A Spatially-Constrained Normalized Gamma Process for Data Clustering

Sotirios P. Chatzis[1], Dimitrios Korkinof[2], and Yiannis Demiris[2]

[1] Cyprus University of Technology, Limassol, Cyprus
[2] Imperial College London, UK

Abstract. In this work, we propose a novel nonparametric Bayesian method for clustering of data with spatial interdependencies. Specifically, we devise a novel normalized Gamma process, regulated by a simplified (pointwise) Markov random field (Gibbsian) distribution with a countably infinite number of states. As a result of its construction, the proposed model allows for introducing spatial dependencies in the clustering mechanics of the normalized Gamma process, thus yielding a novel nonparametric Bayesian method for spatial data clustering. We derive an efficient truncated variational Bayesian algorithm for model inference. We examine the efficacy of our approach by considering an image segmentation application using a real-world dataset. We show that our approach outperforms related methods from the field of Bayesian nonparametrics, including the infinite hidden Markov random field model, and the Dirichlet process prior.

1 Introduction

Nonparametric Bayesian modeling techniques, especially Dirichlet process mixture (DPM) models, have become very popular in statistics over the last few years, for performing nonparametric density estimation [1,2,3]. This theory is based on the observation that an infinite number of component distributions in an ordinary finite mixture model (clustering model) tends on the limit to a Dirichlet process (DP) prior [2,4]. Eventually, the nonparametric Bayesian inference scheme induced by a DPM model yields a posterior distribution on the proper number of model component densities (inferred clusters) [5], rather than selecting a fixed number of mixture components.

Markov random fields (MRFs) [6] are a classical methodology for modeling spatially-interdependent data. In essence, MRFs impose a Gibbsian distribution over the allocation of the modeled data into states (clusters), which enforces the belief that spatially adjacent data are more likely to cluster together. As the Gibbsian prior imposed by MRFs entails complex calculations that make it intractable in real-world problems dealing with large datasets, efficient approximations of the full MRF distribution are usually employed. For example, a pointwise simplification of the MRF prior based on the *mean-field principle* from statistical mechanics [7] was employed in [8]. Recently, MRFs have also

been used in the context of Bayesian nonparametrics yielding the infinite hidden Markov random field (iHMRF) model [9,10].

Inspired by these advances, in this paper we come up with a different approach towards clustering data with spatial interdependencies. We propose a spatially-adaptive random measure, coined the Markov random field normalized Gamma process (MRF-NGP). Our model is based on the introduction of a normalized Gamma process (NGP) controlled by an additionally postulated pointwise Markov random field imposed over the data allocation into model states, obtained by application of the mean-field principle [9]. We evaluate the efficacy of our approach considering an unsupervised image segmentation application using benchmark data.

2   Theoretical Background

2.1   The Dirichlet Process

Dirichlet process (DP) models were first introduced in [11]. A DP is characterized by a base distribution G_0 and a positive scalar α, usually referred to as the innovation parameter, and is denoted as $\mathrm{DP}(\alpha, G_0)$. Essentially, a DP is a distribution placed over a distribution. Let us suppose we randomly draw a sample distribution G from a DP, and, subsequently, we independently draw M random variables $\{\Theta_m^*\}_{m=1}^M$ from G:

$$G|\alpha, G_0 \sim \mathrm{DP}(\alpha, G_0) \tag{1}$$

$$\Theta_m^*|G \sim G, \quad m = 1, \dots M \tag{2}$$

Integrating out G, the joint distribution of the variables $\{\Theta_m^*\}_{m=1}^M$ can be shown to exhibit a clustering effect. Specifically, given the first $M-1$ samples of G, $\{\Theta_m^*\}_{m=1}^{M-1}$, it can be shown that a new sample Θ_M^* is either (a) drawn from the base distribution G_0 with probability $\frac{\alpha}{\alpha+M-1}$, or (b) is selected from the existing draws, according to a multinomial allocation, with probabilities proportional to the number of the previous draws with the same allocation [12].

