Statistics and Its Interface Volume 3 (2010) 513–521

Bayesian co-segmentation of multiple MR images

Jianfeng Xu and Feng Liang*

Segmentation is one of the basic problems in magnetic resonance (MR) image analysis. We consider the problem of simultaneously segmenting multiple MR images, which, for example, can be a series of 2D/3D images of the same tissue scanned over time, different slices of a volume image, or images of symmetric parts. These multiple MR images share common structure information and hence they can assist each other in the segmentation procedure. We propose a Bayesian co-segmentation algorithm where the shared information across multiple images is utilized via a Markov random field prior. An efficient algorithm based on the Swendsen–Wang method is employed for posterior sampling, which is more efficient than the single-site Gibbs sampler. Because our co-segmentation algorithm pulls all the image information into consideration, it provides more accurate and robust results than individual segmentation, as supported by our experimental studies with real examples.

Keywords and phrases: MRI, Co-segmentation, Bayesian, MCMC.

1. INTRODUCTION

A basic problem in magnetic resonance imaging (MRI) is to precisely segment regions of interest (ROI's) from the image data, which is a crucial part of diagnosis, surgery, therapy guidance, and other medical research and applications (Kass et al., 1988; Zou et al., 2001; Frangi et al., 2001). Many segmentation algorithms have been introduced in literature. These existing algorithms can be roughly divided into three categories (Lakare and Kaufman, 2000): the algorithms based on geometric or topological structure models, including Snakes (Kass et al., 1988), Level Set (Yang et al., 2004), and Watershed (Beucher and Meyer, 1993), the algorithms based on statistical models, including the thresholding method (Weszka, 1978), K-means clustering (Wells et al., 1996), Markov random fields based methods (Winkler, 2006), and classification methods, and the algorithms based on hybrid approaches.

Despite such a wide array of literature, MRI segmentation still remains a challenging problem due to the complex structure between ROI and the neighboring parts and imaging artifacts such as noise, motion, contrast, etc. To address these challenges, various methods have been proposed to incorporate more information, either from prior knowledge or from other sources, to improve the accuracy of segmentation. For instance, Cootes et al. (1995) introduced the shape prior into segmentation to keep the deformation of the contour consistent with statistical models from the PCA analysis. Following Cootes' method, several papers have focused on the incorporation of the shape prior information into traditional methods (Yang et al., 2004; Leventon et al., 2000; Tsai et al., 2003). In many situations, however, the use of shape prior models is limited because there are not enough training images to build the prior model. Further, it is not suitable for applications involving a large number of MR images that call for automatic computer-assisted segmentation procedures. Recently, Younis et al. (2007) proposed to combine the information from MRI and MR spectroscopy imaging (MRSI) for segmentation, with the advantage that MRI segmentation can be further corrected or enhanced by MRSI. Such a method, however, is not accessible for single-modality MRI analysis.

In this paper, we propose to incorporate the information from similar images to improve the accuracy of segmentation, instead of replying on information from prior knowledge on the shape of ROI or from other sources (i.e., modalities). We consider the problem of simultaneously segmenting multiple MR images, which, for example, can be a series of MR images scanned over time such as images of liver perfusion or dynamic cardiac motion, spatial slices of a volume, or images of symmetrical tissues such as lung and hippocampus. Due to their similarities, it is beneficial to share the image information with each other when executing segmentation, which we refer to as co-segmentation. Similar problems have been considered by Cheng and Figueiredo (2007). However they treated each image independently and only utilized the spatial information within each image for segmentation.

The term "co-segmentation" was used by others before, but with a slightly different meaning. For example, in Younis et al. (2007), co-segmentation refers to segmenting the brain region based on two modalities, MR and MRSI images. In Rother et al. (2006), co-segmentation refers to extracting a common part (e.g., the foreground) from an image pair, in which what is shared is not the segmentation structure across images but the model for the foreground segment. In other words, some pixels in these two images are assumed to be generated by the same statistical model, but how these pixels are located and how other pixels are generated or located in these two images are totally independent.

*Corresponding author.

In our work, co-segmentation refers to jointly segmenting multiple images of which the segmentation structure (i.e., spatial configuration of the segmentation result) is shared. Note that in our framework, the image data from a shared segment on different images are not necessarily modeled by the same distribution, so our approach is less sensitive to noise and image artifacts. We present a Bayesian co-segmentation procedure, in which the shared structure information across images, as well as the information among neighboring vertices (pixels/voxels) in the same image, is coded in a Markov random field prior. A Swendsen–Wang type algorithm is developed for posterior sampling, which updates a block of vertices simultaneously and is more efficient than the alternative single-site Gibbs sampler.

The remaining sections are organized as follows: Section 2 discusses the model and prior specification; Section 3 presents the Bayesian inference via two Gibbs sampling algorithms; In Section 4, we illustrate the utility of our method on four MRI data sets, and finally close with remarks in Section 5.

2. METHOD

2.1 Gaussian mixture models

For convention, we first introduce some notations. Assume all J MR images have the same size in each of the l dimensions, $(d_1, \ldots, d_l)$, and have totally $n = d_1 \times \cdots \times d_l$ vertices. For example, $l = 2$ for 2D images and $l = 3$ for 3D images. The vertices are often called pixels for 2D images and voxels for 3D images. Denote the intensity measure at vertex i in image j by $Y_{ji} \in \mathbb{R}^p$, for example, an image with RGB format has $p = 3$. In this paper, we focus on gray scale MR images, so $Y_{ji} \in \mathbb{R}$ with $j = 1 : J$ and $i = 1 : n$.

