JOINT SEGMENTATION OF MULTIPLE IMAGES WITH SHARED CLASSES: A BAYESIAN NONPARAMETRICS APPROACH

Jessica Sodjo[1], *Audrey Giremus*[1], *François Caron*[2], *Jean-François Giovannelli*[1], *Nicolas Dobigeon*[3]

[1] IMS (Univ. Bordeaux, CNRS, BINP). F-33400 Talence, France
[2] Univ. Oxford, Department Statistics. Oxford, United Kingdom
[3] University of Toulouse, IRIT/INP-ENSEEIHT, F-31071 Toulouse, France

ABSTRACT

A combination of the hierarchical Dirichlet process (HDP) and the Potts model is proposed for the joint segmentation/classification of a set of images with shared classes. Images are first divided into homogeneous regions that are assumed to belong to the same class when sharing common characteristics. Simultaneously, the Potts model favors configurations defined by neighboring pixels belonging to the same class. This HDP-Potts model is elected as a prior for the images, which allows the best number of classes to be selected automatically. A Gibbs sampler is then designed to approximate the Bayesian estimators, under a maximum a posteriori (MAP) paradigm. Preliminary experimental results are finally reported using a set of synthetic images.

Index Terms— Image, segmentation, Bayesian nonparametrics, hierarchical Dirichlet process, Potts model.

1. INTRODUCTION

Image segmentation is a key step for image analysis and interpretation. It has numerous applications such as video surveillance [1] and medical diagnosis [2], among others. Image segmentation consists of dividing the observed scene in spatially homogeneous regions. It is formalized by assigning a unique label value to the set of pixels sharing common characteristics (e.g., pixel statistics). In this paper, this set is referred to as a class, defined by pixels within a given class distributed according to the same probability density function.

Segmenting a collection of multiple images can be conducted by resorting to a dedicated algorithm applied on each image separately. Among popular segmentation algorithms, the Bayesian ones exploit a prior model on the image. Within the Bayesian framework, a standard approach consists of resorting to the Potts model as a prior for the labels [3]. It strengthens that neighboring pixels share the same label value. However, in this approach, the number of classes should be set beforehand, e.g, using prior knowledge or by cross-validation. Selecting automatically the best number of clusters is a challenge. It can be automatically inferred by using estimation algorithms able to explore parameter spaces with variable dimensions, such as reversible jump Markov chain Monte Carlo (RJMCMC) [4] algorithms, or by comparing the concurrent models of distinct dimensions with an appropriate criterion (e.g., Bayesian information criterion [5]). More recently, Bayesian nonparametrics (BNP) approaches have proven their efficiency to address this issue [6]. BNP methods allow the classes assignments and the corresponding parameters (including the number of classes) to be deduced by exploring the observations which can be infinite. An other method is to combine one of the BNP, like the Dirichlet Process, with the Potts model [7].

This paper considers the specific case of a set of images with shared classes to be segmented jointly. For example, when analyzing a pair of images of urban and suburban areas, respectively, the classes "tree" and "house" are expected to be shared with different proportions. In that case, it could be relevant to take the strong shared statistical information into account to improve the class parameter estimation. The hierarchical Dirichlet process (HDP) [8] is a suitable extension of the Dirichlet process (DP) to address this issue. The first stage of the modeling identifies groups of similar pixels, or *regions*, in each image. Then, in a second stage, the regions with same characteristics among the collection of images to be analyzed are identified as belonging to the same class. Furthermore, in this work, we propose to combine the HDP with the Potts model. Thanks to the use of HDP, as previously mentioned, the number of classes can be automatically inferred while the Potts model exploits spatially correlations within each image. Our contribution is twofold: first, we derive a proper prior model combining the HDP and the Potts model; secondly, a Gibbs sampler is designed to explore the parameter space and approximate the Bayesian estimators associated with the resulting posterior distribution.

The paper is organized as follows. In section 2, the prior model is presented. Section 3 details the posterior and the proposed sampling algorithm. Section 4 reports preliminary simulation results obtained on synthetic data. Section 5 gives a synthesis and some perspectives for our model.

Part of this work has been supported by the BNPSI ANR project no ANR-13-BS-03-0006-01.

