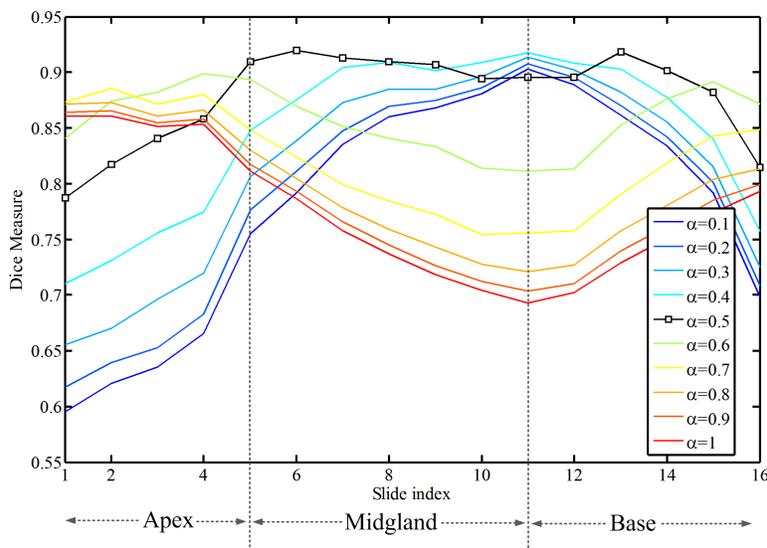


Figure 4. Dice values versus different values of α for segmentation of the prostate in apex, midglad and base when $L = 1$.

5.2 Accuracy of Estimated Parameters

5.2.1 Cramer-Rao Lower Bound

Figure 5 shows the variance of the estimated mean (left) and variance of the estimate variance (right) of prostate versus N_k for different number of features, L . Dotted line in that figure shows the CRLB of the mean estimation (left) and variance estimation (right). The variance of estimated parameters are plotted against the portion of pixels employed for parameter estimation. It is evident, from weak-law of large number theory,³⁴ that employing a larger number of pixels for parameter estimation results in lesser variance of estimated parameters. As Figure 5 illustrates, SpAEM achieves CRLB which implies: (1) SpAEM is an unbiased estimator, and (2) our solution achieves the lowest mean square error. Hence we can surmise that, SpAEM is the minimum variance unbiased (MVU) estimator.

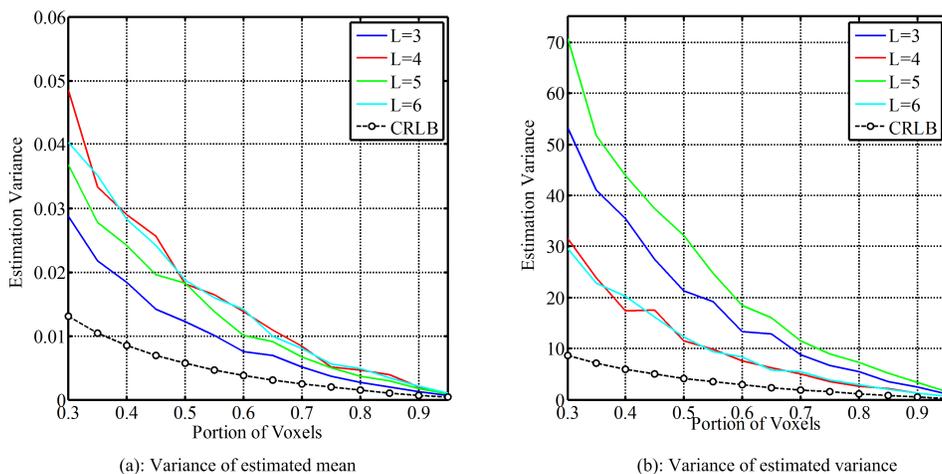


Figure 5. The variance of (a) estimated mean and (b) estimated variance versus the portion of pixels used for parameter estimation.

5.2.2 Cramer-Von Mises Criterion

We wished to measure the goodness of fit of theoretical CDF to SpAEM estimated CDF. We generate a synthetic TRUS image with known mean and variance (of both prostate and background), then we use SpAEM to estimate those parameters. Having estimated parameters, we can write estimated distributions (because mean and variance are sufficient statistics for Gaussian distribution³⁴). Now, we use CVM to measure the difference between estimated distributions and synthetic distributions for both prostate and background. The approach is briefly described in following steps:

Step-1: Generate synthetic TRUS image that satisfies the assumptions in Section 2.