The goal of MRI segmentation is to partition vertices in an image into different clusters, based on the homogeneity of their intensity measures. Suppose there are totally K segments or clusters. For each vertex i in image j, we introduce a latent variable $Z_{ji} \in \{1, \ldots, K\}$. It is a common practice in MRI segmentation to model the intensity measure Y_{ji} in the same cluster with a Gaussian distribution (Wells et al., 1996; Permuter et al., 2006; Lee and Lewicki, 2002),

$$p(Y_{ji} | Z_{ji} = k) = \phi\big(Y_{ji}; \mu_{jk}, \sigma^2_{jk}\big), \tag{1}$$

where $\phi(\cdot; \mu, \sigma^2)$ denotes a normal density function with mean μ and σ^2. Under this framework, segmentation becomes the problem of inferring the latent variables Z_{ji}'s. There are several advantages of this model-based approach. First, it is a soft segmentation approach in the sense that Z_{ji} is not restricted to take one fixed value, but treated as a random variable and allowed to have uncertainty over the K clusters. So it can handle cases that have ambiguities in the structure definition due to the sampling artifacts or poor resolution. Further, such a generative model can be easily extended to a semi-supervised setting where some vertices can have known labels given by experts. At last, with a model-based approach, the selection of hyper-parameters, such as the number of clusters, can be formulated as a model selection problem, thus a range of criteria such as AIC and BIC can be applied.

We write the unknown parameters and latent variables in this model as $(\mathbf{Z}_{\cdot\cdot}, \boldsymbol{\mu}_{\cdot\cdot}, \boldsymbol{\sigma}^2_{\cdot\cdot})$, where the subscript dot is a shorthand notation for the set containing all possible values in that subscript location. In a Bayesian framework, we make our inference of the unknowns based on the posterior distribution

$$\begin{aligned}\pi\big(\mathbf{Z}_{\cdot\cdot}, \boldsymbol{\mu}_{\cdot\cdot}, \boldsymbol{\sigma}^2_{\cdot\cdot} \mid \mathbf{Y}_{\cdot\cdot}\big) \propto \prod_{j=1}^{J}\prod_{i=1}^{n} p\big(Y_{ji} \mid Z_{ji}, \mu_{j\cdot}, \sigma^2_{j\cdot}\big) \\ \times \pi(\mathbf{Z}_{\cdot\cdot}) \times \prod_{j=1}^{J}\prod_{k=1}^{K} \pi\big(\mu_{jk}, \sigma^2_{jk}\big),\end{aligned} \tag{2}$$

where the first line is the likelihood of the data $\mathbf{Y}_{\cdot\cdot}$ and the second line are the prior distributions over the latent variable $\mathbf{Z}_{\cdot\cdot}$ and unknown parameters.

In co-segmentation, prior distributions play a critical role, which will be discussed in the next subsection. The posterior distribution given above is not in closed form and we will employ a MCMC algorithm for posterior sampling, which will be discussed in Section 3.

2.2 Co-segmentation priors

For computation efficiency, we use conjugate normal priors for the cluster mean μ_{jk} and inverse Gamma for variance σ^2_{jk}, namely,

$$\pi(\mu_{jk}) = \mathrm{N}(\mu_0, \tau_0), \qquad \pi\big(\sigma^2_{jk}\big) = \mathrm{InvGa}(\alpha_0, \beta_0).$$

In our empirical experiments, we use default values for these hyper-parameters $(\mu_0, \tau_0, \alpha_0, \beta_0)$, which correspond to non-informative or vague prior choices. For example, in all of our experiments, we set

$$\mu_0 = 0.5, \qquad \tau_0 = 0.1, \qquad \alpha_0 = 1, \qquad \beta_0 = 0.01.$$

In specific clinical scenario, however, the hyper-parameters should be tuned based on the prior knowledge and experience of radiologists, which will make co-segmentation more effective.

In traditional Bayesian mixture models, the latent variables Z_{ji}'s in (1) are modeled as i.i.d. discrete random variables. However, such an independent model does not work well with image data since the spatial dependence among intensity measures Y_{ji}'s from nearby vertices should be incorporated into the model. For example, it is natural to assume that neighboring vertices are likely to belong to the same segment. For co-segmentation, in order to utilize information across images, we further extend the neighborhood on the same image to multiple images (e.g., image pairs, or adjacent images if images are obtained over time). We start

with a graph on the nJ vertices from the J MR images: connect two vertices (ji) and $(j'i')$, if they are neighboring vertices on the same image or vertices on different images but from roughly the same location. For example, we set the neighborhood as $|j-j'| \le 1$ and $|i-i'| \le 1$ in our experiments. Denote all the edges by set E_0, then construct a Potts model on Z_{ji}'s as the following

$$\pi(\mathbf{Z}_{\cdot\cdot}) \propto \exp\left\{ \sum_{(ji)\sim(j'i')} \beta_{(ji)(j'i')} I[Z_{ji} = Z_{j'i'}] \right\} \tag{3}$$

where $I[\cdot]$ is an indicator function, the pair $(ji) \sim (j'i')$ means they are connected by an edge from E_0, and $\beta_{(ji)(j'i')}$ is an edge-dependent tuning parameter representing the interaction strength. Our prior distribution above leads to information sharing over multiple images and meanwhile preserving the spatial structure within each individual image which has been destroyed by vectorization.

Note that if one permutes the labels Z_{ji}'s simultaneously for all images, for example, relabel the 1st cluster as the 2nd and the 2nd as the 1st for all images, the likelihood (1), the prior distribution (3), and therefore the posterior distribution stay the same. This is known as the label-switching issue in Bayesian mixture modeling (McLachlan and Peel, 2000). To make a coherent inference, we need to fix the order of the K clusters on one particular image. In the later analysis, we will order the K clusters on the first image by their cluster means, i.e., $\mu_{11} < \cdots < \mu_{1K}$.

Also note that if one permutes the labels on some (not all) images, the likelihood still stays the same, but the prior and the posterior change, since *a priori* we favor the configuration satisfying $Z_{ji} = Z_{j'i'}$ where (ji) and $(j'i')$ are neighboring vertices on different images. So given a configuration $\mathbf{Z}_{\cdot\cdot}$, it is possible to shuffle the labels on image 2 to image J (the labels on the 1st image have been fixed) to increase the prior (3), and therefore to increase the posterior probability. Finding the optimal order of the labels on the $(J-1)$ images is time-consuming, so instead we give a simple greedy algorithm to find a sub-optimal solution in the Appendix. This shuffling step is not a valid MCMC step, but is used to guide the chain to reach regions with high posterior probabilities, so in the MCMC algorithms in Section 3, this shuffling step is only called in the burning period.

