

# Bayesian Image Segmentation Using MRF's Combined with Hierarchical Prior Models

Kohta Aoki<sup>1</sup> and Hiroshi Nagahashi<sup>2</sup>

<sup>1</sup> Interdisciplinary Graduate School of Science and Engineering,  
Tokyo Institute of Technology

<sup>2</sup> Imaging Science and Engineering Laboratory,  
Tokyo Institute of Technology,  
4259 Nagatsuta-cho, Midori-ku, Yokohama, 226-8503 Japan  
{aoki, longb}@isl.titech.ac.jp

**Abstract.** The problem of image segmentation can be formulated in the framework of Bayesian statistics. We use a Markov random field as the prior model of the spacial relationship between image pixels, and approximate an observed image by a Gaussian mixture model. In this paper, we introduce into the statistical model a hierarchical prior structure from which model parameters are regarded as drawn. This would give an efficient Gibbs sampler for exploring the joint posterior distribution of all parameters given an observed image and could make the estimation more robust.

## 1 Introduction

Image segmentation is an important low-level processing which could lead to higher level tasks such as object recognition, tracking and content-based indexing and retrieval. An observed image is segmented into some homogeneous regions. This can be viewed as a labeling process where each pixel in the image is assigned a label indicating the region to which it should belong. In this paper, the problem of image segmentation is formulated in a Bayesian framework.

In recent years, much attention has been attracted to Bayesian approaches that can give promising solutions to various vision problems. Because a Bayesian framework allows for the formulation of statistical models for image characteristics and the integration of prior knowledge about contextual structures, it has been applied to the design of algorithms for image segmentation [9].

We propose a statistical model with a hierarchical prior structure for image segmentation by treating model parameters as random variables and specifying their prior distributions. This allows the Gibbs sampler [5] to effectively explore a posterior distribution of model parameters derived from Bayes' theorem and provides robust estimation. The spacial relationship between image pixels is modeled as a Markov random field (MRF) prior, and the likelihood of possible pixel values is defined as a Gaussian mixture model (GMM). A line process is also used to represent the discontinuity of pixel values at the boundary between two distinct regions.

Our estimation method is performed based on the model posterior distribution rather than using maximum-likelihood (ML) procedures [10, 11]. By choosing conjugate priors of the parameters, we could present a more efficient Gibbs sampler than the sampling methods proposed in [1, 6].

## 2 Stochastic Models for Image Segmentation

Denoting a rectangular lattice of the size  $M \times N$  by  $\mathcal{S} = \{(i, j) | 1 \leq i \leq M, 1 \leq j \leq N\}$ , an observed image is a set of  $d$  dimensional feature vectors defined on  $\mathcal{S}$ , i.e.,  $\mathcal{Y} = \{\mathbf{y}_s | s \in \mathcal{S}\}$ . If  $k$  is the number of regions constituting the observed image, the set of possible labels can be represented by  $\mathcal{K} = \{1, \dots, k\}$ . When a set of random variables  $\mathcal{X} = \{X_s | s \in \mathcal{S}\}$  is defined on the identical lattice, labels assigned to pixel sites are assumed to be a configuration  $\{\mathbf{X} = \mathbf{x}\} = \{X_s = x_s | s \in \mathcal{S}\}$  (or abbreviated simply as  $\mathbf{x}$ ). Image segmentation is viewed as statistical inference for labeling given an observed image.

### 2.1 Mixture Models

A probability distribution of pixel values in an observed image can be approximated by a mixture model. A pixel value  $\mathbf{y}$  is assumed to be the realization of an independent and identically distributed random variable  $\mathbf{Y}$  with the probability density function

$$p(\mathbf{y} | \pi, \boldsymbol{\theta}) = \sum_{c=1}^k \pi_c f(\mathbf{y}; \boldsymbol{\theta}_c), \quad (1)$$

where  $\pi = (\pi_1, \dots, \pi_k)$  are called the mixture weights which satisfy the following two conditions;  $\pi_c > 0, \sum_{c=1}^k \pi_c = 1$ .

