Image Segmentation using MCMC

Tanya Sharma
tanyasharma16@ucla.edu

Abstract

This paper explores the application of Monte Carlo and Markov chain techniques for image segmentation. We begin by introducing the concept of image segmentation and important concepts such as Markov random fields and optimization methods that are crucial for addressing this problem. With this foundation established, we delve into the realm of unsupervised image segmentation and examine two advanced algorithms that utilize hierarchical Markov random field models to segment both noisy and textured images. Furthermore, we expand beyond Markov random fields and explore a novel approach in which samples are drawn from a posterior distribution on the space of curves, employing non-parametric representations for image segmentation.

Image segmentation, sampling, Markov random fields, conditional simulation

1 Introduction

Image segmentation can be understood as the process of assigning labels to pixels in an image, effectively dividing the image into distinct regions. This partitioning is based on the principle that pixels within the same region share similar characteristics and should be assigned the same label. Kembervo [1] in the survey paper describes image segmentation as the division of an image into homogeneous sub-images and the labeling of pixels based on their respective types. The image segmentation problem has captured my interest due to its significant relevance in current scenarios. With the rapid advancements in Machine Learning and computer vision technologies, image segmentation has gained widespread usage and applications like medical imaging, Object recognition, remote sensing etc to name a few. The image segmentation problem can be categorized as either supervised or unsupervised. Supervised image segmentation occurs when the number of regions into which the image is to be segmented is known beforehand. In contrast, unsupervised image segmentation assumes an unknown number of regions within the image and each region is modelled as individual Markov random field for estimation purposes. To provide a solid foundation, it is essential to formally define the problem of image segmentation before delving further into its intricacies.

1.1 Markov Random Fields

Markov random fields serve as a valuable framework for defining spatial images, offering computational advantages due to their utilization of neighborhood concepts. In this section, our objective is to develop a thorough comprehension of Markov random fields and their significance in the realm of modeling spatial images.
Let, S = set of lattice points.
Let s ∈ S, be a lattice point.
Let G = $\{G_s : s \in S\}$
The family G is the neighbourhood system of S if :

- ◇ for every s in S, s $\notin G_s$
- ◇ for every s and r in S, $s \in G_r$ only iff $r \in G_s$

Put simply, we establish a concept of neighborhood for each point s $\in$ S. This neighborhood is defined in such a manner that the point itself is excluded from its own neighborhood. Moreover, if point "r" is considered a neighbor of point "s," then it follows that "s" is also a neighbor of "r" i.e. the neighbourhood system is symmetric. Next, if number of sites in S are N, then let $\Omega = \{\omega = (x_{s1}, ..., x_{sN})\} : x_{si} \in \Theta, i = 1, ...N\}$
A random vector $X = X_{s \in S}$, with range Ω, is a Markov random field (MRF) over S with respect to G, if there exists a probability measure P on Ω such that :

- ◇ For every $\omega \in \Omega$, $P(X = \omega) > 0$
- ◇ For every s in S, $P(X_s = x_s | X_r = x_r, \forall r \neq s) = P(X_s = x_s | X_r = x_r, \forall r \in G_s)$ for any $\omega = (x_{s1}, ..., x_{sN})$ in Ω.

In a nutshell, within a Markov random field (MRF), for any s $\in$ S, probability of X_s conditioned on all other values of X at positions r $\in$ S and r $\neq$ s, is equal to the probability of X_s conditioned only on the neighbours of s. So, X is characterised only by local interactions and we no longer look at the entire lattice. Considering an image in the context of our discussion, let's assume we have a gray-scale image with dimensions MN, indexed by s and X can be the pixel values in the range $\Omega = (0, ..255)$.

In the above formulation, we have not defined the functional form of $P(X_s = x_x)$. Hammersley-Clifford Theorem provides a method to write the density of a Markov Random field using a Gibbs Distribution.