3. MCMC ALGORITHMS

With the Potts prior on $\mathbf{Z}_{\cdot\cdot}$, the posterior distribution (2) is not in closed form. However, the conditional distributions for each parameter or latent variable given others and the data are from known parametric families, which leads to a simple Gibbs sampling scheme given in Section 3.1. We refer to it as a single-site Gibbs sampler (SSGS), since it updates the cluster membership Z_{ji} sequentially over each vertex. A more efficient algorithm that updates a block of Z_{ji}'s based on Swendsen–Wang method is described in Section 3.2, which we refer to as a SW Gibbs sampler (SWGS).

3.1 A single-site Gibbs sampler

In the single-site co-segmentation, all parameters and latent variables are updated sequentially in each iteration as follows.

I. Initialization. Assign initial values for $(\boldsymbol{\mu}_{\cdot\cdot}, \boldsymbol{\sigma}^2_{\cdot\cdot}, \mathbf{Z}_{\cdot\cdot})$ such that $\mu_{11} < \cdots < \mu_{JK}$. Then, execute the following MCMC steps recursively:

II. At the t-th iteration,

1. Update Z_{ji} sequentially for $i = 1:n$ and $j = 1:J$ with

$$P(Z_{ji} = k | \mathbf{Y}_{\cdot\cdot}, \cdots) = \frac{w_k}{\sum_{l=1}^K w_l},$$

where $(\cdots)$ denotes all other parameters and latent variables (except Z_{ji}) evaluated at their current values, and

$$w_k \propto \phi\big(Y_{ji}; \mu_{jk}, \sigma^2_{jk}\big) \times \exp\left\{ \sum_{(j'i')\in\mathcal{N}(ji)} \beta_{(ji)(j'i')} I[Z_{j'i'} = k] \right\}, \tag{4}$$

where the set $\mathcal{N}(ji)$ denotes all the vertices $(j'i')$ that are neighbors of (ji) as defined by the initial edge set E_0.

2. For each cluster $k = 1:K$ on the j-th image $(j = 1:J)$, let $\overline{Y}_{j[Z=k]}$ denote the cluster mean for the k-th cluster and $n_{jk} = \sum_{i=1}^n I[Z_{ji} = k]$ the corresponding cluster size. Update

$$\mu_{jk} | \mathbf{Y}_{\cdot\cdot}, \cdots \sim \mathrm{N}\big(\mu, \tau^2\big)$$

where $\mu = r\overline{Y}_{j[Z=k]} + (1-r)\mu_0$, and

$$r = \frac{\tau_0^2}{\sigma^2_{jk}/n_{jk} + \tau_0^2}, \qquad \frac{1}{\tau^2} = \frac{1}{\sigma^2_{jk}/n_{jk}} + \frac{1}{\tau_0^2}.$$

Then update

$$\sigma^2_{jk} | \mathbf{Y}_{\cdot\cdot}, \cdots \sim \mathrm{InvGa}\bigg(\frac{n_{jk}}{2} + \alpha_0, \frac{1}{2}\sum\nolimits_{i:Z_{ji}=k} \big(Y_{ji} - \mu_{jk}\big)^2 + \beta_0 \bigg).$$

3. Label-switching: relabel the K clusters such that $\mu_{11} < \mu_{12} < \cdots < \mu_{1K}$.

4. Shuffle the labels on image 2 to image J when in the burning period (see Appendix).

3.2 A Swendsen–Wang Gibbs sampler

In SSGS, the labels Z_{ji}'s are updated sequentially. Such a single-site operation is inefficient for segmenting MR images in two aspects: (1) Traversing all vertices sequentially in each iteration is time-consuming due to the high dimension of MR images; (2) It might take many iterations to update a set of coupled vertices (Barbu and Zhu, 2005). In this subsection, we present a more efficient algorithm, which still operates image by image but on each image updates a block of vertices simultaneously.

Let $\mathbf{Z}_{j\cdot}$ denote all the labels on the j-th image. Recall that the joint conditional distribution of $\mathbf{Z}_{j\cdot}$ given the data and other unknowns is

$$\pi(\mathbf{Z}_{j\cdot}|\mathbf{Y}_{\cdot\cdot},\cdots) \propto \exp\left\{\sum_{i=1}^{n} g_{ji}(Z_{ji})\right\} \times \exp\left\{\sum_{(ji)\sim(ji')} \beta_{(ji)(ji')} I[Z_{ji}=Z_{ji'}]\right\}, \tag{5}$$

where

$$g_{ji}(k) = \log \phi\big(Y_{ji};\mu_{jk},\sigma^2_{jk}\big) + \sum_{(ji)\sim(j'i)} \beta_{(ji)(j'i)} I[Z_{j'i}=k].$$

The distribution (5) remains in the Potts family with the likelihood and prior contribution from other images as the external field term, So for each image we can sample a block of coupled Z_{ji}'s simultaneously using the Swendsen–Wang (SW) algorithm.

The SW algorithm, proposed by Swendsen and Wang (1987), is an efficient sampling method for Potts models. Instead of updating the Z values vertex by vertex, it updates the values block by block as follows. Given the current configuration of Z's, connect neighboring vertices with the same Z values with certain probability that depends on β in (5). The connected vertices form disjointed blocks, and then SW algorithm updates the Z values in a block simultaneously. The SW algorithm can be justified by augmenting the space of Z's by bonding variables (Edwards and Sokal, 1988; Higdon, 1997). Recently, Barbu and Zhu (2005) provided another justification for SW from the aspect of Metropolis–Hastings, which leads to the use of SW algorithm for non-Potts models.