We allow the density functions of a mixture model to follow the multivariate normal distribution  $\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  with a mean vector  $\boldsymbol{\mu}$  and a covariance matrix  $\boldsymbol{\Sigma}$ , that is, for  $c = 1, \dots, k$ ,

$$f(\mathbf{y}; \boldsymbol{\theta}_c) = f(\mathbf{y}; \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) = \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}_c|}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}_c)^T \boldsymbol{\Sigma}_c^{-1} (\mathbf{y} - \boldsymbol{\mu}_c) \right\}. \quad (2)$$

A pixel value is drawn from the distribution indicated by the corresponding label. In other words, the distribution of pixel values for each partial region comprising the observed image is approximated by the corresponding normal distribution of a mixture model.

### 2.2 Markov Random Fields

We regard labels as random variables and consider the conditional distribution of a label at each site given those at all the other sites. It is natural to think that if two sites are away from each other, there is little or no relationship between



**Fig. 1.** 6 possible patterns of edges and corresponding potentials (a)  $V_e = -2$ , (b)  $V_e = 2$ , (c)  $V_e = 1$ , (d)  $V_e = -1$ , (e)  $V_e = 1$ , (f)  $V_e = 2$ . (Pixel sites are represented by a circle and line process sites are depicted by a square.)

their labels. If the label at site  $s$  is conditioned on the configuration of a set of sites  $\eta_s$ , then  $\eta_s$  is called the neighborhood of  $s$ .

An MRF defines the Markov property of a spatial process. Although this is characterized by the local conditional distributions involving a neighborhood system, a specific representation of its joint probability distribution is derived from the MRF-Gibbs distribution equivalence [2]. Assuming the 4-neighborhood system, its corresponding clique [5] and homogeneity, the local conditional probability of a label  $x_s$  is defined as

$$p(x_s | \phi, \mathbf{x}_r, r \in \eta_s) = \frac{\exp \left\{ -\alpha_{x_s} - \sum_{r \in \eta_s} \beta V_c(x_s, x_r) \right\}}{\sum_{c \in \mathcal{K}} \exp \left\{ -\alpha_c - \sum_{r \in \eta_s} \beta V_c(c, x_r) \right\}}, \quad (3)$$

where  $\phi = (\alpha_1, \dots, \alpha_k, \beta)$  are the parameters of an MRF, and  $V_c(x_s, x_r) = -1$  if  $x_s$  and  $x_r$  have the same label and 1 otherwise. The energy function of the corresponding Gibbs distribution takes the following form

$$U(\mathcal{X} | \phi) = \sum_{s \in \mathcal{S}} \alpha_{x_s} + \sum_{s \in \mathcal{S}} \sum_{r \in \eta_s} \beta V_c(x_s, x_r). \quad (4)$$

To calculate the Gibbs distribution, the partition function that is the sum over all possible configurations in the label configuration space  $\mathbb{X}$  needs to be evaluated. The evaluation however is generally intractable, so we can approximate it by a pseudo-likelihood [3], which is provided by taking over all pixel sites the product of the conditional probability given by Eq. 3.

### 2.3 Line Processes

A line process [5] comprises a lattice of random variables  $\mathcal{L} = \{L_t | t \in \mathcal{D}\}$  located between vertical or horizontal pair of pixel sites, which can be used to prevent *oversmoothing* that may occur at discontinuities among neighboring pixel sites. The process takes either value of the binary labels  $\{0, 1\}$  indicating that an edge element is absent or present, respectively. The joint prior on both labels and lines is factored as  $p(\mathcal{X}, \mathcal{L} | \phi) = p(\mathcal{X} | \phi, \mathcal{L}) p(\mathcal{L})$ . This formulation describes that the label process prior is conditioned on the configuration of the line process.