2.2   Markov Random Fields

We consider an alphabet $Q = \{1, ..., K\}$. Let S be a finite index set, $S = \{1, ..., N\}$; we shall refer to this set, S, as the set of sites or locations. Let us consider for every site $j \in S$ a finite space $\mathcal{Z}_j$ of states z_j taking on Q. The product space $\mathcal{Z} = \prod_{j=1}^N \mathcal{Z}_j$ will be denoted as the space of the configurations of the state values of the considered sites set, $\boldsymbol{z} = (z_j)_{j \in S}$. A strictly positive probability distribution, $p(\boldsymbol{z})$, $\boldsymbol{z} \in \mathcal{Z}$, on the product space $\mathcal{Z}$ is called a random field [13].

Let ∂ denote a neighborhood system on S, i.e. a collection $\partial = \{\partial_j : j \in S\}$ of sets, such as $j \notin \partial_j$ and $l \in \partial_j$ if and only if $j \in \partial_l \ \forall l, j \in S$. Then, the

previously considered random field, $p(\boldsymbol{z})$, is a Markov random field with respect to the introduced neighborhood system ∂ if [14] $p(z_j|\boldsymbol{z}_{S-\{j\}}) = p(z_j|\boldsymbol{z}_{\partial_j}) \ \forall j \in S$. The distribution $p(\boldsymbol{z})$ of a Markov random field can be shown to be of a Gibbsian form [15]:

$$p(\boldsymbol{z}) \triangleq \frac{1}{W(\gamma)} \exp\left(-\sum_{c\in\mathcal{C}} V_c(\boldsymbol{z}|\gamma)\right) \tag{3}$$

where γ is the inverse temperature of the model, $W(\gamma)$ is the (normalizing) partition function of the model, $V_c(\boldsymbol{z}|\gamma)$ are the clique potentials of the model, and $\mathcal{C}$ is the set of the cliques included in the model neighborhood system.

A significant problem of MRF models concerns computational tractability, as the normalizing term $W(\gamma)$ is hard to compute in applications dealing with large datasets. One way to resolve these issues is the mean-field approximation [7,16]. It is based on the idea of neglecting the fluctuations of the sites interacting with a considered site, so that the resulting system behaves as one composed of independent variables for which computation becomes tractable. That is, given an estimate $\hat{\boldsymbol{z}}$ of the unknown site labels vector $\boldsymbol{z}$, obtained by means of a stochastic restoration criterion, such as the iterative conditional modes (ICM) or the marginal posterior modes (MPM) algorithm (see, e.g., [14,16]), we make the hypothesis [17]

$$p(\boldsymbol{z}) = \prod_{j=1}^{N} p(z_j|\hat{\boldsymbol{z}}_{\partial_j};\gamma) \tag{4}$$

where

$$p(z_j = i|\hat{\boldsymbol{z}}_{\partial_j};\gamma) = \frac{\exp(-\sum_{c\ni j} V_c(\tilde{\boldsymbol{z}}_{ij}|\gamma))}{\sum_{h=1}^{K} \exp(-\sum_{c\ni j} V_c(\tilde{\boldsymbol{z}}_{hj}|\gamma))} \tag{5}$$

$\tilde{\boldsymbol{z}}_{ij} \triangleq (z_j = i, \hat{\boldsymbol{z}}_{\partial_j})$, $\hat{\boldsymbol{z}}_{\partial_j}$ is the estimate of the jth site neighborhood, and the indexes c refer to the cliques that contain the jth site.

3   Proposed Approach

3.1   Model Formulation

Let us consider a set of observations $Y = \{\boldsymbol{y}_n\}_{n=1}^{N}$, $\boldsymbol{y}_n \in \mathcal{Y}$, measured over a set of sites $\mathcal{S} = \{1, ..., S\}$ on which a neighborhood system ∂ is defined. Let us denote as $X = \{x_n\}_{n=1}^{N}$, $x_n \in \mathcal{S}$, the sites where the observed data points $\{\boldsymbol{y}_n\}_{n=1}^{N}$ were measured. Let us introduce the latent variables $\{z_n\}_{n=1}^{N}$ denoting the model state (cluster) where an observed data point $\boldsymbol{y}_n$ measured at the location x_n is assigned by our model. Motivated by the merits and the theory of the DP discussed in the previous section, to derive the sought model, we make the key-assumption, based on the mean-field-based approximation of the MRF distribution, that for any given site x_n, we have available an estimate $\hat{\boldsymbol{z}}_{\partial_n}$ of the value $\boldsymbol{z}_{\partial_n} \triangleq (z_m)_{m\in\partial_n}$ of the latent cluster assignment variables of the observations measured at sites in the neighborhood of site x_n. Apparently, this assumption entails obtaining an initial estimate of the latent variables $\{z_n\}_{n=1}^{N}$ for the modeled data, as discussed