Our SW Gibbs sampling algorithm, abbreviated as SWGS, is similar to SSGS given in the previous subsection, except at step II(1) we update $\mathbf{Z}_{j\cdot}$ for $j = 1:J$ as follows. Given the current labeling $\mathbf{Z}_{j\cdot}$ for the j-th image, connect two neighboring vertices (ji) and (ji') that have the same Z value with probability $1-\exp\{-\beta_{(ji)(ji')}\}$ where $\beta_{(ji)(ji')}$ is the edge-dependent interaction parameter as in (3) and (5). Then the vertices on the j-th image are divided into m disconnected components $(V_1,\ldots,V_m)$, where $V_l \cap V_{l'} = \emptyset$ and $\bigcup_{l=1}^{m} V_l = \{1,\ldots,n\}$. For each component V, we update their labels $\mathbf{Z}_{jV}$ simultaneously with

$$P(\mathbf{Z}_{jV}=k|\mathbf{Y}_{\cdot\cdot},\cdots) \propto \exp\left\{\sum_{i\in V} g_{ji}(k)\right\}. \tag{6}$$

Although our SWGS algorithm is designed for cases where the conditional distributions are from the Potts family, it can be easily extended to sample arbitrary posterior distributions. For example, for the bounding box prior in Lempitsky et al. (2009), and the prior used in Barbu and Zhu (2005) that encourages large and connected segments, the corresponding conditional distribution of $\mathbf{Z}_{j\cdot}$ is no longer from the Potts family. Nevertheless, we can still update the labels block by block through a Metropolis–Hastings step. We formulate the disjoint blocks as described before, then for block V, propose to assign $\mathbf{Z}_{jV}$ a new label. The acceptance ratio for the new configuration $\mathbf{Z}'_{j\cdot}$ is given by

$$\alpha(\mathbf{Z}_{j\cdot} \to \mathbf{Z}'_{j\cdot}) = \min\left\{1, \frac{q(\mathbf{Z}_{j\cdot},V|\mathbf{Z}'_{j\cdot})\ \pi(\mathbf{Z}_{j\cdot}|\mathbf{Y}_{\cdot\cdot},\cdots)}{q(\mathbf{Z}'_{j\cdot},V|\mathbf{Z}_{j\cdot})\ \pi(\mathbf{Z}'_{j\cdot}|\mathbf{Y}_{\cdot\cdot},\cdots)}\right\}.$$

The proposal density $q(\mathbf{Z}_{j\cdot},V|\mathbf{Z}'_{j\cdot})$ is difficult to evaluate since there are many different ways to obtain the same vertices set V. An important result from Barbu and Zhu (2005) showed that the ratio of the two proposal densities $q(\mathbf{Z}_{j\cdot},V|\mathbf{Z}'_{j\cdot})/q(\mathbf{Z}'_{j\cdot},V|\mathbf{Z}_{j\cdot})$ is of simple form. So the acceptance ratio can be computed easily. Further, with a particular choice of the proposal distribution for assigning labels for $\mathbf{Z}'_{jV}$, the acceptance ratio is 1.

3.3 Posterior inference and choice of hyper-parameters

In the Bayesian framework, parameters estimation and inference on latent variable Z_{ji}'s become a matter of summarizing the posterior distribution. In general, we can report either posterior mean, median or model over the MCMC samples. In our experimental studies, we adopted posterior means for parameter estimation and posterior modes for Z_{ji}'s.

Our algorithm requires the number of clusters K to be pre-given. In some applications, we can set K to be the number of different tissues in MR images. For example, it is common to set $K=3$ in segmenting MR brain images, which correspond to gray-matter, white-matter, and cerebrospinal-fluid tissue. Alternatively, we can use model selection criteria such as AIC and BIC to select K.

The choice of $\beta_{(ji)(j'i')}$, the edge-dependent interaction parameter as in (3), plays an important role in co-segmentation: large value leads to high influence of neighboring vertices. To the best of our knowledge, there is no optimal way to select β. We could try different β values and then select them by some model selection criteria, which,

however, is time-consuming since there are many β's (same as the number of edges). In our empirical study, we set

$$\beta_{(ji)(j'i')} = -\log\left(1 - e^{-h|Y_{ji} - Y_{j'i'}|}\right),$$

where h is a positive tuning parameter, implying high influence of neighboring vertices when their intensity measures are close. Then choose h as follows. For each image, we plotted the histogram for the intensity difference among neighboring vertices, $|Y_{ji} - Y_{j'i'}|$, where (ji) and $(j'i')$ are neighbors. (The intensity measures have been normalized with $Y_{ji} \in [0, 1]$.) We found that there was no valleys between 0 and 0.05 in all histograms, that is, two vertices would be classified in the same cluster when $|Y_{ji} - Y_{j'i'}| < 0.05$, if using a simple thresholding rule. Then we chose h such that neighboring vertices with intensity difference equal to 0.05 had a 50% chance to be connected in the SW algorithm, that is, we set $1 - \exp\{-\beta\} = e^{-0.05 \times h} \approx 0.5$. Such an equation leads to $h \approx 14$. Eventually, we adopted a less informative choice $h = 15$.

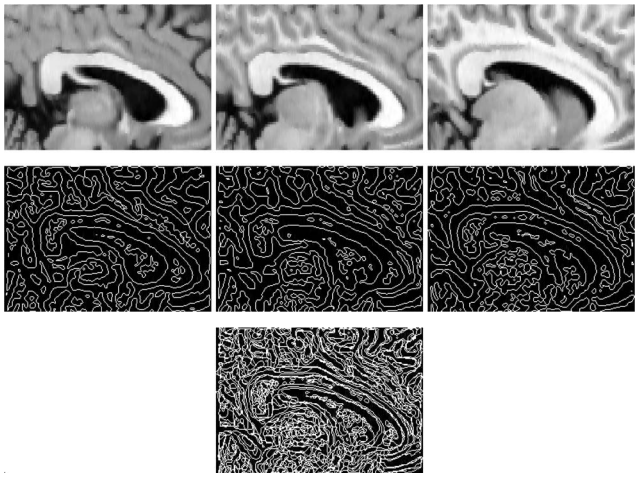

Figure 1. The top row are the original sequential MR images; the second row are their closed contours, which form super-vertices; the bottom plot is the final super-vertices for the three images.