Step-2: Estimate parameters via SpAEM and EM.

Step-3: Calculate CVM for SpAEM and EM using (17).

Figure 6 (a) is synthetic TRUS image when $\mu_1 = 1$, $\mu_2 = 4$ are mean of the intensity of prostate and background, respectively. Corresponding variances are identical and equal 1, i.e. $\Sigma_1^2 = \Sigma_2^2 = 1$. Figure 6 (b) compares CVM for SpAEM and EM. As it is evident, the difference between theoretical CDF and empirical CDF is less in SpAEM in compare to EM.

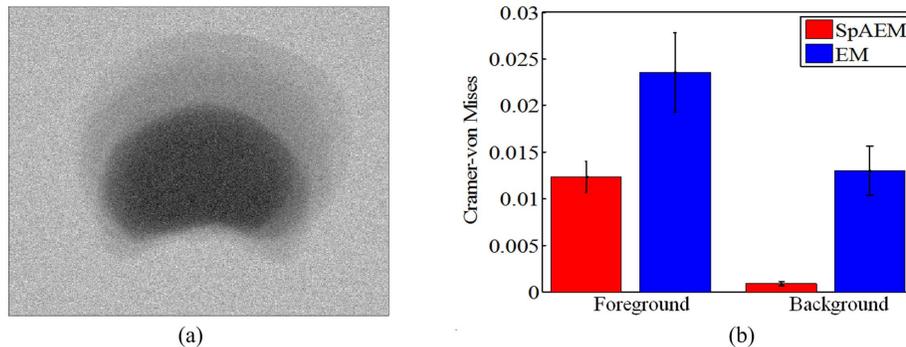


Figure 6. (a) Synthetic TRUS image generated as described in Section 5.2.2, (b) Comparing CVM of prostate and background using SpAEM and EM for estimating the parameters of synthetic TRUS images.

6. CONCLUDING REMARKS

In this work, we introduced a novel Spatially Aware Expectation-Maximization (SpAEM) algorithm for parameter estimation. SpAEM can be applied to those specific scenarios where we need to model image distributions as a mixture of a finite number of distinct distributions with unknown parameters. We defined a prior spatial probability matrix to complement existing parameter estimation methods. There was one-to-one correspondence between each row of prior spatial probability matrix and each pixel of an image, such that k th entity of n th row of prior spatial probability matrix demonstrated the prior probability that n th pixel belonged to k th class. When estimating the parameters for a class, spatial prior probability improved the accuracy of parameter estimation by:

- Increasing the contribution of pixels with higher probability of belonging to the class of interest.
- Excluding pixels for which the probability of belonging to the class of interest is zero.

In order to evaluate SpAEM, we employed it for the task of prostate capsule segmentation in TRUS images. The performance was measured quantitatively via Sorensen-Dice coefficient as well as quantitatively via comparing with ground truth. Experimental results showed SpAEM outperformed traditional EM. We finally used Cramer-Rao Lower Bound and Cramer-Von Mises Criterion as a quantitative metric to demonstrate the

accuracy of introduced parameter estimation scheme. Experimental results showed the proficiency of SpAEM over traditional EM.

Our approach and study did have some limitations. Our results can be improved by first, considering more accurate distribution of features. We assumed the features are independent multivariate Gaussian which is not always true. For example, the intensity value of ultrasound images is governed by a Rayleigh distribution.²⁷ Secondly, an optimization problem for parameter α may also be required to improve the accuracy of the segmentation. And finally, we assumed pixels are independent which is not an accurate assumption for adjacent pixels. As part of future work, we aim to include Markov prior (similar to¹¹) to further boost the parameter estimation.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under award numbers R01CA136535-01, R01CA140772-01, and R21CA167811-01; the National Institute of Diabetes and Digestive and Kidney Diseases under award number R01DK098503-02, the DOD Prostate Cancer Synergistic Idea Development Award (PC120857); the QED award from the University City Science Center and Rutgers University, the Ohio Third Frontier Technology development Grant. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. In addition, data collection and sharing for the Alzheimers Disease Dataset was funded by the Alzheimers Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from a number of associations and companies*.