in Section 2.2. Further, we consider the following predictor (location)-dependent random measure

$$G(x) = \sum_{i=1}^{\infty} \varpi_i(x) \delta_{\Theta_i} \tag{6}$$

where

$$\varpi_i(x) = \frac{\Lambda_i(x)}{\sum_{j=1}^{\infty} \Lambda_j(x)} \tag{7}$$

the random variables Λ_i follow a Gamma distribution as

$$\Lambda_i|x_n \sim \mathcal{G}(\alpha k_i(x_n; \hat{\boldsymbol{z}}_{\partial_n}), 1) \tag{8}$$

α is the innovation parameter of the process, $k_i(x_n; \hat{\boldsymbol{z}}_{\partial_n})$ is the probability of the nth site being assigned to the ith cluster as computed by the employed pointwise MRF distribution

$$k_i(x_n; \hat{\boldsymbol{z}}_{\partial_n}) \triangleq p(z_n = i | \hat{\boldsymbol{z}}_{\partial_n}; \gamma) = \frac{\exp(-\sum_{c \ni x_n} V_c(\tilde{\boldsymbol{z}}_{ni}|\gamma))}{\sum_{h=1}^{\infty} \exp(-\sum_{c \ni x_n} V_c(\tilde{\boldsymbol{z}}_{nh}|\gamma))} \tag{9}$$

$\tilde{\boldsymbol{z}}_{ni} \triangleq (z_n = i, \hat{\boldsymbol{z}}_{\partial_n})$, $\hat{\boldsymbol{z}}_{\partial_n}$ is the current estimate of the nth site neighborhood, $V_c(\cdot)$ are the employed clique potential functions, and the indexes c refer to the cliques that include the nth site, x_n. The utility of the pointwise MRF distribution $k_i(x_n; \hat{\boldsymbol{z}}_{\partial_n})$ in our model consists in reducing the probability (discounting) of clusters that seem rather unlikely from the viewpoint of the postulated neighborhood system. We dub this random probability measure $G(x)$ the MRF-NGP process. A proof that the normalizing constant in the denominator of (7) is finite almost surely is provided in the Appendix.

3.2   Variational Bayesian Inference

Let us a consider a set of observations $Y = \{\boldsymbol{y}_n\}_{n=1}^N$ with corresponding locations $X = \{x_n\}_{n=1}^N$. We postulate for our observed data a likelihood function of the form

$$p(\boldsymbol{y}_n|z_n = i) = p(\boldsymbol{y}_n|\boldsymbol{\theta}_i) \tag{10}$$

while for the latent assignment variables z_n we consider

$$p(z_n = i|x_n) = \varpi_i(x_n) \tag{11}$$

where the $\varpi_i(x)$ are given by (7), with the prior over the $\Lambda_i(x)$ given by (8). Regarding the likelihood parameters $\boldsymbol{\theta}_i$, we impose a suitable conjugate exponential prior over them; for instance, in case of a Gaussian likelihood function

$$p(\boldsymbol{y}_n|\boldsymbol{\theta}_i) = \mathcal{N}(\boldsymbol{y}_n|\boldsymbol{\mu}_i, \boldsymbol{R}_i) \tag{12}$$

we impose a Normal-Wishart prior over the likelihood parameters $\boldsymbol{\theta}_i = \{\boldsymbol{\mu}_i, \boldsymbol{R}_i\}$, i.e.

$$p(\boldsymbol{\mu}_i, \boldsymbol{R}_i) = \mathcal{NW}(\boldsymbol{\mu}_i, \boldsymbol{R}_i|\lambda_i, \boldsymbol{m}_i, \omega_i, \boldsymbol{\Omega}_i) \tag{13}$$

Regarding the MRF temperature parameter γ, and the innovation parameter α, we choose to optimize them as model hyperparameters, as part of the variational inference procedure discussed next.