4. EXPERIMENTS

We demonstrate the efficiency and accuracy of our co-segmentation algorithm on four different types of MRI data sets. We compare the performance of co-segmentation with individual segmentation on the first three data sets: (i) a pair of synthetic brain images; (ii) five real corpus callosum MRI slices from a volume data, and (iii) three 2D abdominal MR images scanned over time. We then demonstrate the efficiency of SWGS versus SSGS on a 3D dynamic cardiac data set. The individual segmentation procedure is almost the same as the co-segmentation, except that $\beta_{(ji)(j'i')}$ in (3) is set to be zero when $j \neq j'$, that is, information is shared among nearby vertices on the same image, but not across images.

Evaluating the accuracy of medical images segmentation is difficult due to the lack of "true" segmentation results. Since segmentation for medical images is mainly used to identify ROI's and true ROI's can be provided by experts, after obtaining the segmentation result, we label some clusters as ROI and the other clusters as non-ROI. In all the figures, we display only this two-class segmentation result: ROI versus non-ROI. We employ quantitative evaluations based on the proportion of correctly identified ROI, as introduced in Fenster and Chiu (2005). Let V_T and V_S denote the regions enclosed by the true boundary of ROI (provided by human experts) and the estimated boundary (from the segmentation algorithm) respectively. Define the true positive (TP) volume as the volume enclosed by both the true and estimated boundaries, i.e., $V_{TP} = V_S \cap V_T$, the false positive (FP) volume is $V_{FP} = V_S - V_T$, the false negative (FN) volume is $V_{FN} = V_T - V_S$, and V denotes the total region. Then define

$$\text{True Position Fraction (TPF)} = \frac{V_{TP}}{V_T},$$

$$\text{False Position Fraction (FPF)} = \frac{V_{FP}}{V - V_T},$$

$$\text{False Fraction (FF)} = 1 - \frac{V_{FP} + V_{FN}}{V_T}.$$

Following Fenster and Chiu (2005), we name the last measure "False Fraction", although it is the measure that we prefer to be large.

For each data set, we first normalize the intensity measures such that $Y_{ji} \in [0, 1]$. Then apply some standard preprocessing in image segmentation to reduce the number of vertices. We first use Laplacian of Gaussian (a compound operator that combines a smoothing Gaussian-shape operation with a differentiation Laplacian operation) to obtain the closed contours of boundary for each image. The contours give rise to a fine partition of an image and vertices falling into the same boundary can be treated as a super-vertex. To keep the consistency of super-vertices over J images, we use the overlap of the J partitions to form super-vertices. Figure 1 depicts the final super-vertices of three sequential images.

4.1 Synthetic contaminated brain image pair

In this experiment, we construct a synthetic contaminated image pair based on a brain image (of size 181×217) from the Simulated Brain Database[1], which is derived from an average of 27 T1-weighted images of a normal brain. We

[1] http://www.bic.mni.mcgill.ca/brainweb/

Table 1. Evaluation of individual segmentation and co-segmentation

	TPF		FPF		FF	
	Co-seg	Individual	Co-seg	Individual	Co-seg	Individual
Brain 1	0.98	0.81	0.03	0.01	0.97	0.81
Brain 2	0.99	0.82	0.04	0.01	0.96	0.82
CC 1	0.96	0.98	0.00	0.01	0.94	0.93
CC 2	0.98	0.99	0.01	0.01	0.93	0.89
CC 3	0.98	1.00	0.01	0.02	0.91	0.79
CC 4	0.94	0.99	0.01	0.02	0.88	0.75
CC 5	0.87	0.98	0.00	0.17	0.83	−0.98
Liver 1	0.94	0.95	0.01	0.01	0.88	0.87
Liver 2	0.93	0.93	0.00	0.04	0.90	0.48
Liver 3	0.92	0.95	0.00	0.017	0.90	0.77

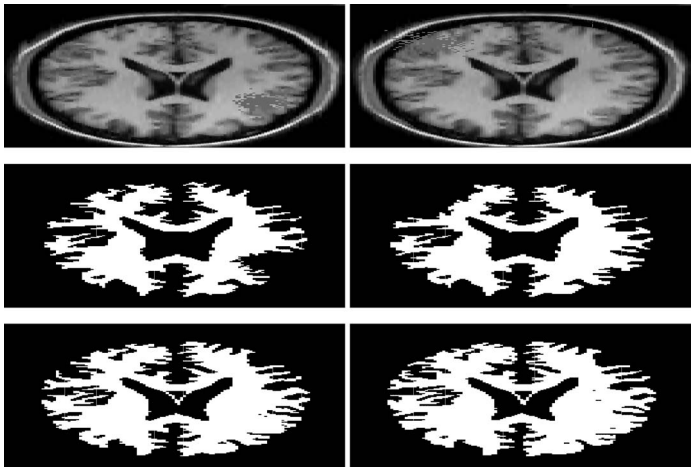

Figure 2. Data and results for the contaminated brain image pair: original images (top row), results from individual segmentation (middle row), and results from co-segmentation (bottom row).