1.2 Gibbs Distribution for Markov Random Fields

Let S be a set of lattice points and let G be a neighbourhood system of S. Let Ω be the configuration space and P be the probability measure on Ω. X will be a random vector. X is a Markov Random Field on S w.r.t to G iff X has a Gibbs distribution over Ω i.e.

$$P(X = \omega) = \frac{exp(U(\omega)/T)}{Z}$$
$$\text{where, } U(\omega) = \sum_{c:s \in C} V_c(\omega)) \text{ and}$$
$$Z = \sum_{\omega \in \Omega} exp(-U(\omega)/T)$$

Here, U(.) is called the Energy Function, Z(.) is called the Partition Function, V(.)'s are called the Potentials and T is called the Temperature. Energy function is defined as the sum of potentials over "c". "c" is a subset of S and it is called clique. A clique is a set of points such that any two of its elements are neighbours i.e. $\forall r, s \in c, r \in G_s$ and $s \in G_r$.
If X has a Gibbs distribution then,

$$P(X_s = x_s | X_r = x_r, r \neq s) = \frac{P(X=\omega)}{P(X_r = x_r, r \neq s)}$$

And it can be proved that (refer -),

$$P(X_s = x_s | X_r = x_r, r \neq s) = \frac{exp(-U(\omega)/T)}{\sum_{x \in \Delta} exp(-U(\omega^s)/T)}$$

The advantage of using Gibbs Distribution is that the potentials within a Gibbs distribution capture the spatial consistency of the image. They provide insight into how the probability of a pixel having a specific value is affected by its neighboring pixels, whether considered individually or in various combinations.

1.3 Hierarchical Markov Random Fields

Thus far, we have explored the concept of Markov Random Fields (MRF) and their application in image modeling. However, it is worth noting that MRFs have limitations when it comes to capturing complicated prior knowledge. To address this challenge, Geman and Geman introduced hierarchical MRF models. These models incorporate a hierarchical structure that signifies the type and level of prior knowledge pertaining to the image. By incorporating this hierarchy, the models can effectively incorporate and represent complex prior knowledge. In this new set up,

$X = (X^P, X^E, X^L....)$ where each X^i is a Markov Random Field in itself. While X^P represent the image intensities, remaining X's model other structural attributes of the image. Hierarchical Markov Random Fields are used to model noisy and textured images which we will discuss in the next section.

2 Review

In this section we will review unsupervised image segmentation and examine two advanced algorithms that utilize hierarchical Markov random field models to segment noisy and textured images. Then, we venture beyond the conventional framework of Markov random fields and review a different approach. In this approach samples are drawn from a posterior distribution on the space of curves, employing non-parametric representations for image segmentation.

2.1 Unsupervised image segmentation

The algorithms examined in this section will identify the following:

- ◇ Most likely number of classes
- ◇ Associated model parameters
- ◇ Generate a corresponding image segmentation into classes

We can systematically divide this entire process into distinct components. Firstly, we will establish a formal definition of the problem within the Bayesian framework. Subsequently, we will utilize Bayes' law to break down the problem into essential elements, namely the posterior, likelihood, and priors. Finally, we will proceed to define the functional form of our posterior. Once we have our set up in place, we can employ the MAP criterion to maximise the posterior distribution.

2.1.1 Defining the problem

Let $\Omega = \{(i,j); 1 \leq i \leq M, 1 \leq j \leq N\}$ i.e. Ω is a M*N lattice.

Let $Y = \{Y_s = y_s; s \in \Omega\}$ i.e. Y represents the observed grayscale image where pixel values $\in$ (0,1].

Let $X = \{X_s = x_s; s \in \Omega\}$ are the labels of the underlying markov random field and take values from $\Theta = 0, 1, ...k - 1$. Since this problem is unsupervised the number of labels is unknown.

Two neighbourhood structures are defined : η_s and ρ_s.

η_s is defined on the neighbourhood at site s while ρ_s is defined on the observed image Y. Therefore, x_η is defined as the vector of labels defining the neighbourhood s and y_{ρ_s} is the vector of pixel grayscale values over that neighbourhood.