Our variational Bayesian inference formalism consists in derivation of a family of variational posterior distributions $q(.)$ which approximate the true posterior distribution over the infinite sets $\{z_n\}_{n=1}^N$, $\{\Lambda_i(x_n)\}_{i,n=1}^{\infty,N}$, and $\{\boldsymbol{\theta}_i\}_{i=1}^\infty$. Apparently, under this infinite dimensional setting, Bayesian inference is not tractable. For this reason, we fix a value K and we let the variational posterior over the $\Lambda_k(x)$ have the property $q(\Lambda_{k>K}(x) = 0) = 1$, $\forall x \in \mathcal{S}$. In other words, we set $\varpi_k(x)$ equal to zero for $k > K$, $\forall x \in \mathcal{S}$. Note that, under this setting, the treated model involves a full MRF-NGP prior; truncation is not imposed on the MRF-NGP prior itself, but only on the variational distribution to allow for a tractable inference procedure [18].

Letting W be the parameters of our model, and Ξ the entailed hyperparameters, including γ, α, and the hyperparameters of the prior over the likelihood parameters, variational Bayesian inference consists in derivation of an approximate posterior $q(W)$ by maximization (in an iterative fashion) of the variational free energy

$$\mathcal{L}(q) = \int \mathrm{d}W q(W) \log \frac{p(X, Y, W|\Xi)}{q(W)} \tag{14}$$

which provides a lower bound to the computationally intractable log marginal likelihood (log evidence), $\log p(X, Y)$, of the model [19]. The derived algorithm is in essence an expectation-maximization-like algorithm. Each iteration comprises an E-step, on which the variational posteriors over the model latent variables are computed, and an M-step, on which the variational posteriors over the model parameters are updated. Let us denote as $\langle . \rangle$ the posterior expectation of a quantity.

M-Step. This step comprises the updates of the Gamma-distributed variables $\Lambda_i(x_n)$

$$q(\Lambda_i(x_n)) = \mathcal{G}(\Lambda_i(x_n)|\beta_{ni}, \xi_{ni}) \tag{15}$$

where

$$\beta_{ni} = \alpha k_i(x_n; \hat{\boldsymbol{z}}_{\partial_n}) + q(z_n = i) \tag{16}$$

$$\xi_{ni} = 1 + \frac{1}{\sum_{j=1}^K \langle \Lambda_j(x_n) \rangle} \tag{17}$$

and

$$\langle \Lambda_j(x_n) \rangle = \frac{\beta_{nj}}{\xi_{nj}} \tag{18}$$

as well as of the parameters $\boldsymbol{\theta}_i$, for which we obtain the general solution

$$\log q(\boldsymbol{\theta}_i) \propto \log p(\boldsymbol{\theta}_i) + \sum_{n=1}^N q(z_n = i) \log p(\boldsymbol{y}_n|\boldsymbol{\theta}_i) \tag{19}$$

which is similar to the corresponding solution for models imposing simple DP priors over their cluster assignment distributions.

E-Step. This step comprises the updates of the posteriors $q(z_n = i)$:

$$q(z_n = i) \propto \exp\left(\langle \log \Lambda_i(x_n) \rangle\right) \exp\left(\varphi_{ni}\right) \tag{20}$$

where

$$\langle \log \Lambda_i(x_n) \rangle = \psi(\beta_{ni}) - \log \xi_{ni} \tag{21}$$

and

$$\varphi_{nj} = \langle \log p(\boldsymbol{y}_n | \boldsymbol{\theta}_j) \rangle \tag{22}$$