The corresponding Gibbs energy function is written as

$$\begin{aligned} U(\mathcal{X}, \mathcal{L} | \phi) &= U(\mathcal{X} | \phi, \mathcal{L}) + U(\mathcal{L}) \\ &= \sum_{s \in \mathcal{S}} \alpha_{x_s} + \sum_{s \in \mathcal{S}} \sum_{s' \in \eta_s} (1 - l_{s, s'}) \beta V_c(x_s, x_{s'}) + \sum_{t \in \mathcal{D}} \gamma V_e(l_t, l_{t'} | t' \in \tau_t), \end{aligned} \quad (5)$$

where  $\gamma$  is the constant parameter which manage the balance between a line process and a label process, and  $\tau_t$  is the neighborhood system of a line site  $t$  as in [5].  $V_e(\cdot, \cdot)$  is the potential function involving six types of possible edges (shown in Fig. 1), which may be specified to reflect empirical knowledge about the characteristics found in edges: for example, they tend to lie on a straight line, while isolated elements, endings, and corners are less likely to be seen in due order. The Gibbs distribution for the line process could be approximated by a pseudo-likelihood as with the prior on the label process.

### 3 Hierarchical Prior Models

We introduce a hierarchical structure into the model prior by treating the parameters of the mixture model and two Markov random fields as random variables. The mixture model with a hierarchical prior structure for univariate variables proposed by Richardson & Green [7] can be extended to the multivariate case [8]. Furthermore, combining with an MRF which represents a spatial independence among variables, our model is applied to the segmentation of a color image.

#### 3.1 The Priors on Mixture Model Parameters

The priors on parameters  $\mu_c$  and  $\Sigma_c$  are respectively the normal and Wishart distributions

$$\mu_c \sim \mathcal{N}_d(\xi, \kappa^{-1}) \quad \text{and} \quad \Sigma_c^{-1} | \rho \sim \mathcal{W}_d(a, \rho^{-1}). \quad (6)$$

We choose here three dimensional ( $d = 3$ ) vectors in the  $L^*u^*v^*$  color space as a set of data being observed. Then,  $\xi = [\xi_1 \ \xi_2 \ \xi_3]^T$  are computed so that each element is the midpoint of the observed range of the corresponding data element, and letting  $R_i$  ( $i = 1, 2, 3$ ) be the length of this variable interval,  $\kappa$  is obtained as the diagonal matrix  $\kappa = \text{diag}(R_1^{-2}, R_2^{-2}, R_3^{-2})$ . Although these priors are not “natural conjugate” for  $(\mu_c, \Sigma_c)$ , where the parameters within each pair are *a priori* dependent [7], their conjugacy are useful to implement the Gibbs sampler.

When priors’ parameters, or hyperparameters, are considered as constant, there could be an inherent uncertainty about their values. Therefore an additional hierarchical structure which could deal with this uncertainty enables us to develop flexible and robust estimation techniques. The hyperparameter  $\rho$  is regarded as a random variable being drawn from the Wishart distribution

$$\rho \sim \mathcal{W}_d(q, \nu^{-1}), \quad (7)$$

where  $\nu = v\kappa$ .

#### 3.2 The Prior on MRF Parameters

The external field parameters  $\alpha = (\alpha_1, \dots, \alpha_k)$  used in an MRF can be correlated with the mixture weights  $\pi = (\pi_1, \dots, \pi_k)$  of a mixture model as

$$\alpha_c = -\log \pi_c, \quad c = 1, \dots, k. \quad (8)$$

Ignoring the term of pairwise clique potentials in Eq. 3 which represents the dependence between two neighboring sites, the probability of a label assigned to site  $s$  is represented as

$$p(x_s) = \frac{e^{-\alpha_{x_s}}}{\sum_{c \in \mathcal{K}} e^{-\alpha_c}} = \frac{\pi_{x_s}}{\sum_{c \in \mathcal{K}} \pi_c} = \pi_{x_s}. \quad (9)$$

Thus the MRF parameters  $\alpha$  turn out to be equivalent to the mixture weights  $\pi$ . We select the following Dirichlet distribution as the conjugate prior of  $\pi$ ;

$$(\pi_1, \dots, \pi_k) \sim \mathcal{D}(u, \dots, u). \quad (10)$$

Due to the approximation of a likelihood by the pseudo-likelihood, the posterior distribution of  $\beta$  will be improper under particular configurations of labels, so  $\beta$  is set *a priori* [1].