Ψ = parameter vector and all the model parameters are included into it.

2.1.2 Bayesian Framework

Since a Gibbs distribution is used to model the likelihood of observing the image Y given the labels X and also all the prior knowledge, therefore,

$$p(Y_s = y_s, X_s = x_s | y_{\rho_s}, x_\eta, \Psi) = \frac{1}{Z(y_{\rho_s}, x_\eta, \Psi)} e^{-U(Y_s = y_s, X_s = x_s | y_{\rho_s}, x_\eta, \Psi)}$$

After segmenting the parameter space $\Psi = [\Phi_c : \{c \in \Theta\}, \gamma]$, where Φ_c represents the model parameters of likelihood of observing pixel image given the neighbourhood and γ_c is the vector of hyper-parameters defining the prior on the label field and replacing the Gibbs distribution with its approximation called Pseudo-Likelihood, the above conditional probability becomes,

$$PL(Y = y, X = x | \Psi) = \Pi_{s \in \Omega} \frac{1}{Z(\Phi_{x_s})} exp\{-U_1(Y_s = y_s | y_{\rho_s}, X_s = x_s, \Phi_{x_s})\} * \frac{exp\{-\sum_{s \in \Omega} U_2(X_s = x_s | x_{\eta_s}, \gamma)\}}{\Pi_{s \in \Omega} \sum_c exp\{-U_2(X_S = c | x_{\eta_s}, \gamma)\}}$$

Since we know from the Bayes Rule,

Posterior $\propto$ Likelihood*Prior

Therefore, we get the following :

$$P(X = x, \psi, k | Y = y) \propto PL(Y = y, X = x | \Psi) * p_r(k) p_r(\gamma) \Pi_{c=0}^{k-1} p_r(\Phi_c)$$

where $p_r(.)$ are the prior distributions.

2.1.3 Functional form of Posterior distribution

The paper [2] applies an isotropic Markov random field model for modeling noisy images and a Gaussian Markov random field model for modeling textured images. In both models, Potts models are employed to establish the prior distributions on the label field X. However, the distinction between the two models lies in how they define the likelihood of observing pixel gray-scale values given the label field. In the isotropic Markov random field model, the likelihood of observing pixel gray-scale values given the label field is given by the Gaussian distribution. In the Gaussian MRF, hierarchical image models are used where each individual texture is a Gaussian Markov Random Field while the interaction between the regions comprising these textures is modelled by Potts Model.
After defining the posterior distribution of the image, the subsequent task involves maximizing the distribution to obtain the Maximum A Posteriori (MAP) estimate. Nevertheless, due to the high dimension of images, searching through an extensive range of configurations becomes computationally expensive. Furthermore, the partition function becomes intractable when dealing with a Gibbs distribution. Therefore, MCMC algorithms are formulated which will allow direct sampling of the parameters from the posterior distribution.

2.1.4 MCMC Sampling from the posterior distribution

In this referenced paper[2], the sampling scheme relies on the combination of Gibbs sampler and Metropolis-Hastings sub-chains. This integration allows for the efficient sampling of both model parameters and the number of classes. The sampling process consists of the following steps -

- Re-segment the image : Gibbs sampling is used to sample from the label field.
- Sample noise model parameters : Metropolis-Hastings sampling is used to sample and update noise parameters.
- Sample MRF model parameters : Metropolis-Hastings sampling is used to sample and update MRF Model parameters.
- Sample the number of classes : Sampling the number of classes is not a straightforward task. The difficulty arises from the fact that altering the model order, which refers to the number of classes in the segmentation, directly impacts the dimension of the parameter vector. It increases or decreases depending on whether we increase or decrease the number of classes. To address this challenge, reversible jumps are incorporated into the Markov Chains. The reversible jump algorithm described in the paper consists of the following steps :
  - Decide whether to increase or decrease the number of classes.
  - Generate new vector of model parameters based on the number of classes from step 1.
  - reallocate hidden data labels into the new number of classes.
  - Evaluate the acceptance ratio generated to decide whether to accept or reject this proposal, ensuring the preservation of detailed balance.

