Smooth Image Segmentation by Nonparametric Bayesian Inference

Peter Orbanz and Joachim M. Buhmann
Institute of Computational Science, ETH Zurich
European Conference on Computer Vision (ECCV), 2006

The long version will appear on International Journal of Computer Vision (IJCV), 2009

Presented by Lan Du
Jan. 15, 2009

Outline

- Introduction
- Dirichlet process mixture (MDP) models
- Markov random fields (MRF)
- Dirichlet process mixtures constrained by Markov random fields (MDP/MRF)
- Application to image processing: Histogram clustering
- Experimental results
- Conclusions

Introduction

- Statistical approaches to image segmentation usually differ in two difficult design decisions, i.e. the statistical model for an individual segment and the number of segments.
  - Model choices: *k*-means clustering, histogram clustering, mixtures of Gaussians
  - Number of clusters: a priori knowledge, educated guessing, model selection methods (MDL or cross-validation)
- In this paper, the authors proposed a nonparametric Bayesian model for histogram clustering based on Dirichlet process mixture (MDP) and Markov random field (MRF) for image segmentation.
  - The model can automatically determine the number of segments.
  - The spatial smoothness constraints on the class assignments are enforced by a Markov random field.

Dirichlet process mixture models

- Data generation process

$$\boldsymbol{x}_i \sim F(\cdot|\theta_i), \quad \theta_i \sim G; \quad i = 1, \cdots, n$$

$$G \sim \mathrm{DP}(\alpha G_0)$$

- Property: MDP model is capable of adjusting the number of classes without switching models.

$$P(\theta_i|\theta_{-i}) = \sum_{k=1}^{N_c} \frac{n_k^{-i}}{n-1+\alpha} \delta_{\theta_k^*}(\theta_i) + \frac{\alpha}{n-1+\alpha} G_0(\theta_i)$$

x_i	:	data sample;	$F(\cdot)$	:	a parametric likelihood function
θ_i	:	mixture parameter of data sample x_i;	θ_{-i}	:=	$\{\theta_1, \cdots, \theta_{i-1}, \theta_{i+1}, \cdots, \theta_n\}$
G_0	:	the base measure;	α	:	a positive constant

n_k^{-i} : the number of samples in group k with the additional superscript indicating the exclusion of θ_i

N_c : the number of classes

Markov random fields

- Markov random fields provide an approach to modeling systems of dependent random variables.

$\mathcal{N} = \{V_{\mathcal{N}}, E_{\mathcal{N}}\}$: an undrirected neighborhood graph,

$V_{\mathcal{N}} = \{v_1, \cdots, v_n\}$: set of vertices; $\mathcal{C}$: completely connected subsets of $V_{\mathcal{N}}$

v_i : the vertex (site) is associated with an observation x_i and $x_i \sim F(\cdot|\theta_i)$

$E_{\mathcal{N}}$: set of graph edges

$\partial(i) := \{j|(i,j) \in E_{\mathcal{N}}\}$: dentes the index set of neighbors of v_i in $\mathcal{N}$

M : a MRF contribution term; H : a cost function defined on the neighborhood graph $\mathcal{N}$

$M(\theta_i|\theta_{-i}) \propto \exp(-H(\theta_i|\theta_{-i}));$ e.g. $H(\theta_i|\theta_{-i}) := \sum_{l\in\partial(i)} \|\theta_i - \theta_l\|^2$

$M(\theta_1, \cdots, \theta_n) := \frac{1}{Z_M} \exp\left(-\sum_{C\in\mathcal{C}} H_C(\theta_C)\right);$ $H(\theta_1, \cdots, \theta_n) = \sum_{C\in\mathcal{C}} H_C(\theta_C)$

- The conditional prior contribution will favor similar parameter values at sites which are neighbors.