### 3.3 The Posterior Distribution

By applying the Bayes' theorem, the joint posterior density for our hierarchical model is expressed as

$$\begin{aligned} p(\mathcal{X}, \mathcal{L}, \theta, \phi, \rho | \mathcal{Y}) &\propto p(\rho)p(\phi)p(\theta|\rho)p(\mathcal{Y}|\mathcal{X}, \theta)p(\mathcal{X}|\phi, \mathcal{L})p(\mathcal{L}) \\ &\approx \prod_{s \in \mathcal{S}} |\Sigma_{x_s}|^{-1} \exp \left\{ -\frac{1}{2} (\mathbf{y}_s - \boldsymbol{\mu}_{x_s})^T \Sigma_{x_s}^{-1} (\mathbf{y}_s - \boldsymbol{\mu}_{x_s}) \right\} \\ &\times \prod_{s \in \mathcal{S}} \frac{\exp \left\{ -\alpha_{x_s} - \sum_{s' \in \eta_s} (1 - l_{s,s'}) \beta V_c(x_s, x_{s'}) \right\}}{\sum_{c \in \mathcal{K}} \exp \left\{ -\alpha_c - \sum_{s' \in \eta_s} (1 - l_{s,s'}) \beta V_c(c, x_{s'}) \right\}} \\ &\times \prod_{t \in \mathcal{D}} \frac{\exp \{ -\gamma V_e(l_t, \mathbf{l}_{t'}; t' \in \tau_t) \}}{\sum_{e \in \{0,1\}} \exp \{ -\gamma V_e(e, \mathbf{l}_{t'}; t' \in \tau_t) \}} \times p(\rho) \prod_{c \in \mathcal{K}} p(\boldsymbol{\mu}_c) p(\Sigma_c) p(\alpha_c). \quad (11) \end{aligned}$$

This would differ from the model for gray-scale image segmentation proposed by Barker [1] and the extended model for color image presented by Kato [6] in the point that the hierarchical prior structure attached to MRF's could allow for flexible and robust estimation.

## 4 Bayesian Estimation

A posterior distribution can be explored by using Markov chain Monte Carlo (MCMC) methods, especially the Gibbs sampler [5] in this work. Below, all the parameters to be estimated are denoted as  $\boldsymbol{\vartheta} = (\mathcal{X}, \mathcal{L}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \alpha, \rho)$ .

### 4.1 Markov Chain Monte Carlo Methods

Bayesian inference of  $\boldsymbol{\vartheta}$  is based on the joint posterior distribution  $p(\boldsymbol{\vartheta}|\mathcal{Y})$  derived from the priors and the likelihood functions via Bayes' theorem. The posterior marginal distribution of each label  $x_s$  given an observed image is followed by

$$p(x_s = c|\mathcal{Y}) = \sum_{\mathcal{X} \in \mathbb{X}} \delta(x_s, c)p(\mathcal{X}|\mathcal{Y}), \quad (12)$$

where  $\delta(i, j)$  denotes the Kronecker delta. One could find  $p(\mathcal{X}|\mathcal{Y})$  by marginalizing the joint posterior probability  $p(\boldsymbol{\vartheta}|\mathcal{Y}) = p(\mathcal{X}, \mathcal{L}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \alpha, \boldsymbol{\rho}|\mathcal{Y})$  over all the parameters except for  $\mathcal{X}$ . Because the sum over all possible configurations of both labels and lines needs to be evaluated in Eq. 12, it is not possible to calculate this probability analytically even for images of moderate sizes. Thus, in this paper, we adopt the Gibbs sampler to perform the estimation.