It also consists in updating the estimates of the assignment variables $\hat{\boldsymbol{z}} = (\hat{z}_n)_{n=1}^N$ which are used in computing the pointwise MRF priors employed in our model to regulate cluster discounting. For this purpose, we simply set

$$\hat{z}_n = \operatorname{argmax}_{i=1}^K q(z_n = i) \tag{23}$$

Finally, regarding the model hyperparameters Ξ, in this work we obtain estimates of the hyperparameter γ by maximization of the lower bound $\mathcal{L}(q)$, and we heuristically select the values of the rest of the model hyperparameters.
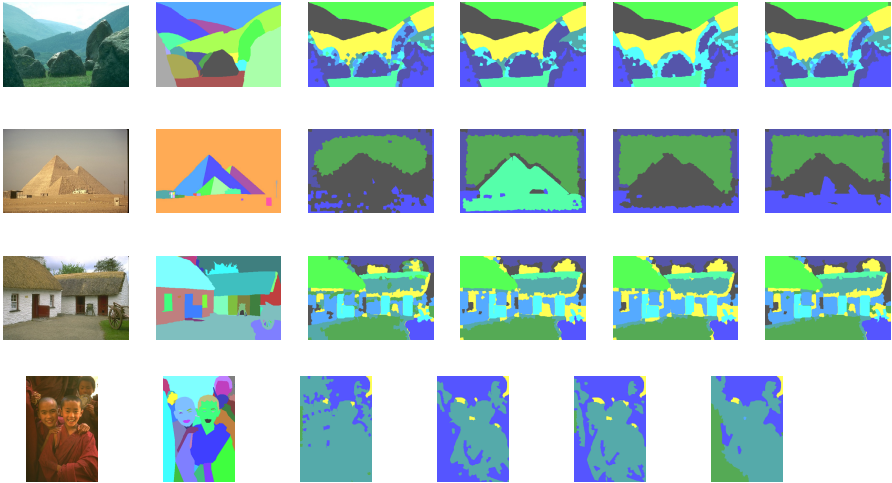

Fig. 1. Few selected 321x481 images from the Berkeley image segmentation dataset. **Left-to-right**: a) Original image, b) One human groundtruth, c) K-means initialization, d) iHMRF, e) MRF-NGP. **Top-to-bottom**: a) #241004, b) #161062, c) #385028, d) #246053.

Table 1. Obtained PRI results for the considered subset of the Berkeley benchmark

Image #	DPM	iHMRF	MRF-NGP	Image #	DPM	iHMRF	MRF-NGP
159029	0.7688	0.7727	0.7842	246053	0.5896	0.6006	0.6526
20008	0.8376	0.8514	0.8478	239096	0.7570	0.7957	0.7871
100075	0.7851	0.7795	0.7861	323016	0.8283	0.8436	0.8479
301007	0.8438	0.8432	0.8460	231015	0.8019	0.8185	0.8138
122048	0.7396	0.7520	0.7421	25098	0.8270	0.8231	0.8394
145053	0.6189	0.6315	0.7304	8143	0.6294	0.6605	0.7011
236017	0.5997	0.6035	0.6346	35010	0.7701	0.7854	0.8051
170054	0.6841	0.7453	0.7628	15004	0.7561	0.7865	0.7994
385028	0.8520	0.8393	0.8544	100080	0.8067	0.8093	0.8036
67079	0.7344	0.7347	0.7599	161062	0.6610	0.6280	0.6742
209070	0.6335	0.7006	0.7380	159045	0.6984	0.7315	0.7482
27059	0.8359	0.8470	0.8669	170057	0.6948	0.7243	0.7498
176019	0.6930	0.7470	0.7343	89072	0.7377	0.7590	0.7841
103070	0.7164	0.7307	0.7530	175032	0.5594	0.6783	0.6686
241004	0.8643	0.8642	0.8738	86016	0.7379	0.7664	0.7573

4 Experimental Evaluation

We investigate the efficacy of our approach considering an unsupervised image segmentation application. Specifically, we consider segmentation of real-world images, using a subset of the Berkeley image segmentation benchmark [20]. The Berkeley image segmentation benchmark comprises real-world color images along with their segmentation maps provided by different individuals. To obtain an objective performance evaluation of the proposed algorithm, we employ the probabilistic rand index (PR index or PRI) [21]. The PR index counts the fraction of pairs of pixels whose labelings are consistent between a computed segmentation and the given groundtruth, averaging across multiple groundtruth segmentations to account for scale variation in human perception. In all our experiments, we use Gaussian likelihoods, and choose to impose a Normal-Wishart prior over the likelihood parameters.