Dirichlet process mixtures constrained by Markov random fields

- Data generation process

$$(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n) \sim \prod_{i=1}^n F(\boldsymbol{x}_i|\theta_i)$$

$$(\theta_1, \cdots, \theta_n) \sim M(\theta_1, \cdots, \theta_n) \prod_{i=1}^n G(\theta_i)$$

$$G \sim \mathrm{DP}(\alpha G_0)$$

$$P(\theta_1, \ldots, \theta_n|G) = \prod_{i=1}^n G(\theta_i)$$

$$\Pi(\theta_1, \cdots, \theta_n) \propto P(\theta_1, \cdots, \theta_n) M(\theta_1, \cdots, \theta_n)$$

$$\Pi(\theta_i|\theta_{-i}) \propto P(\theta_i|\theta_{-i}) M(\theta_i|\theta_{-i})$$

- Property:

$$\Pi(\theta_i|\theta_{-i}) \propto M(\theta_i|\theta_{-i}) \sum_{k=1}^{N_c} n_k^{-i} \delta_{\theta_k^*}(\theta_i) + \alpha M(\theta_i|\theta_{-i}) G_0(\theta_i)$$

$$\Pi(\theta_i|\theta_{-i}) \propto M(\theta_i|\theta_{-i}) \sum_{k=1}^{N_c} n_k^{-i} \delta_{\theta_k^*}(\theta_i) + \frac{\alpha}{Z_H} G_0(\theta_i)$$

  - The smoothness constraints on cluster assignments encourage consistent assignments within neighborhoods.
  - The site corresponding to a new draw from the base measurement will not be affected by the smoothness constraint.

Algorithm 1 (MDP/MRF Sampling)

Initialize: Generate a single cluster containing all points:

$$\theta_1^* \sim G_0(\theta_1^*) \prod_{i=1}^{n} F(\mathbf{x}_i|\theta_1^*) \ . \quad (34)$$

Repeat:

1. Generate a random permutation σ of the data indices.
2. *Assignment step.* For $i = \sigma(1), \ldots, \sigma(n)$:
   (a) If $\mathbf{x}_i$ is the only observation assigned to its cluster $k = S_i$, remove this cluster.
   (b) Compute the cluster probabilities

$$q_{i0} \propto \alpha \int_{\Omega_\theta} F(\mathbf{x}_i|\theta)\, G_0(\theta)\, d\theta$$
$$q_{ik} \propto n_k^{-i} \exp\left(-H(\theta_k^*|\theta_{-i})\right) F(\mathbf{x}_i|\theta_k^*) \quad (35)$$

   for $k = 1, \ldots, N_C$.
   (c) Draw a random index k according to the finite distribution $(q_{i0}, \ldots, q_{iN_C})$.
   (d) Assignment:
      – If $k \in \{1, \ldots, N_C\}$, assign $\mathbf{x}_i$ to cluster k.
      – If $k = 0$, create a new cluster for $\mathbf{x}_i$.
3. *Parameter update step.* For each cluster $k = 1, \ldots, N_C$: Update the cluster parameters θ_k^* given the class assignments $S_1, \ldots, S_n$ by sampling

$$\theta_k^* \sim G_0(\theta_k^*) \prod_{i|S_i=k} F(\mathbf{x}_i|\theta_k^*) \ . \quad (36)$$

Estimate assignment mode: For each point, choose the cluster it was assigned to most frequently during a given final number of iterations.

DP Formulation

$$q_{i0} \propto \alpha \int_{\Omega_\theta} F(\boldsymbol{x}_i|\theta)G_0(\theta)\mathrm{d}\theta$$

$$q_{ik} \propto n_k^{-i} F(\boldsymbol{x}_i|\theta_k^*)$$

Application to image processing: Histogram clustering

- The likelihood, base distribution and cost function:

$$F(\boldsymbol{h}_i|\boldsymbol{\theta}_i) = \frac{1}{Z_M(\boldsymbol{h}_i)} \exp\left(\sum_{j=1}^{N_{\text{bins}}} h_{ij} \log(\theta_{ij})\right) \qquad -- \qquad \text{Multinomial Distribution}$$

$$G_0(\boldsymbol{\theta}_i|\beta, \boldsymbol{\pi}) = \frac{1}{Z_D(\beta,\boldsymbol{\pi})} \exp\left(\sum_{j=1}^{N_{\text{bins}}} (\beta\pi_j - 1) \log(\theta_{ij})\right) \qquad -- \qquad \text{Dirichlet Distribution}$$

$$H(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{-i}) = -\lambda \sum_{l\in\partial(i)} \delta_{\boldsymbol{\theta}_i,\boldsymbol{\theta}_l}; \quad M(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{-i}) = \exp(-H(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{-i}))$$

$\boldsymbol{h}_i = (h_{i1}, \cdots, h_{iN_{\text{bins}}})$: a histogram drawn from the intensity values of all pixels within the ith patch

N_{bins} : the number of histogram bins

θ_{ij} : the probability for a value to occur in bin j of a histogram at site i

β : a positive scalar

$\boldsymbol{\pi}$: a N_{bins} – dimensional probability vector

- The resulting MRF will make a larger local contribution if more neighbors of a site are assigned to the same class, thereby encouraging spatial smoothness of cluster assignments.