Markov chain Monte Carlo (MCMC) methods can be used to construct a sequence of variables  $\{\boldsymbol{\vartheta}^{(i)} | i \in \mathbb{N}\}$  that follow a target distribution, or  $p(\boldsymbol{\vartheta}|\mathcal{Y})$  in this case, as an invariant distribution. That is, generated samples using an MCMC method can be regarded as asymptotically distributed according to the target distribution. The posterior distribution given by Eq. 12 can therefore be evaluated from such generated samples by

$$\hat{p}(x_s = c|\mathcal{Y}) = \frac{1}{n} \sum_{i=1}^n \delta(x_s^{(i)}, c). \quad (13)$$

When the problem of image segmentation is formulated in a Bayesian framework, the approach to finding the maximum *a posteriori* (MAP) estimate for the site labels is often used (for example, in [1, 6]). Because of the direct implementation of the Gibbs sampler as mentioned below, a labeling that maximizes the posterior marginal (MPM) at each pixel site is considered here as an optimal result, such as

$$x_s^* = \arg \max_{c \in \mathcal{K}} \hat{p}(x_s = c|\mathcal{Y}), \quad s \in \mathcal{S}. \quad (14)$$

In the same way, we can estimate an optimal edge element for each line site  $t$  as

$$l_t^* = \arg \max_{e \in \{0,1\}} \hat{p}(l_t = e|\mathcal{Y}) = \begin{cases} 1 & \text{if } \#\{i | l_t^{(i)} = 0\} < \#\{i | l_t^{(i)} = 1\} \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

In other words, if the number of generated samples taking the value of 1 is larger than that of samples taking the value of 0, then  $l_t^* = 1$ . Otherwise,  $l_t^* = 0$ .

## 4.2 The Gibbs Sampler

It is not possible to directly draw samples from the model posterior distribution  $p(\boldsymbol{\vartheta}|\mathcal{Y})$ . By partitioning  $\boldsymbol{\vartheta}$  into six parts for which conjugate priors are selected respectively, the full conditional posterior distribution of each parameter where all the other parameters and the observed image are given and fixed can be obtained analytically as in [8]. Thus the Gibbs sampler for our model is composed of the following five procedures:

- (a) updating the MRF parameters  $\alpha$  (or the mixture weights  $\pi$ );
- (b) updating the mixture component parameters  $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ;
- (c) updating the labels  $\boldsymbol{x}$ ;

- (d) updating the lines  $l$ ;
- (e) updating the hyperparameter  $\rho$ .

Getting started with an initial value  $\boldsymbol{\vartheta}^{(0)}$  drawn from the priors and then iterating the above process, a Markov chain  $\{\boldsymbol{\vartheta}^{(i)} | i \in \mathbb{N}\}$  with the invariant distribution  $p(\boldsymbol{\vartheta} | \mathcal{Y})$  can be constructed.

Each procedure is described below. Notice that the symbol ‘ $|\dots$ ’ being used from here denotes conditioning on the values of all other parameters and on the obtained data.

**Updating the MRF Parameters  $\alpha$ .** In updating the MRF parameters  $\alpha$ , the mixture weights  $\pi$  are first drawn from the conditional distribution as in

$$(\pi_1, \dots, \pi_k) | \dots \sim \mathcal{D}(u + m_1, \dots, u + m_k), \quad (16)$$

where  $m_c$  is the number of pixel sites with label  $c$ , that is,  $m_c = \#\{s \in \mathcal{S} | x_s = c\}$ . Then  $\alpha$  can be computed from Eq. 8.

**Updating the Mixture Component Parameters  $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .** Updating of the mixture model parameters  $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  is performed respectively, following that

$$\boldsymbol{\mu}_c | \dots \sim \mathcal{N}_d(\boldsymbol{\kappa}_c^{-1}(m_c \boldsymbol{\Sigma}_c^{-1} \bar{\mathbf{y}}_c + \boldsymbol{\kappa} \boldsymbol{\xi}), \boldsymbol{\kappa}_c^{-1}), \quad (17)$$

$$\boldsymbol{\Sigma}_c^{-1} | \dots \sim \mathcal{W}_d\left(a + m_c, \left[\boldsymbol{\rho} + \sum_{s: x_s = c} (\mathbf{y}_s - \boldsymbol{\mu}_c)(\mathbf{y}_s - \boldsymbol{\mu}_c)^T\right]^{-1}\right), \quad (18)$$

where  $\bar{\mathbf{y}}_c$  is the cluster mean of the values of pixels with label  $c$ , which is computed as

$$\bar{\mathbf{y}}_c = \frac{1}{n_c} \sum_{s: x_s = c} \mathbf{y}_s, \quad s \in \mathcal{S}, \quad (19)$$

and  $\boldsymbol{\kappa}_c = m_c \boldsymbol{\Sigma}_c^{-1} + \boldsymbol{\kappa}$ .