We compare the performance of our approach to iHMRF and DPM, both using the same likelihood function and prior over the likelihood function parameters as in the case of our model. All the evaluated algorithms are initialized by means of the k-means algorithm. Regarding the potential functions of the imposed pointwise MRFs for both the evaluated iHMRF and MRF-NGP models, we opt for a simple Potts model with a second order (8-neighbors) neighborhood system, yielding

$$p(z_n = c | \hat{\boldsymbol{z}}_{\partial_n}; \gamma) = \frac{\exp(\gamma \sum_{l \in \partial_n} \delta(c - \hat{z}_l))}{\sum_{h=1}^{K} \exp(\gamma \sum_{l \in \partial_n} \delta(h - \hat{z}_l))} \tag{24}$$

for the pointwise MRF priors, where K is the truncation threshold, and $\delta(.)$ stands for the Kronecker's delta function.

Feature extraction is effected as follows: First, each image is segmented into approximately $N = 1000$ superpixels using the method proposed in [22]. We then compute feature vectors at superpixel level, comprising RGB and HSV color information along with the values of the Maximum Response (MR) filter banks [23]. The truncation level of the variational Bayesian algorithm for all the treated models is set to $K = 10$.

To account for the effect of poor model initialization, which may lead model training to yield bad local optima as model estimators, we execute our experiments multiple times for each image, with different initializations each time, common for all the evaluated algorithms. The visual segmentation result is presented for 4 selected images in Fig. 1, along with the original image, one human groundtruth, and the initialization. The mean PRI results (over the executed repetitions) for the whole considered dataset are presented in Table 1. Total results across all images are presented in Table 2. Based on the obtained PRI metric results, we can conclude that the MRF-NGP performs considerably better than all the considered rival methods. Note also that small differences in the values of the PRI metric correspond to significant differences in the quality of the obtained segmentation results [24]. The illustrated segmentation results vouch for this assertion.

Table 2. Mean and median of the PRI metric across the considered subset of the Berkeley benchmark

PRI(%)	DPM	iHMRF	MRF-NGP
Mean	73.54	75.51	77.15
Median	73.88	76.27	77.34

5   Conclusions

In this paper, we proposed a method for nonparametric clustering of data with general spatial interdependencies. Our method, coined the MRF-NGP, consists in postulating a normalized Gamma-process, the cluster prior probabilities of which are discounted by means of a simplified pointwise Markov random field imposed over data point allocation into clusters. As a result of this construction, the MRF-NGP imposes the belief that spatially proximate data are more likely to cluster together. To examine the efficacy of our approach, we evaluated it in unsupervised image segmentation tasks using a real-life benchmark dataset, namely the Berkeley image segmentation benchmark. We showed that it yields a considerable improvement in the obtained performance of the clustering algorithm compared to both the DPM, and the recently proposed iHMRF model.

Appendix

Here, we prove the almost sure finiteness of the normalizing factor $\sum_{j=1}^{\infty} \Lambda_j(x_n)$ in (7). Let

$$S_T \triangleq \sum_{j=1}^{T} \Lambda_j(x_n) \tag{25}$$

It follows that $S_1 \leq S_2 \leq \cdots \leq S_T \leq \cdots \leq S$, where

$$S \triangleq \lim_{T \to \infty} S_T \tag{26}$$

since the random variables $\Lambda_j(x_n)$ are non-negative, as they follow a Beta distribution.

Then, to prove that S is finite almost surely, we only need to prove that $\mathbb{E}[S]$ is finite. From the monotone convergence theorem, we yield

$$\mathbb{E}[S] = \lim_{T \to \infty} \mathbb{E}[S_T] = \lim_{T \to \infty} \sum_{j=1}^{T} \mathbb{E}[\Lambda_j(x_n)] = \alpha \tag{27}$$

since $\lim_{T \to \infty} \sum_{j=1}^{T} k_j(x_n; \hat{\boldsymbol{z}}_{\partial_n}) = 1$, as the $k_j(x_n; \hat{\boldsymbol{z}}_{\partial_n})$ comprise prior MRF-derived probabilities of the observation at the nth site being assigned to any of the postulated model states. Hence, we have proven that S is finite almost surely.