While the higher-level steps remain consistent, the precise algorithmic formulation for both isotropic segmentation and GMRF segmentation can be accessed in detail within the paper. The paper provides a comprehensive account of the specific procedures and methodologies employed for these segmentation techniques.

2.1.5 Experimental Results

For Isotropic Segmentation, when β defined double pixel clique coefficient is held fixed a priori to the value of 1.5, convergence occurs in 200 iterations of a gray-scale image with additive Gaussian noise with 6 classes. Even when β is not held fixed (as can be seen in Figure 1 and Figure 2), in around 300 iterations, MRF converged to correct number of classes. However, it is important to note here that, a normal prior is imposed on β when it is not fixed at a particular value. This is because, in the absence of the prior, β goes to infinity due to Pseudo-Likelihood approximation of the Gibbs distribution. The estimated β in this case will be incorrect but the segmentation was found to be stable.

For Gaussian Markov Random Field Segmentation (Figure 3), when β is held fixed a priori to the value of 1.5, in 600 iterations the algorithm converges to four classes after starting beginning with an arbitrary initial guess of three classes. While the number of classes in actual image are 6. So it reflects that it is not completly capturing the segments in the image. Also, it is noted that as the temperature falls to around the critical temperature, then only the spatial correlations in the image are getting modelled and this is attributed as a consequence of using Potts Model.

2.2 Curve Sampling and Geometric Conditional Simulation

In this section we expand beyond Markov random fields and explore a novel approach introduced in the paper[3] in which samples are drawn from a posterior distribution on the space of curves, employing non-parametric representations for image segmentation

2.2.1 Defining the problem

Let $\Omega \subset R^2$ be the image domain.

Let $I : \Omega \rightarrow R$ be scaler valued image.

Let $\overrightarrow{C} : [0,1] \rightarrow \Omega$ be a closed curve.

Then,

$$E(\overrightarrow{C}) = \int_{R(\overrightarrow{C})} (I(x) - m_o)^2 dx + \int_{\Omega/R(\overrightarrow{C})} (I(x) - m_1)^2 dx + \alpha \int_{\overrightarrow{C}} ds$$

where, $R(\overrightarrow{C})$ is the region enclosed by the curve $\overrightarrow{C}$ and m_o and m_1 are mean intensities inside and outside the curve.
The paper highlights an important observation that if we attempt to search for the global optimum using optimization methods then mostly we will be stuck in some local optimum and it gives little insight on the closeness to the global optimum.
So alternatively they propose, $E(\overrightarrow{C})$ as the negative log of the probability density -

$$\pi(\overrightarrow{C}|I) \propto exp(-E(\overrightarrow{C}; I))$$

The probabilistic formulation is used so that sampling methods can be employed by drawing samples from π thereby mitigating the issue of getting trapped in local minima. However, $\pi(\overrightarrow{C}|I)$ is complex and therefore is not very straightforward to sample from. So they propose the MCMC algorithm to sample from π described in the section.

2.2.2 Metropolis-Hastings to sample from curves

The authors use Metropolis-Hastings algorithm to sample from curves. They define a proposal distribution, acceptance ratio, update the curves using level sets methods and demonstrate the computation of transition probabilities to ensure sampling from posterior. They do not choose the proposal arbitrarily as it can lead to slow mixing rate. Instead, they implicitly define it so that there is a high probability that it generates samples from π.