2. JOINT SEGMENTATION MODELING

2.1. Observation model

Let us define J as the number of images to be jointly segmented. Image j ($j = 1, \ldots, J$) is composed of N_j pixels whose observed gray levels, denoted y_{jn} ($n = 1, \ldots, N_j$), are assumed to be distributed according to a probability distribution $f(\cdot|\cdot)$ parameterized by θ_{jn}

$$y_{jn}|\theta_{jn} \sim f(\cdot|\theta_{jn}).$$

These J images are assumed to be composed of K homogeneous classes. Thus, all pixels (j, n) belonging to the class k are characterized by a common statistical parameter $\theta_{jn} = \phi_k$, with ϕ_k the parameter corresponding to class k. Note that the distribution over the whole image is then a mixture of distributions. In a given image, a region is defined as a set of pixels that own the same region label value; a class k is the union of all the regions assigned the class label k in all images. In the sequel of this paper, the following notations are used

- $c_{jn} = t$: pixel n of image j is in region labeled t
- $d_{jt} = k$: the region labeled t of image j is assigned the class label k
- ψ_{jt}: parameter of the region t in image j
- ν_{jt}: number of pixels in region t in image j
- m_{jk}: number of regions with class label k in image j
- $m_{j\cdot} = \sum_k m_{jk}$: number of regions in image j
- $m_{\cdot k} = \sum_j m_{jk}$: number of regions with class label k
- K: number of classes

2.2. Hierarchical Dirichlet Process

Let us define $\mathbb{G}_j$ as the probability distribution of the parameters θ_{jn}. That probability density $\mathbb{G}_j$ is considered unknown. In a Bayesian setting, it is assigned a prior distribution. Thus, prior distributions on the space of prior distributions are required. Bayesian nonparametrics models were considered as a solution to this issue and more particularly the DP

$$\mathbb{G}_j \sim DP(\alpha_0, \mathbb{G}_0) \quad \text{for} \quad j \in \{1, \ldots, J\}$$
$$\text{with } \mathbb{G}_j = \sum_{t=1}^{\infty} \tau_{jt} \delta_{\psi_{jt}}$$

Within this setting, $\mathbb{G}_j$ is an infinite sum of Dirac measures on ψ_{jt} ($\psi_{jt} \sim \mathbb{G}_0$) with weights τ_{jt}. For classes to be shared among the images, the base distributions should have common atoms ϕ_k. $\mathbb{G}_0$ is then defined as a discrete prior whose masses are concentrated on the values of the atoms. These atoms are independently distributed according to a defined probability measure H with probability density function h, $\phi_k \sim H$; the number of atoms K is supposed unknown. A DP can be chosen as a prior for $\mathbb{G}_0$,

$$\mathbb{G}_0 \sim DP(\gamma, H) \text{ for } \mathbb{G}_0 = \sum_{k=1}^{\infty} \pi_k \, \delta_{\phi_k}$$

This model has been introduced as hierarchical Dirichlet process [8]. To better understand the model, let us introduce a metaphor, namely the Chinese restaurant Process (CRP). The CRP is a description of the DP model based on the Pólya urn properties [9]. The equivalent of this metaphor for the HDP is the Chinese restaurant franchise (CRF). A franchise of J restaurants is considered, serving the same menu with a possibly infinite number of dishes. Here, a customer corresponds to a pixel, a table to a region, a dish to a class and a restaurant to an image. Thus, we have a set of images with a possibly infinite number of classes that can be shared among the images. The CRF is explained directly using the analogy.

Let us consider pixel n in image j. This pixel has a positive probability to be located in an existing region t ($t = 1, \ldots, m_{j\cdot}$), proportional to the number ν_{jt} of pixels in the region or to be assigned to a new one, proportional to α_0. If the pixel is associated to the region t, it will be parameterized by ψ_{jt} of the region; indeed, all pixels in a region are in the same class. If a new region is chosen, then the associated class parameter must be sampled. Here also, an existing parameter ϕ_k can be picked proportionally to the number of regions in the set that are assigned to the class $m_{\cdot k}$ and a new one proportionally to γ. It is written:

$$\theta_{jn}|\boldsymbol{\theta}_j^{-n}, \alpha_0, \mathbb{G}_0 \sim \sum_{t=1}^{m_{j\cdot}} \frac{\nu_{jt}}{N_j - 1 + \alpha_0} \delta_{\psi_{jt}} + \frac{\alpha_0}{N_j - 1 + \alpha_0} \mathbb{G}_0$$

$$\psi_{jt}|\boldsymbol{\psi}^{-jt}, \gamma, H \sim \sum_{k=1}^{K} \frac{m_{\cdot k}}{m_{\cdot\cdot} + \gamma} \delta_{\phi_k} + \frac{\gamma}{m_{\cdot\cdot} + \gamma} H$$

with $\boldsymbol{\theta}_j^{-n} = \{\theta_{jn'} \mid n' = 1, \ldots, N_j, n' \neq n\}$ and $\boldsymbol{\psi}^{-jt} = \{\psi_{j't'} \mid j'=1, \ldots, J; t'=1, \ldots, m_{j'\cdot}; (j', t') \neq (j, t)\}$. It clearly appears that α_0 and γ tune the probability of having a new table and a new dish, respectively. Thus they adjust the average number of tables in the restaurants and the average number of proposed dishes.