Algorithm 1 (MDP/MRF Sampling)

Initialize: Generate a single cluster containing all points:

$$\theta_1^* \sim G_0(\theta_1^*) \prod_{i=1}^{n} F(\mathbf{x}_i|\theta_1^*) . \qquad (34)$$

Repeat:

1. Generate a random permutation σ of the data indices.
2. *Assignment step.* For $i = \sigma(1), \ldots, \sigma(n)$:
   (a) If $\mathbf{x}_i$ is the only observation assigned to its cluster $k = S_i$, remove this cluster.
   (b) Compute the cluster probabilities

$$q_{i0} \propto \alpha \int_{\Omega_\theta} F(\mathbf{x}_i|\theta)\, G_0(\theta)\, d\theta \qquad (35)$$
$$q_{ik} \propto n_k^{-i} \exp\left(-H(\theta_k^*|\theta_{-i})\right) F(\mathbf{x}_i|\theta_k^*)$$

   for $k = 1, \ldots, N_C$.
   (c) Draw a random index k according to the finite distribution $(q_{i0}, \ldots, q_{iN_C})$.
   (d) Assignment:
   - If $k \in \{1, \ldots, N_C\}$, assign $\mathbf{x}_i$ to cluster k.
   - If $k = 0$, create a new cluster for $\mathbf{x}_i$.
3. *Parameter update step.* For each cluster $k = 1, \ldots, N_C$: Update the cluster parameters θ_k^* given the class assignments $S_1, \ldots, S_n$ by sampling

$$\theta_k^* \sim G_0(\theta_k^*) \prod_{i|S_i=k} F(\mathbf{x}_i|\theta_k^*) \qquad (36)$$

Estimate assignment mode: For each point, choose the cluster it was assigned to most frequently during a given final number of iterations.

- Update equations:

$$\tilde{q}_{i0} := \frac{Z_D(\boldsymbol{h}_i + \beta\boldsymbol{\pi})}{Z_D(\beta\boldsymbol{\pi})}$$

$$\tilde{q}_{ik} := n_k^{-i} \exp\left(\lambda \sum_{l\in\partial(i)} \delta_{\boldsymbol{\theta}_i,\boldsymbol{\theta}_j} + \sum_j h_{ij} \log(\theta_{kj}^*)\right)$$

$$G_0(\boldsymbol{\theta}_k^*|\beta\boldsymbol{\pi}) \prod_{i|S_i=k} F(\boldsymbol{h}_i|\boldsymbol{\theta}_k^*) \propto G_0\left(\boldsymbol{\theta}_k^*|\beta\boldsymbol{\pi} + \sum_{i|S_i=k} \boldsymbol{h}_i\right)$$

Experimental results

- The unconstrained MDP model is applied to natural images (from the Corel database), which is sufficiently smooth not to require spatial constraints.