References

1. Walker, S., Damien, P., Laud, P., Smith, A.: Bayesian nonparametric inference for random distributions and related functions. J. Roy. Statist. Soc. B 61(3), 485–527 (1999)
2. Neal, R.: Markov chain sampling methods for Dirichlet process mixture models. J. Comput. Graph. Statist. 9, 249–265 (2000)
3. Muller, P., Quintana, F.: Nonparametric Bayesian data analysis. Statist. Sci. 19(1), 95–110 (2004)
4. Antoniak, C.: Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. The Annals of Statistics 2(6), 1152–1174 (1974)
5. Blei, D., Jordan, M.: Variational methods for the Dirichlet process. In: 21st Int. Conf. Machine Learning, New York, NY, USA, pp. 12–19 (July 2004)
6. Orbanz, P., Buhmann, J.: Nonparametric Bayes image segmentation. International Journal of Computer Vision 77, 25–45 (2008)
7. Zhang, J.: The mean field theory in EM procedures for Markov random fields. IEEE Transactions on Image Processing 2(1), 27–40 (1993)
8. Celeux, G., Forbes, F., Peyrard, N.: EM procedures using mean field-like approximations for Markov model-based image segmentation. Pattern Recognition 36(1), 131–144 (2003)
9. Chatzis, S.P., Tsechpenakis, G.: The infinite hidden Markov random field model. In: Proc. 12th International IEEE Conference on Computer Vision (ICCV), Kyoto, Japan, pp. 654–661 (September 2009)

10. Chatzis, S.P., Tsechpenakis, G.: The infinite hidden Markov random field model. IEEE Transactions on Neural Networks 21(6), 1004–1014 (2010)
11. Ferguson, T.: A Bayesian analysis of some nonparametric problems. The Annals of Statistics 1, 209–230 (1973)
12. Blackwell, D., MacQueen, J.: Ferguson distributions via Pólya urn schemes. The Annals of Statistics 1(2), 353–355 (1973)
13. Maroquin, J., Mitte, S., Poggio, T.: Probabilistic solution of ill-posed problems in computational vision. Journal of the American Statistical Assocation 82, 76–89 (1987)
14. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 721–741 (1984)
15. Clifford, P.: Markov random fields in statistics. In: Grimmett, G., Welsh, D. (eds.) Disorder in Physical Systems. A volume in Honour of John M. Hammersley on the Occasion of His 70th Birthday. Oxford Science Publication, Clarendon Press, Oxford (1990)
16. Chatzis, S.P., Varvarigou, T.A.: A fuzzy clustering approach toward hidden Markov random field models for enhanced spatially constrained image segmentation. IEEE Transactions on Fuzzy Systems 16(5), 1351–1361 (2008)
17. Qian, W., Titterington, D.: Estimation of parameters in hidden Markov models. Philosophical Transactions of the Royal Society of London A 337, 407–428 (1991)
18. Blei, D.M., Jordan, M.I.: Variational inference for Dirichlet process mixtures. Bayesian Analysis 1(1), 121–144 (2006)
19. Jordan, M., Ghahramani, Z., Jaakkola, T., Saul, L.: An introduction to variational methods for graphical models. In: Jordan, M. (ed.) Learning in Graphical Models, pp. 105–162. Kluwer, Dordrecht (1998)
20. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proc. 8th Int'l Conf. Computer Vision, Vancouver, Canada, pp. 416–423 (July 2001)
21. Unnikrishnan, R., Pantofaru, C., Hebert, M.: A measure for objective evaluation of image segmentation algorithms. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, San Diego, CA, USA, pp. 34–41 (June 2005)
22. Mori, G.: Guiding model search using segmentation. In: Proc. 10th IEEE Int. Conf. on Computer Vision, ICCV (2005)
23. Varma, M., Zisserman, A.: Classifying Images of Materials: Achieving Viewpoint and Illumination Independence. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part III. LNCS, vol. 2352, pp. 255–271. Springer, Heidelberg (2002)
24. Nikou, C., Galatsanos, N., Likas, A.: A class-adaptive spatially variant mixture model for image segmentation. IEEE Transactions on Image Processing 16(4), 1121–1130 (2007)