- ⋄ Start with an initial curve $\overrightarrow{C}$ and set t=1.
- ⋄ Create $f^{(t)}(p) = h \bigotimes n^{(t)}(p) + \mu_{\overrightarrow{C}^{(t-1)}}(p)$
- ⋄ Compute $\overrightarrow{\Gamma}^{(t)}(p) = \overrightarrow{C}_a^{(t-1)}(p) + f^{(t)}(p)\overrightarrow{N}_{\overrightarrow{C}^{(t-1)}}(p)$.
- ⋄ Evaluate Hastings ratio $\eta(\overrightarrow{\Gamma}^{(t)}|\overrightarrow{C}^{(t-1)})$.
- ⋄ $\overrightarrow{C}^{(t)} = \overrightarrow{\Gamma}^{(t)}$ with probability $\eta(\overrightarrow{\Gamma}^{(t)}|\overrightarrow{C}^{(t-1)})$. otherwise, $\overrightarrow{C}^{(t)} = \overrightarrow{C}^{(t-1)}$
- ⋄ Increment t and return to step 2.

2.2.3 Conditional Simulation

In this section, the authors propose a semi-automatic segmentation algorithm that combines -partial user segmentation with conditional simulation to estimate the missing portions of the curve. The core idea is that we sample from a multi-dimensional probability distribution conditioned on the fact that some of the dimensions are known.

2.2.4 Experimental Results

In order to test the performance of the proposed algorithms, the authors implement them on two main applications - Prostate magnetic resonance segmentation and Geological Inversion Problem. In both cases, the images had a dimension of 256x256 pixels and 1000 samples were generated from $\pi(\overrightarrow{C}|I)$. They compare results by plotting most likely samples, histogram images and Marginal confidence bounds.

In Figure 4, the results are for a T1-weighted prostate MR image. The segmentation should correctly identify the spectrum and rectum. The segmentation takes place based on the pixel intensity inside and outside the curves. There can be three types of segmentation : correct prostate segmentation, correct rectum segmentation and one with both the prostate and rectum. It is noted that the most likely segmentation is the one that encompasses both the prostate and rectum.
Additionally, the paper highlights that conditional simulation exhibits superior performance when applied to estimation problems that involve global coupling. It used the gravity inversion problem to illustrate this point.

3 Discussion

In the preceding sections, we encountered two distinct approaches to image segmentation, each employing a different methodology. One approach centered around Markov Random Fields, which leveraged local properties to perform segmentation by considering the characteristics of neighboring pixels. In contrast, the other approach, known as curve sampling, took a more global perspective, incorporating geometric properties to guide the segmentation process.

During my review of the papers, I encountered certain limitations related to the generalizability of the proposed methods. In the case of the MRF model, although the algorithm demonstrated efficient performance, it was primarily tested on relatively simple images. This raises concerns about how well the algorithm would perform on highly complex images that contain intricate structures and diverse visual characteristics. Therefore, the generalizability of the MRF model to more challenging image data-sets remains uncertain. Similarly, the curve sampling algorithm for image segmentation showcased its effectiveness in specific use cases. However, the limited range of applications in which the algorithm was applied raises questions about its applicability and generalizability to diverse image data-sets and real-world scenarios. It is crucial to assess the algorithm's performance on a broader range of images to evaluate its potential limitations and to ascertain its adaptability to different types of segmentation tasks.
Another aspect that seemed to be missing from both papers was a thorough performance comparison with other existing algorithms in the field. A comparative analysis with similar algorithms used for unsupervised image segmentation or curve sampling (with gradient based approach) would have provided a more comprehensive evaluation of the proposed approaches. Such comparisons are crucial in understanding how well the algorithms perform in terms of computational efficiency and overall effectiveness.By including a performance comparison, researchers can obtain a clearer perspective on

the strengths and weaknesses of their proposed algorithms in relation to alternative methods. This would enable a more informed assessment of the computational aspects, such as processing time, resource utilization, and scalability, which are vital considerations for practical implementation.
Another notable limitation of the curve sampling algorithm is its inability to effectively segment objects that are in close proximity to each other. This limitation arises from the algorithm's reliance on pixel intensities inside and outside the curves to perform segmentation. Moreover and even the paper acknowledges the inefficiency of the algorithm in accurately segmenting image regions with high curvatures. This limitation suggests that the proposed method may struggle to effectively capture and delineate intricate or highly curved structures within images. Such regions pose a challenge due to their complex geometries, which may lead to inaccurate or incomplete segmentation results.