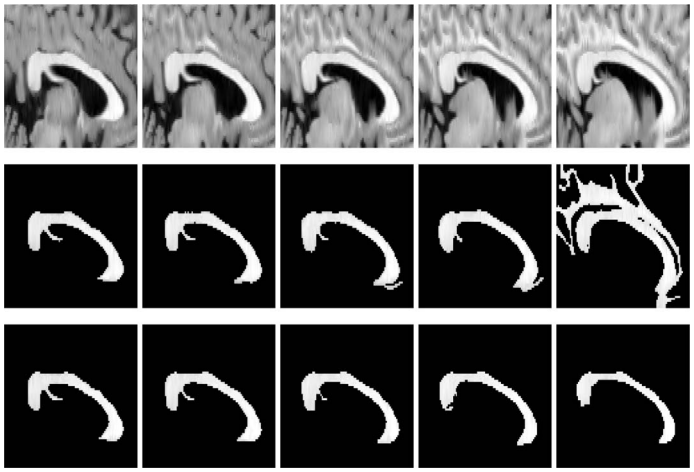

Figure 3. Data and results for CC slices: original images (top row), results from individual segmentation (middle row), and results from co-segmentation (bottom row).

add white noises at two different regions of the brain, one contaminated region on one image. This pair of synthetic data is shown on the top row of Figure 2. We want to demonstrate that the contaminated region in each image could be retrieved much better under the supervision of the other one.

We set $K = 3$ that correspond to GM (gray matter), WM (white matter) and CSF (cerebrospinal-fluid tissue). The results for individual segmentation and co-segmentation for WM are shown in Figure 2 and the quantitative evaluation is in Table 1. Although co-segmentation does slightly worse than individual segmentation on FF (approximately 2% higher than the individual segmentation), it extracts 17% more ROI than the individual segmentation as shown in TPF. As we expect, co-segmentation provides a much better overall result than individual segmentation since it utilizes the information from both images that are known to be similar.

4.2 Multiple corpus callosum slices

Several studies have indicated that the size and shape of the corpus callosum (CC) in human brain are correlated to sex, age, brain growth, and various types of brain dysfunction. In order to find such correlations, computer-assisted segmentation is needed (Lundervold et al., 1999). Five CC slices, each of size 70×100, from the Simulated Brain Database are shown in Figure 3 (top row).

We applied co-segmentation and individual segmentation algorithms on this data set with various choice of K. As discussed in Lundervold et al. (1999), the number of tissue types in CC images is less than 10. So K ranges from 2 to 10 as in the plot of AIC and BIC versus K shown in Figure 4. The most significant reduction of AIC and BIC occurs when K changes from 4 to 5. Both criteria continue decreasing after $K = 5$, but with a much slower rate. So we select $K = 5$. The segmentation results are shown in the middle and the last rows in Figure 3. We can see that co-

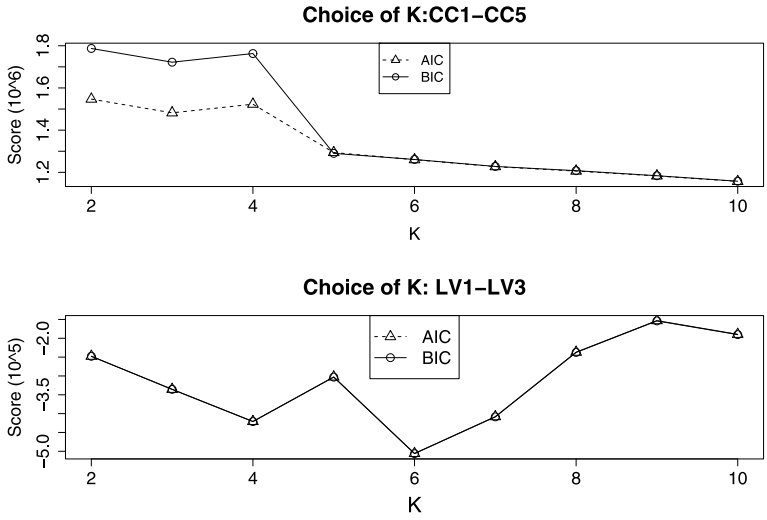

Figure 4. The plot of AIC and BIC versus various K for CC slices (top panel) and sequential liver (lower panel) images.

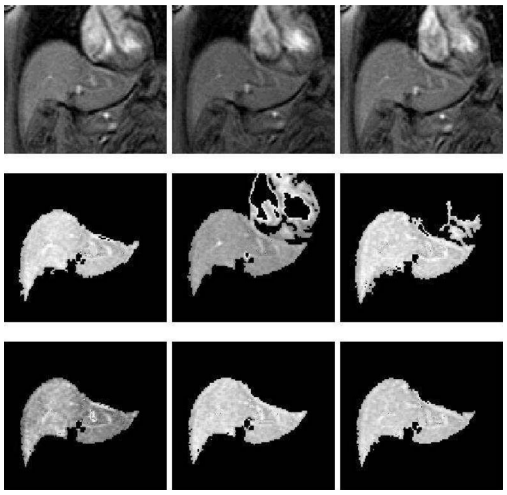

Figure 5. Data and results for sequential liver images: original images (top row), results from individual segmentation (middle row), and results from co-segmentation (bottom row).

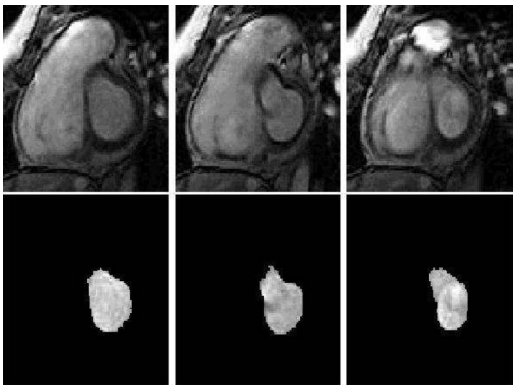

Figure 6. Data and results for the cardiac data: one slice of the 3D image is presented at 3 time points (top panel, from left to right), and results from co-segmentation (lower panel).

segmentation is roughly the same as the individual segmentation on the first four images, but is much better on the last one. As shown in Table 1, individual segmentation of the last image CC5 has FPF = 0.17 and FF = −0.98 indicating serve over segmentation, i.e., a huge false positive error. However, this is avoided by our co-segmentation procedure, which has FPF = 0.00 and FF = 0.83 for CC5, since it utilizes the information across images.