**Updating the Labels  $x$ .** The label at each site  $s \in \mathcal{S}$  is updated according to its conditional posterior distribution,

$$p(x_s = c | \dots) \propto |\boldsymbol{\Sigma}_c|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{y}_s - \boldsymbol{\mu}_c)^T \boldsymbol{\Sigma}_c^{-1}(\mathbf{y}_s - \boldsymbol{\mu}_c) - \alpha_c - \sum_{s' \in \eta_s} (1 - l_{s, s'}) \beta V_c(c, x_{s'})\right\}. \quad (20)$$

Compared with conditional posterior distributions for allocation variables (or the missing data) used in common mixture models [8], this contains the term where both the dependence and the discontinuity between neighboring pixel sites are taken into account.

**Updating the Lines  $l$ .** The label at each line site  $t$  in the line process  $\mathcal{D}$  is updated by

$$p(l_t = e | \dots) \propto \exp \{ -(1 - e)\beta V_c(x_s, x_{s'}) - \gamma V_e(e, \mathbf{l}_{t'}; t' \in \tau_t) \}, \quad (21)$$

where  $x_s$  and  $x_{s'}$  are the labels at pixel site  $s$  and  $s'$ , respectively, which lie to the sides of line site  $t$ .

**Updating the Hyperparameter  $\rho$ .** The hyperparameter  $\rho$  is drawn from its conditional distribution,

$$\rho | \dots \sim \mathcal{W}_d \left( q + ka, \left[ \boldsymbol{\nu} + \sum_{c=1}^k \boldsymbol{\Sigma}_c^{-1} \right]^{-1} \right). \quad (22)$$

### 4.3 Convergence Diagnostics

An initial sample  $\boldsymbol{\vartheta}^{(0)}$  of the parameters might be drawn from the corresponding priors. To reduce the dependency of an estimate given by Eq. 13 on the initial sample, the first  $m$  iterations of the Gibbs sampler would be discarded. In other words, inference is performed using only subsequent samples after they are thought to achieve convergence to an invariant distribution. Many different approaches have been proposed for determining a suitable value of  $m$ , or for diagnosing whether or not the chains converge; see, for example, the expository and comparative review by Cowles and Carlin [4].

We adopt here one of the simplest methods, where the label changing rate between consecutive samples is computed from obtained label samples  $\mathbf{x}^{(t)}$  by

$$r = 1 - \frac{1}{MN} \sum_{s \in \mathcal{S}} \delta(x_s^{(t-1)}, x_s^{(t)}), \quad (23)$$

and then convergence diagnostics is done for several sample paths by assessing whether each the changing rate can stably fluctuate without depending on the initial value.

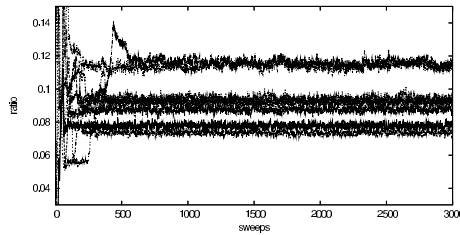
## 5 Experiments

The proposed statistical model has been used to segment several standard color images. The pixel  $RGB$  values in an observed image are converted into coordinates in the  $L^*u^*v^*$  color space which is designed to be perceptually uniform, and then are normalized so that each element has the mean of zero and the variance of one. The number of regions comprising an image is suitably specified: we here assume that  $k = 6$ . The constant parameters are set referring to the literature [7], such as  $a = d + 3$ ,  $q = d + 1$ ,  $v = 100q/a$ ,  $u = 1$ . In our experiments, the pairwise clique coefficient  $\beta$  and the parameter  $\gamma$  are both set to the value of 0.1.