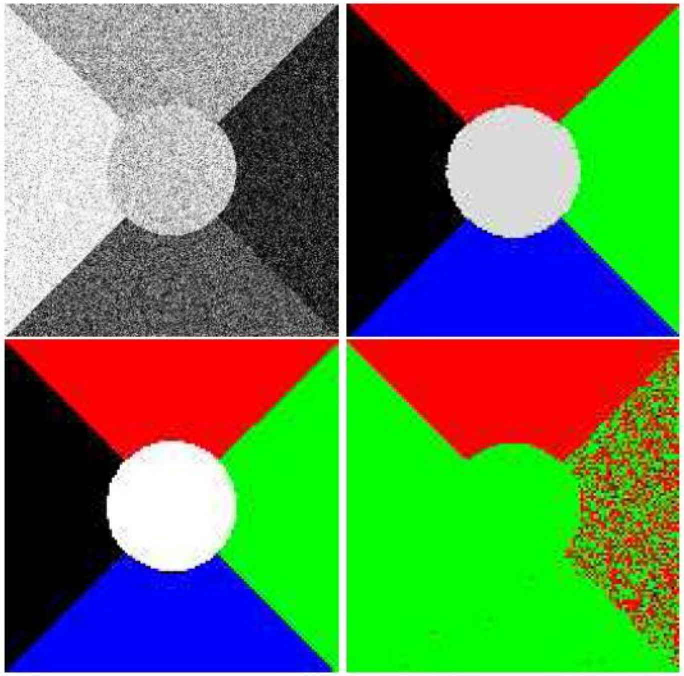

Fig. 1 Behavior of the unconstrained MDP sampler on an image with clearly defined segments. Upper row: Input image (left) and segmentation result for $\alpha = 10$ (right). Bottom row: Segmentation results for $\alpha = 10^{-4}$ (left) and $\alpha = 10^{-10}$.

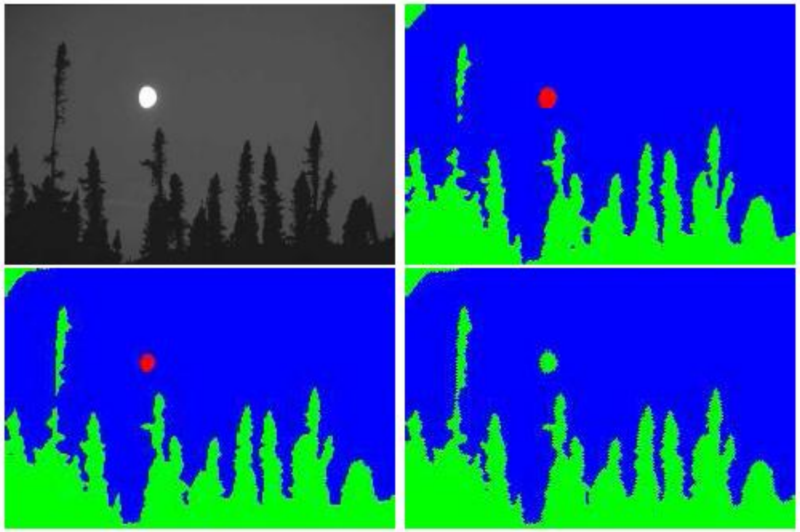

Fig. 2 Unconstrained MDP results on a simple natural image (Corel database): Original image (upper left), MDP results with $\alpha = 10^{-2}$ (upper right), $\alpha = 10^{-7}$ (bottom left), $\alpha = 10^{-9}$ (bottom right).

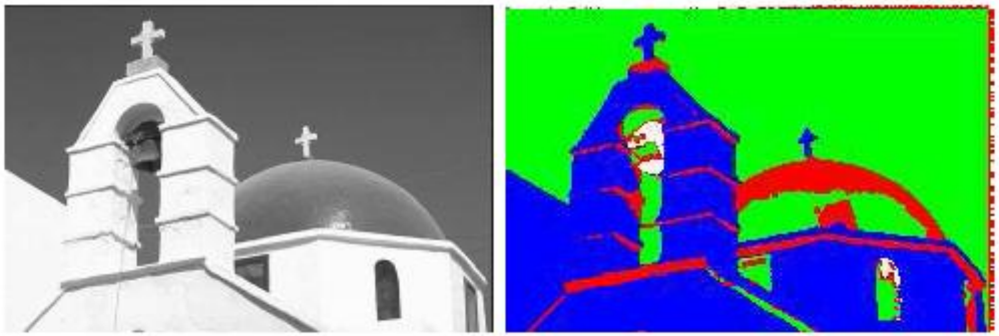

Fig. 3 Natural image (Corel database, left) and unconstrained MDP segmentation result (right).

- The MDP/MRF model is applied to synthetic aperture radar (SAR) images and magnetic resonance imaging (MRI) data.