4.3 Sequential liver images

Liver perfusion is a quantitative measurement of blood flow in the liver, which provides useful information on the assessment and treatment of liver diseases. Precise segmentation of livers under perfusion is an important preliminary step for further analysis. The difficulty of liver segmentation is the ambiguity of the boundaries that are connected with other organs like the heart. In this experiment, we use sequential liver images from Chen and Gu (2006), which are 2D abdominal MR images (of size 145×118) scanned over time and are shown in Figure 5.

The bottom panel of Figure 4 presents the selection of K using AIC and BIC, both of which select $K = 6$. According to Figure 5 and Table 1, our co-segmentation algorithm outperforms the individual one.

4.4 4D cardiac data

Cardiovascular disease is one of the leading death causes in the world. Segmenting cardiac MR images is a common

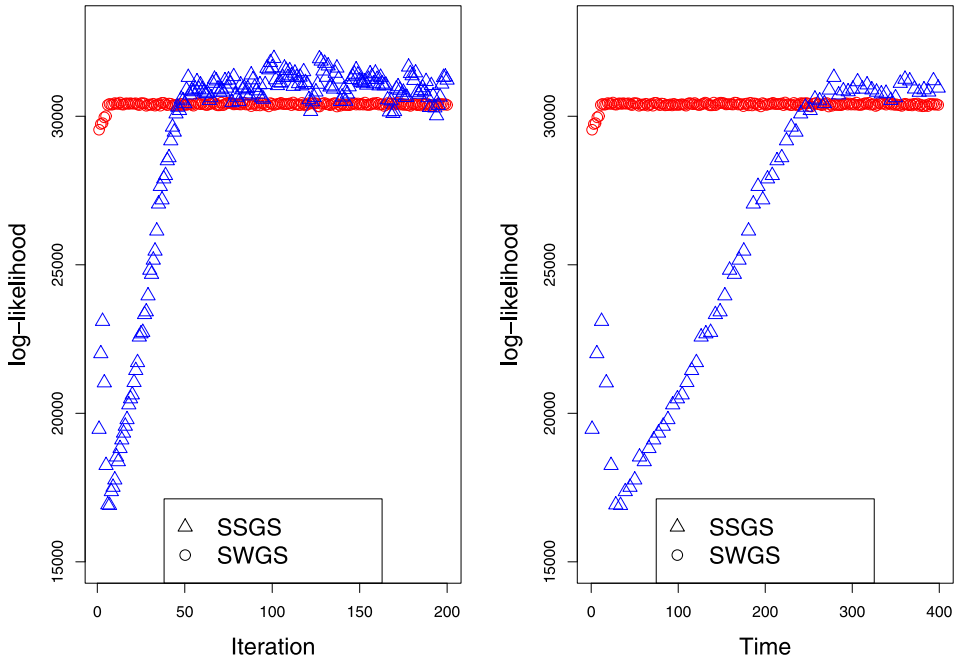

Figure 7. Convergence comparison between SSGS and SWGS. The left plot is the convergence rates of SSGS and SWGS which indicates SWGS needs only 1/10 iterations steps to be convergent. The right plot shows the speed in time for convergency which evaluates the SWGS is around 25 time faster than SSGS.

medical practice since it provides clinically useful indicators of heart function. In this experiment, we use a series of three 3D MR images (of size $101 \times 101 \times 12$) from Andreopoulos and Tsotsos (2008), which are scanned along a cardiac circle. The purpose of segmentation here is to delimitate the endocardium. As shown in Figure 6, the nearby images (along the time domain) have strong similarities, which motivates the use of co-segmentation.

For this set of data, the two co-segmentation algorithms, SWGS and SSGS, output roughly the same result in terms of accuracy. SWGS, however, has an apparent advantage in terms of speed. Figure 7 (left) compares their convergence rates in terms of the number of iterations, and indicates that SSGS needs almost 9 times more iterations than SWGC to reach convergence. This is because in each iteration, SWGS updates the cluster labels block by block, while SSGS does it pixel by pixel. Further, the way SWGS updates the cluster labels requires less computation, therefore less time, to finish an iteration than SSGS, as supported by the right panel of Figure 7 that compares the convergence speed in terms of time.

5. DISCUSSION

A new Bayesian model-based approach is introduced to segment multiple MR images simultaneously. The major contribution of our work is to utilize the shared structures among multiple images, which make the segmentation more accurate and robust. An efficient Gibbs sampler based on the Swendsen–Wang method is employed for posterior sampling, which updates the segmentation indicators block by block and is more efficient than the single-site Gibbs sampler.

APPENDIX: SHUFFLING

For $j = 1 : (J-1)$, perform the following steps sequentially,

1. Save the current configuration $\mathbf{Z}_{\cdot\cdot}$ as $\tilde{\mathbf{Z}}_{\cdot\cdot}$.
2. Based on $\mathbf{Z}_{j\cdot}$ and $\mathbf{Z}_{(j+1)\cdot}$, construct a $K \times K$ table T whose (a, b)-th entry is defined to be

$$T_{a,b} = \sum_{i=1}^{n} I[Z_{ji} = a] \times I[Z_{(j+1)i} = b].$$

   Here the row indices correspond to the clusters from image j and the column indices correspond to the clusters from image $(j+1)$.
3. Let $(a_0, b_0) = \arg\max_{a \in 1:K, b \in 1:K} T_{a,b}$.
4. Switch the labels a_0 and b_0 for $\mathbf{Z}_{(j+1)\cdot}$.
5. Delete the a_0-th row and b_0-th column from table T, and repeat the above procedure on the remaining $(K-1) \times (K-1)$ table, until reaching a table with only one entry.

6. If $\pi(\mathbf{Z}_{\cdot\cdot}) > \pi(\tilde{\mathbf{Z}}_{\cdot\cdot})$, that is,

$$\sum_{(ji)\sim(j'i'):j'=j+1} \beta_{(ji)(j'i')}\big(I[Z_{ji} = Z_{j'i'}] - I[\tilde{Z}_{ji} = \tilde{Z}_{j'i'}]\big) > 0,$$

accept the new configuration, otherwise, reset $\mathbf{Z}_{\cdot\cdot} = \tilde{\mathbf{Z}}_{\cdot\cdot}$.