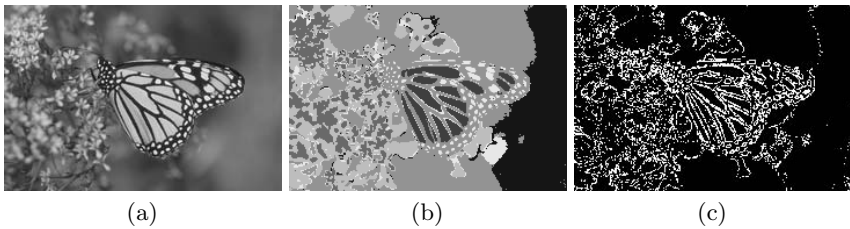


For the image shown in Fig. 3(a), several sample paths were constructed using the Gibbs sampler, where each path consists of 3,000 sample points. The variations of the label changing rates computed from Eq. 23 are presented in Fig. 2. This shows that the first about 1,000 sample points might be dependent on initial values and the subsequent points could be thought to be distributed according to the invariant distribution. The same thing can be said about the image shown in Fig. 4(a). Hence the first 1,000 sample points are discarded and the remaining 2,000 points are used to estimate both optimal labels and lines from Eq. 13, 14 and 15.

Fig. 3(b) and 4(b) show the segmentation results of the images in Fig. 3(a) and 4(a), respectively. The corresponding estimates of edge elements given by



**Fig. 2.** Variations of the label changing rates for sample paths generated by Gibbs sampling



**Fig. 3.** Results for image *monarch* (a) Original (b) Segmentation (c) Edge elements



**Fig. 4.** Results for image *home3* (a) Original (b) Segmentation (c) Edge elements

Eq. 15 are shown in Fig. 3(c) and 4(c). We could use the same parameter settings as these experiments to segment other images. In other words, this method robustly performs image segmentation without laborious parameter tuning.

## 6 Conclusions

In this paper, we have formulated the problem of image segmentation in the framework of Bayesian statistics and have proposed an approach to robustly estimating the posterior distribution of all the model parameters given an observed image. We have developed the hierarchical prior structure by regarding the model parameters as random variables and assuming the suitable distributions. This enables us to effectively implement the Gibbs sampler and to perform robust estimation. Experimental results show that our approach could achieve the favorable image segmentation.

## References

1. S. A. Barker, Image Segmentation using Markov Random Field Models, PhD thesis, University of Cambridge, 1998.
2. J. E. Besag, Spatial Interaction and The Statistical Analysis of Lattice Systems, *Journal of the Royal Statistical Society B*, vol. 36, pp. 192-236, 1974.
3. J. E. Besag, On The Statistical Analysis of Dirty Pictures, *Journal of the Royal Statistical Society B*, vol. 48, pp. 259-302, 1986.
4. M. K. Cowles and B. P. Carlin, Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review, *Journal of the American Statistical Association*, vol. 91, pp. 883-904, 1996.
5. S. Geman and D. Geman, Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721-741, 1984.
6. Z. Kato, Bayesian Color Image Segmentation Using Reversible Jump Markov Chain Monte Carlo, Research Report PNA-R9902, CWI, 1999.
7. S. Richardson and P. J. Green, On Bayesian Analysis of Mixtures with An Unknown Number of Components, *Journal of the the Royal Statistical Society B*, vol. 59, pp. 731-792, 1997.
8. M. Stephens, Bayesian Methods for Mixture of Normal Distributions, PhD thesis, University of Oxford, 1997.
9. Z. Tu and S.-C. Zhu, Image Segmentation By Data-Driven Markov Chain Monte Carlo, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 657-673, 2002.
10. J. Zhang, The Mean Field Theory in EM Procedures for Markov Random Fields, *IEEE Transactions on Signal Processing*, vol. 40, no. 10, pp. 2570-2583, 1992.
11. Y. Zhang, M. Brady, and S. Smith, Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm, *IEEE Transactions on Medical Imaging*, vol. 20, no. 1, pp. 45-57, 2001.