REFERENCES

- [1] Bock, H. G., Carraro, T., Jager, W., Korkel, S., Rannacher, R., and Schloder, J. P., [*Model Based Parameter Estimation, Theory and Applications*], Springer-Verlag Berlin Heidelberg (2013).
- [2] Lee, H. and Singh, R., "Unsupervised kernel parameter estimation by constrained nonlinear optimization for clustering nonlinear biological data," in [*Bioinformatics and Biomedicine (BIBM), 2012 IEEE International Conference on*], 1–6 (Oct 2012).
- [3] van der Heijden, F., [*Classification, parameter estimation, and state estimation: an engineering approach using MATLAB*], Wiley (2004).
- [4] Lakshmanan, S. and Derin, H., "Simultaneous parameter estimation and segmentation of gibbs random fields using simulated annealing," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **11**, 799–813 (Aug 1989).
- [5] Liang, Z., Jaszczak, R., and Coleman, E., "Parameter estimation of finite mixtures for image processing using the em algorithm and information criteria," in [*Nuclear Science Symposium and Medical Imaging Conference, 1991., Conference Record of the 1991 IEEE*], 2171–2175 vol.3 (Nov 1991).
- [6] Kay, S., [*Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*], Prentice Hall (1993).
- [7] Molina, R., Vega, M., Abad, J., and Katsaggelos, A., "Parameter estimation in bayesian high-resolution image reconstruction with multisensors," *Image Processing, IEEE Transactions on* **12**, 1655–1667 (Dec 2003).
- [8] Bishop, C. M., [*Pattern Recognition and Machine Learning (Information Science and Statistics)*], Springer-Verlag New York, Inc., Secaucus, NJ, USA (2013).
- [9] Carson, C., Belongie, S., Greenspan, H., and Malik, J., "Blobworld: image segmentation using expectation-maximization and its application to image querying," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **24**(8), 1026–1038 (2002).
- [10] Xu, J., Monaco, J. P., and Madabhushi, A., "Markov random field driven region-based active contour model (maracel): application to medical image segmentation," in [*Proceedings of the 13th international conference on Medical image computing and computer-assisted intervention: Part III, MICCAI'10*], 197–204, Springer-Verlag, Berlin, Heidelberg (2010).