Hence, the algorithms proposed in both papers demonstrate innovation and significant advancements in their respective fields. However, it is crucial to acknowledge that there are certain limitations that, if addressed, would contribute to a more comprehensive analysis and understanding of the research.

References

[1] Gerda Kamberova,"Markov Random Field Models: A Bayesian Approach to Computer Vision Problems", April 1992.

[2] S.A. Barker, P.J.W. Rayner, "Unsupervised image segmentation using Markov random field models", March 1999.

[3] Ayres Fan, John W. Fisher III, Jonathan Kane, and Alan S. Willsky, "MCMC Curve Sampling and Geometric Conditional Simulation", February 2008.

[4] Rupprecht, Christian and Peter, Loic and Navab, Nassir, "Image Segmentation in Twenty Questions", June 2015.

[5] Zoltan Kato and Ting-Chuen Pong, "A Markov random field image segmentation model for color textured images", March 2006.

[6] Donald Geman and Stuart Geman, "Bayesian Image Analysis", 1986.

[7] Hedibert Freitas Lopes, "A note on Reversible Jump Markov Chain Monte Carlo", February 2006.

[8] Azadeh Mohebi and Paul Fieguth, "Posterior Sampling of Scientific Images", 2006.

4 Appendix

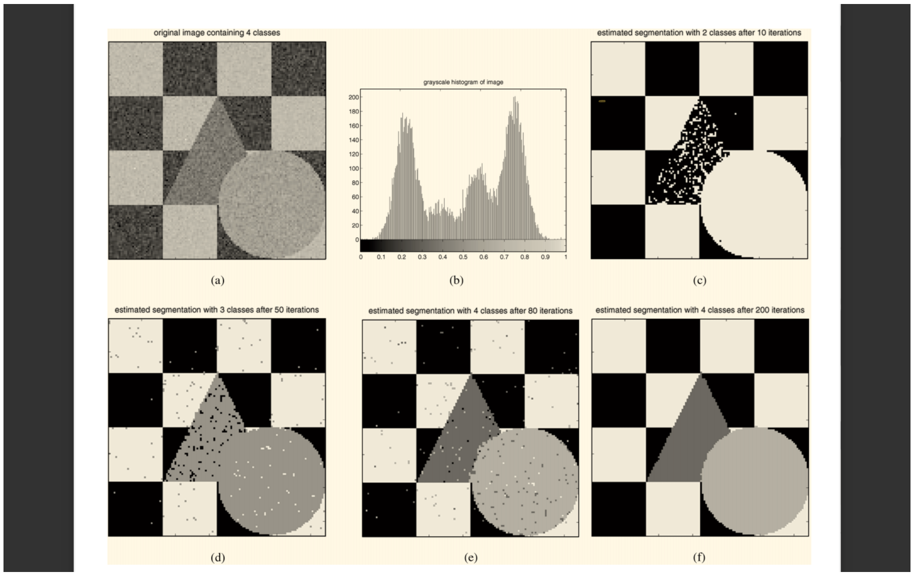

Figure 1: 200 iteration experiment on an image with four classes and beginning the algorithm with an arbitrary initial guess of two classes