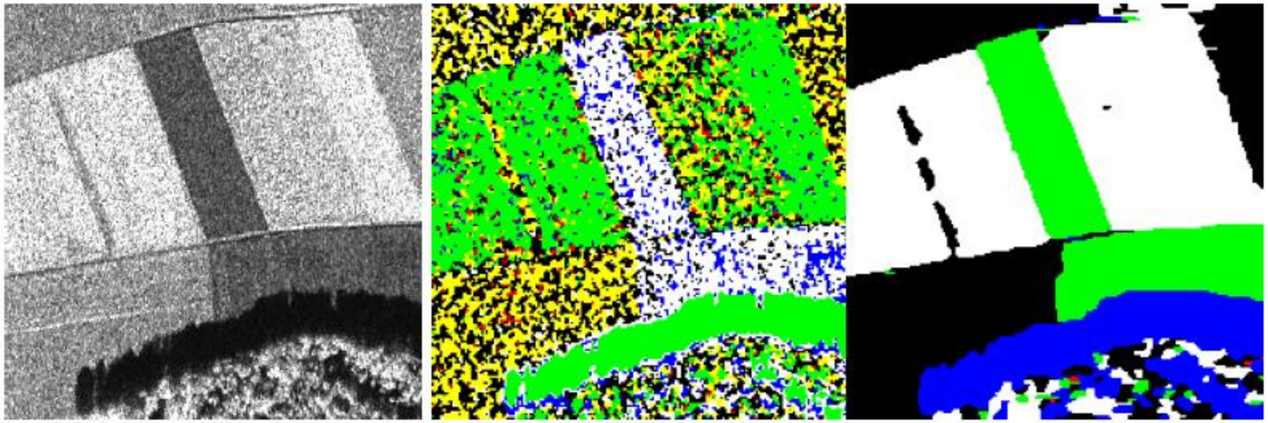

Fig. 1. Segmentation results on real-world radar data. Original image (left), unconstrained MDP segmentation (middle), MDP segmentation with smoothness constraint (right).

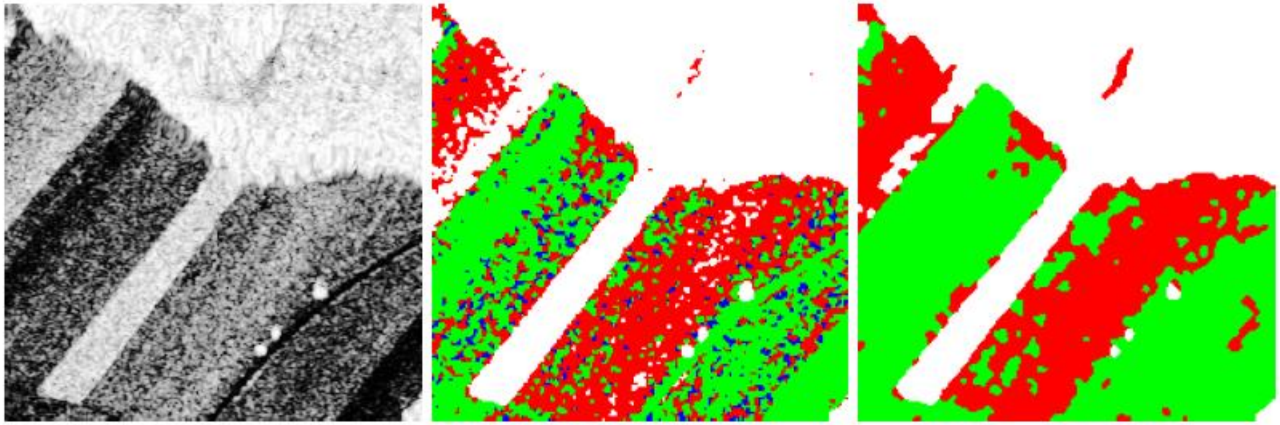

Fig. 2. A SAR image with a high noise level and ambiguous segments (left). Solutions without (middle) and with smoothing (right).