ACKNOWLEDGEMENTS

The authors would like to thank Adrian Abarbu at Florida State University for his valuable discussions.

Received 19 February 2010

REFERENCES

Andreopoulos, A. and J. K. Tsotsos (2008). Efficient and Generalizable Statistical Models of Shape and Appearance for Analysis of Cardiac MRI. *IEEE Trans. Image Processing 12*(3), 335–357.

Barbu, A. and S. Zhu (2005). Generalizing Swendsen–Wang to Sampling Arbitrary Posterior Probabilities. *IEEE Trans. Pattern Analysis and Machine Intelligence 27*(8), 1239–1253.

Beucher, S. and F. Meyer (1993). The Morphological Approach to Segmentation: The Watershed Transformation. In *Math, Morphology in Image Processing*, pp. 433–481.

Chen, G. and L. Gu (2006). A Novel Liver Perfusion Analysis Based on Active Contours and Chamfer Matching. In *3rd International Workshop on Medical Imaging and Augmented Reality*, pp. 108–115.

Cheng, D. S. and M. A. T. Figueiredo (2007). Cosegmentation for Image Sequences. In *14th International Conference on Image Analysis and Processing*, pp. 635–640.

Cootes, T. F., C. J. Taylor, D. Cooper, and J. Graham (1995). Active Shape Models — Their Training and Applications. *Computer Vision and Image Understanding 61*(1), 38–59.

Edwards, R. G. and A. D. Sokal (1988). Generalization of the Fortuin–Kasteleyn–Swendsen–Wang Representation and Monte Carlo Algorithm. *Phys. Rev. D 38*(6), 2009–2012. MR0965465

Fenster, A. and B. Chiu (2005). Evaluation of Segmentation Algorithms for Medical Imaging. In *27th Annual International Conference of the Engineering in Medicine and Biology Society*, pp. 7186–7189.

Frangi, A. F., W. J. Niessen, and M. A. Viergever (2001). Three-Dimensional Modeling for Functional Analysis of Cardiac Images: A Review. *IEEE Trans. Med. Imaging 20*(1), 2–25.

Higdon, D. M. (1997). Auxiliary Variable Methods for Markov Chain Monte Carlo with Applications. *Journal of the American Statistical Association 93*, 585–595.

Kass, M., A. Witkin, and D. Terzopoulos (1988). Snakes: "Active Contour Models". *IEEE Trans. Med. Imaging*, 321–331.

Lakare, S. and A. Kaufman (2000). 3D Segmentation Techniques for Medical Volumes. *Research Proficiency Exam, Center Visual Comput., Dept. Comput. Sci., State Univ.*

Lee, T. and M. Lewicki (2002). Unsupervised Image Classification, Segmentation, and Enhancement Using ICA Mixture Models. *IEEE Trans. Image Process 11*(3), 270–279.

Lempitsky, V., P. Kohli, C. Rother, and T. Sharp (2009). Image Segmentation with a Bounding Box Prior. *IEEE International Conference on Computer Vision (ICCV)*.

Leventon, M., E. Grimson, and O. Faugeras (2000). Statistical Shape Influence in Geodesic Active Contours. *IEEE Conf. Comput. Vis. Pattern Recog. 1*, 316–323.

Lundervold, A., T. Taxt, N. Duta, and A. K. Jain (1999). Model-Guided Segmentation of Corpus Callosum in MR Images. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Volume 1, pp. 1231–1237.

McLachlan, G. and D. Peel (2000). *Finite Mixture Models*. USA: Wiley. MR1789474

Permuter, H., J. Francos, and I. Jermyn (2006). A Study of Gaussian Mixture Models of Color and Texture Features for Image Classification and Segmentation. *Pattern Recogn. 39*(4), 695–706.

Rother, C., T. Minka, A. Blake, and V. Kolmogorov (2006). Cosegmentation of Image Pairs by Histogram Matching — Incorporating a Global Constraint into MRFs. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Volume 1, pp. 993–1000.

Swendsen, R. H. and J. S. Wang (1987). Nonuniversal Critical Dynamics in Monte Carlo Simulations. *Physical Review Lettes 58*(2), 86–88.

Tsai, A., W. Wells, C. Tempany, E. Grimson, and A. Willsky (2003). Coupled Multi-Shape Model and Mutual Information for Medical Image Segmentation. *Inf. Process. Med. Imaging*, 185–197.

Wells, W., W. Grimson, R. Kikinis, et al. (1996). Adaptive Segmentation of MRI Data. *IEEE Trans. Med. Imaging 15*(4), 429–442.

Weszka, J. S. (1978). A Survey of Thresholding Techniques. *Computer Graphics and Image Processing 7*, 259–265.

Winkler, G. (2006). *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction (Stochastic Modelling and Applied Probability)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc. MR1950762

Yang, J., L. H. Staib, and J. S. Duncan (2004). Neighbor-Constrain. *IEEE Trans. Med. Imaging 23*(8), 940–948.

Younis, A., A. Soliman, and N. John (2007). Co-segmentation of MR and MR Spectroscopy Imaging Using Hidden Markov Models. In *IEEE/NIH Life Science Systems and Applications Workshop (LISA 2007)*, pp. 188–191.

Zou, Q., K. C. Keong, N. W. Sing, and Y. Chen (2001). MRI Head Segmentation for Object Based Volume Visualization. In *Proceeding of the seventh Australia and New Zealand Intelligence Information System Conference*, pp. 361–366.

Jianfeng Xu
Department of Statistics
University of Illinois at Urbana-Champaign
Champaign, IL
USA
E-mail address: jxu9@illinois.edu

Feng Liang
Department of Statistics
University of Illinois at Urbana-Champaign
Champaign, IL
USA
E-mail address: liangf@illinois.edu