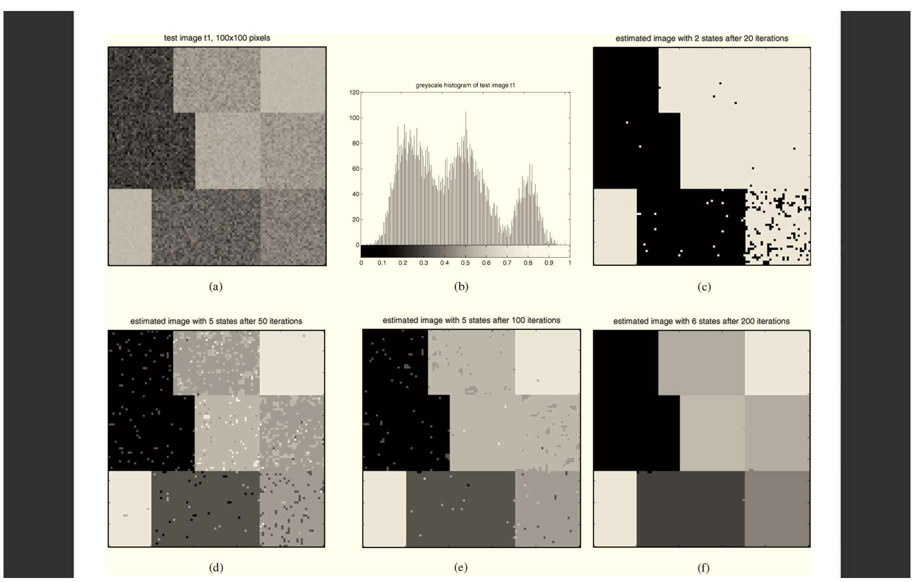

Figure 2: 200 iteration experiment on an image with six classes and beginning the algorithm with an arbitrary initial guess of two classes

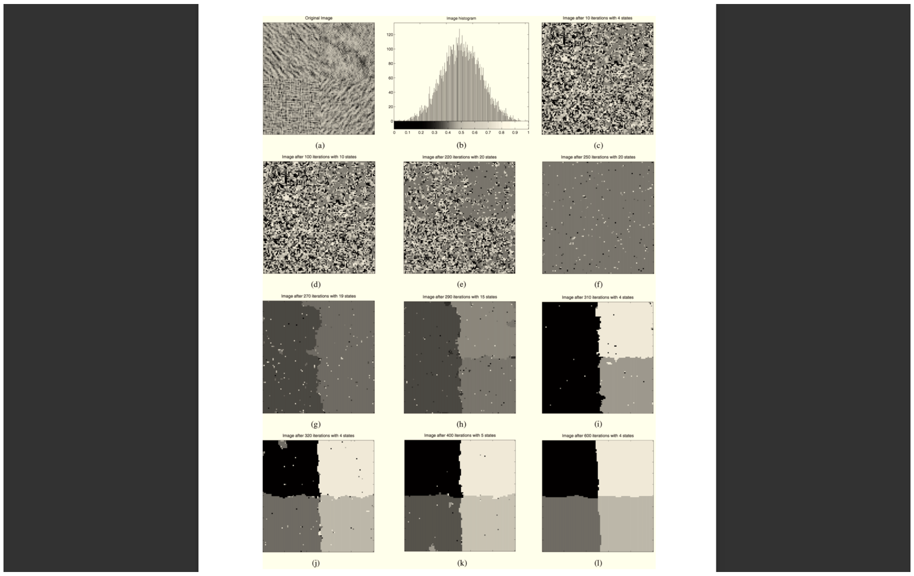

Figure 3: 600 iteration experiment on a four state synthetic GMRF image. The algorithm begins with an arbitrary initial guess of three classes and converges to four.

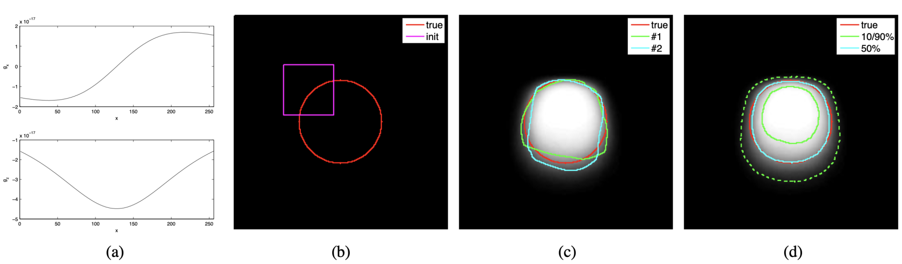

Figure 4: Prostate Segmentation (a) Observed gravity profile (b)Initial curve and true boundary (c)Two most likely samples (d)Marginal confidence bounds and histogram image