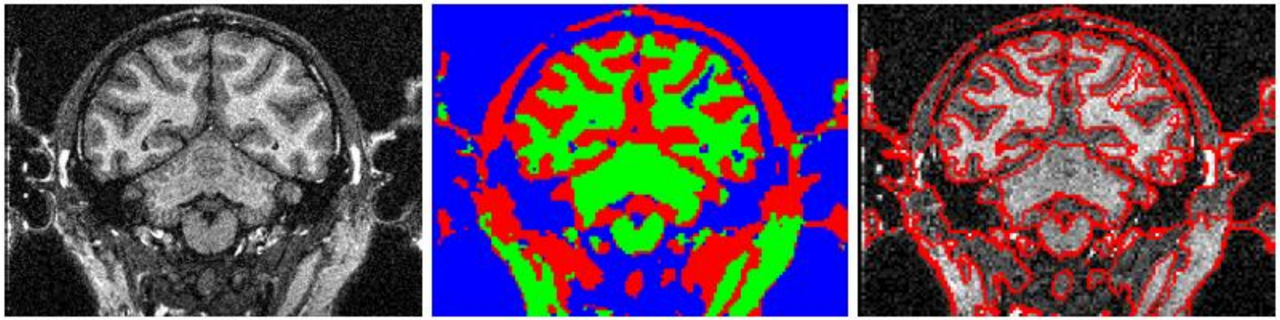

Fig. 3. MR frontal view image of a monkey's head. Original image (left), smoothed MDP segmentation (middle), original image overlaid with segment boundaries (right).

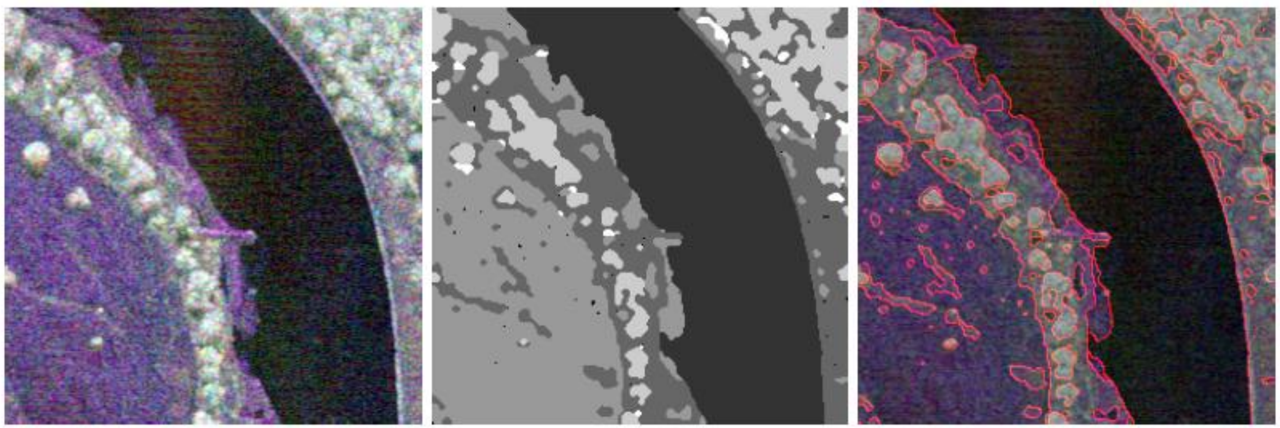

Fig. 4. Segmentation result for multichannel data: A SAR image with three channels (left), segmentation result obtained with the MDP/MRF model, and the original image overlaid with segment boundaries (right)

Stability is a cross-validation based wrapper method for an arbitrary clustering algorithm chosen by the user. An instability index is computed for different number of clusters, which measures how unstable cluster solutions are under the random split procedure. The chosen model is the one for which the instability index is minimal. Usually, a local rather than the global minimum is chosen. To obtain a valid comparison, the algorithm chosen for use with stability is an EM algorithm named ACM.

Table 1. Number of clusters chosen by the algorithm on two radar images for different values of the hyperparameter

α		1e-10	1e-9	1e-8	1e-7	1e-6	1e-5	1e-4	1e-3
Image Fig. 1	MDP	2	4	4	6	5	4	5	6
	smoothed	2	2	3	4	4	4	4	4
Images Fig. 2	MDP	4	3	4	7	6	5	5	9
	smoothed	2	2	3	4	5	3	3	5

Table 2. Stability indices computed with ACM clustering on two radar images for different numbers of clusters