*<http://adni.loni.usc.edu/about/funding/>

- [11] Monaco, J., Hipp, J., Lucas, D., Smith, S., Balis, U., and Madabhushi, A., "Image segmentation with implicit color standardization using spatially constrained expectation maximization: Detection of nuclei," in [*Proceedings of the 15th International Conference on Medical Image Computing and Computer-Assisted Intervention - Volume Part I*], MICCAI'12, 365–372, Springer-Verlag, Berlin, Heidelberg (2012).
- [12] Pyun, K., Lim, J., Won, C. S., and Gray, R., "Image segmentation using hidden markov gauss mixture models," *Image Processing, IEEE Transactions on* **16**, 1902–1911 (July 2007).
- [13] Noble, J. and Boukerroui, D., "Ultrasound image segmentation: a survey," *Medical Imaging, IEEE Transactions on* **25**, 987–1010 (Aug 2006).
- [14] Zhu, Y., Williams, S., and Zwigelaar, R., "Computer technology in detection and staging of prostate carcinoma: A review," *Medical Image Analysis* **10**(2), 178 – 199 (2006).
- [15] Shao, F., Ling, K. V., Ng, W. S., and Wu, R. Y., "Prostate boundary detection from ultrasonographic images," *Journal of Ultrasound in Medicine* **22**(6), 605–623 (2003).
- [16] Ghose, S., Mitra, J., Oliver, A., Mart, R., Llad, X., Freixenet, J., Vilanova, J., Comet, J., Sidib, D., and Meriaudeau, F., "A supervised learning framework for automatic prostate segmentation in trans rectal ultrasound images," in [*Advanced Concepts for Intelligent Vision Systems*], Blanc-Talon, J., Philips, W., Popescu, D., Scheunders, P., and Zemk, P., eds., *Lecture Notes in Computer Science* **7517**, 190–200, Springer Berlin Heidelberg (2012).
- [17] Ghose, S., Oliver, A., Mart, R., Llad, X., Vilanova, J. C., Freixenet, J., Mitra, J., Sidib, D., and Meriaudeau, F., "A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images," *Computer Methods and Programs in Biomedicine* **108**(1), pp. 262–287 (2012).
- [18] Hulsmans, F.-j. J. H., Castelijns, J. A., Reeders, J. W. A. J., and Tytgat, G. N. J., "Review of artifacts associated with transrectal ultrasound: Understanding, recognition, and prevention of misinterpretation," *Journal of Clinical Ultrasound* **23**(8), 483–494 (1995).
- [19] Qiu, W., Yuan, J., Ukwatta, E., Tessier, D., and Fenster, A., "Three-dimensional prostate segmentation using level set with shape constraint based on rotational slices for 3d end-firing trus guided biopsy," *Medical Image Computing and Computer-Assisted Intervention MICCAI 2012* **1** (2012).
- [20] Sparks, R., Nicolas Bloch, C., Feleppa, E., Barratt, D., and Madabhushi, A., "Fully automated prostate magnetic resonance imaging and transrectal ultrasound fusion via a probabilistic registration metric," in [*Proceedings of the SPIE*], **8671** (2013).
- [21] Abolmaesumi, P. and Sirouspour, M., "An interacting multiple model probabilistic data association filter for cavity boundary extraction from ultrasound images," *Medical Imaging, IEEE Transactions on* **23**(6), 772–784 (2004).
- [22] Yezzi, A. J., Kichenassamy, S., Kumar, A., Olver, P., and Tannenbaum, A., "A geometric snake model for segmentation of medical imagery," *Medical Imaging, IEEE Transactions on* **16**(2), 199–209 (1997).
- [23] Yu, Y., Molloy, J. A., and Acton, S., "Segmentation of the prostate from suprapubic ultrasound images," *Medical Physics* **1** (2004).
- [24] Mahdavi, S. S., Moradi, M., Morris, W. J., Goldenberg, S. L., and Salcudean, S. E., "Fusion of ultrasound b-mode and vibro-elastography images for automatic 3-d segmentation of the prostate," *Medical Imaging, IEEE Transactions on* **31**, 2073–2082 (Nov 2012).
- [25] Cramér, H., [*Mathematical Methods of Statistics*], Princeton landmarks in mathematics and physics, Princeton University Press (1999).
- [26] Cramr, H., "On the composition of elementary errors," *Scandinavian Actuarial Journal* **1928**(1), 141–180 (1928).
- [27] Goodman, J., [*Speckle Phenomena in Optics: Theory and Applications*], Roberts & Company (2007).
- [28] Naim, I. and Gildea, D., "Convergence of the em algorithm for gaussian mixtures with unbalanced mixing coefficients," in [*Proceedings of the 29th International Conference on Machine Learning (ICML-12)*], Langford, J. and Pineau, J., eds., 1655–1662, ACM, New York, NY, USA (2012).
- [29] Burckhardt, C. B., "Speckle in ultrasound B-mode scans," *IEEE Transactions on Sonics and Ultrasonics* **25**(1), 1–6 (1978).
- [30] Yacoub, M. D., Bautista, J. E. V., and Guerra de Rezende Guedes, L., "On higher order statistics of the Nakagami-m distribution," *IEEE Transactions on Vehicular Technology* **48**(3), 790–794 (1999).

- [31] Greenwood, J. A. and Durand, D., "Aids for fitting the gamma distribution by maximum likelihood," *Technometrics* **2**(1), pp. 55–65 (1960).
- [32] Sørensen, T., [*A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application to Analyses of the Vegetation on Danish Commons*], Biologiske skrifter, I kommission hos E. Munksgaard (1948).
- [33] Edgeworth, F. Y., "On the probable errors of frequency-constants (contd.)," *Journal of the Royal Statistical Society* **71**(3), pp. 499–512 (1908).
- [34] Papoulis, A. and Pillai, S., [*Probability, random variables, and stochastic processes*], McGraw-Hill electrical and electronic engineering series, McGraw-Hill (2002).