N_C	Stability index		N_C	Stability index	
	Image Fig. 1	Image Fig. 2		Image Fig. 1	Image Fig. 2
2	0.0012 ± 0.0009	0.0003 ± 0.3341	6	0.4740 ± 0.0867	0.2933 ± 0.3437
3	0.3359 ± 0.2324	0.1765 ± 0.2856	7	0.5164 ± 0.0434	0.2907 ± 0.3007
4	0.3204 ± 0.2113	0.1233 ± 0.3481	8	0.5598 ± 0.0728	0.3532 ± 0.2889
5	0.2947 ± 0.0884	0.1436 ± 0.1929	9	0.6637 ± 0.0512	0.3378 ± 0.2801

For Image Fig. 1, the outcome of stability method is comparable to the result of the smoothed MDP model, which (except for very small values of α) selects three or four clusters. Both the MDP/MRF approach and the stability method give unreliable results on an image with a high noise level (Image Fig. 2) and poorly discernible segments.

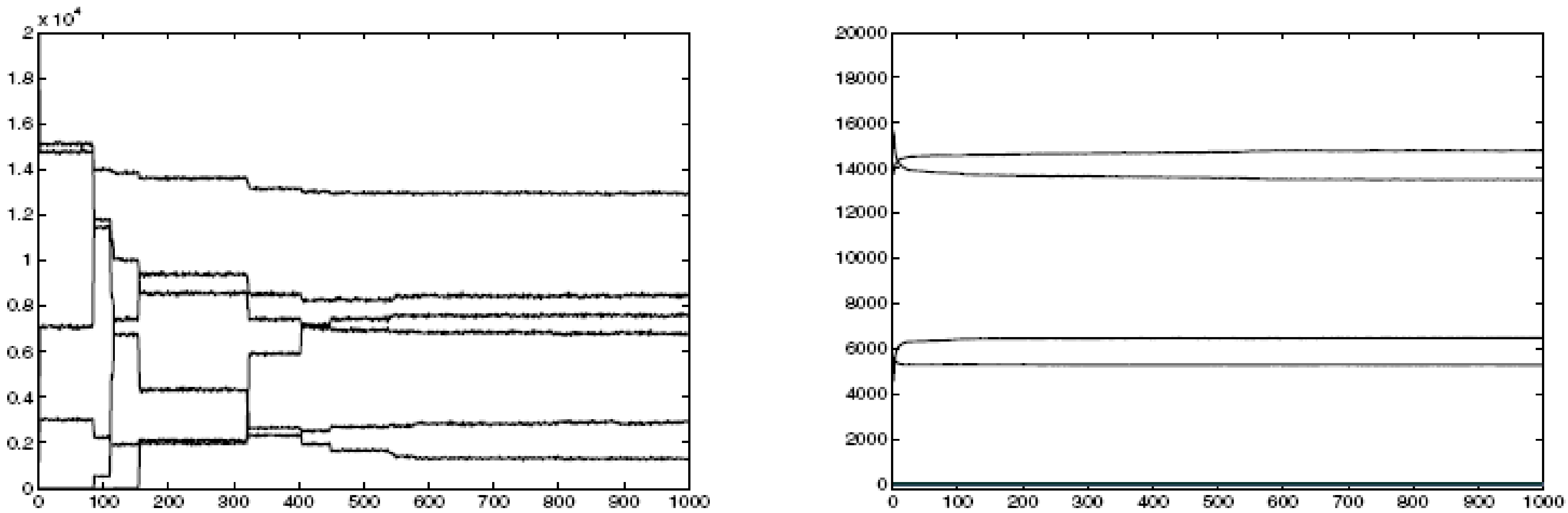

Fig. 5. Cluster sizes during the sampling process for the unconstrained and smoothed version of the MDP method. The number of sites assigned to each cluster (vertical) are drawn against the number of iterations (horizontal), with each graph representing a cluster. Left: Radar image (Fig. 1), no smoothing. Right: Same image, with smoothing.

Without smoothing, large batches of sites are suddenly reassigned from one cluster to another; while with smoothness constraints, clusters change gradually. Both the algorithms take about 600 iterations to stabilize.

Conclusions

- This paper summarizes the first attempt both to apply the Dirichlet nonparametric approach to image segmentation, and to combine it with Markov random fields.
- It is easy to extend such an MDP-based model to multiple images. The number of segments may vary from image to image, but the images are drawn from the same source or very similar sources.
- How to develop an efficient large-scale inference (*i.e.* VB) scheme for this model?