2.3. Prior model

The main objective of the proposed strategy consists of segmenting a collection of images. As a consequence, the class parameters can be considered as nuisance parameters and are marginalized out from the joint posterior distribution, i.e., only inference of the label variables $\boldsymbol{c} = \{c_{jn} \mid j = 1, \ldots, J; n = 1, \ldots, N_j\}$ and $\boldsymbol{d} = \{d_{jt} \mid j = 1, \ldots, J; t = 1, \ldots, m_{j\cdot}\}$ is conducted. Furthermore, to promote spatially compact regions within an image, we propose to resort to a Markov random field prior for these labels. The latter can be divided into two factors, the first denoted φ corresponding to the terms induced by unique cliques, where a clique is a set of pixels in the same neighborhood. The second term ρ is defined by the cliques of more than one elements for combining Potts-Markov model and DP [7]

$$Pr(\boldsymbol{c}, \boldsymbol{d}) \propto \varphi(\boldsymbol{c}, \boldsymbol{d}) \rho(\boldsymbol{c}, \boldsymbol{d}) \tag{1}$$

where $\varphi(\boldsymbol{c}, \boldsymbol{d})$ is the factor corresponding to the singletons and is defined as the prior induced by the HDP. In (1), $\rho(\boldsymbol{c}, \boldsymbol{d})$

is set to be a Potts model on the class assignments. Thus,

$$\varphi(\boldsymbol{c},\boldsymbol{d}) = \prod_{j=1}^{J} \left\{ \left[\prod_{n=1}^{N_j} \frac{1}{(\alpha_0+n-1)} \right] \alpha_0^{m_{j\cdot}} \left[\prod_{t=1}^{m_{j\cdot}} \Gamma(\nu_{jt}) \right] \right\}$$
$$\left[\prod_{t=1}^{m_{\cdot\cdot}} \frac{1}{(\gamma+t-1)} \right] \gamma^K \left[\prod_{k=1}^{K} \Gamma(m_{\cdot k}) \right] \tag{2}$$

and

$$\rho(\boldsymbol{c},\boldsymbol{d}) = \prod_{j=1}^{J} \exp\left(\sum_{n\sim q} \beta\, \delta(d_{jc_{jn}}, d_{jc_{jq}}) \right) \tag{3}$$

where $n \sim q$ means that pixel q is a neighbor of pixel n and $\delta(d_{jc_{jn}}, d_{jc_{jq}}) = 1$ if $d_{jc_{jn}} = d_{jc_{jq}}$ and $\delta(d_{jc_{jn}}, d_{jc_{jq}}) = 0$ otherwise. The granularity parameter β tunes the spatial correlation. It is assumed to be fixed in this work, but could be included within the Bayesian model to be estimated jointly with the parameters of interest following, e.g., the strategy proposed in [10].

The observations are assumed to be independent conditionally upon the class they belong to. Thus, the likelihood can be factorized as:

$$f(\boldsymbol{y}|\boldsymbol{c},\boldsymbol{d}) = \prod_{k=1}^{K} f(\boldsymbol{y}_{A_k})$$

where $A_k = \{(j,n)|d_{jc_{jn}} = k\}$ is the set of pixels in class k, $\boldsymbol{y}_{A_k} = \{y_{jn}|(j,n) \in A_k\}$ and

$$f(\boldsymbol{y}_{A_k}) = \int \left[\prod_{(j,n)\in A_k} f(y_{jn}|\phi_k) \right] h(\phi_k) \mathrm{d}\phi_k \tag{4}$$

which can be calculated analytically in the conjugate case that we choose.

3. HIERARCHICAL IMAGE SEGMENTATION

In a Bayesian setting, the estimates of the parameters of interest are computed from the posterior distribution: $Pr(\boldsymbol{c},\boldsymbol{d}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{c},\boldsymbol{d})Pr(\boldsymbol{c},\boldsymbol{d})$. In this work, the latter is not analytically tractable. Classically, we propose to explore it using MCMC methods. The proposed algorithm is a single-site Gibbs sampler [11] which consists in sampling one after the other the unknown variables conditionally upon the observed images and all the other variables. In the considered framework, we are only interested in the segmentation, hence the class parameter vectors are marginalized. It should be noted that they could be easily sampled afterwards conditionally upon the class assignment variables. As for the region and class assignments, they are repeatedly sampled according to a Chinese Restaurant Franchise [8], which is modified to account for the Potts interaction between neighboring pixels by including a term of the form $\exp(\sum_{n\sim q} \beta\, \delta(d_{jc_{jq}}, k))$, see equations (5) to (8). An iteration of the proposed sampler is described on Algo. 1.

Algorithm 1 Gibbs sampler

```
for j = 1, ..., J do
    for n = 1, ..., N_j do
        Sample c_jn ~ p(c_jn | c^{-jn}, d, y)
    end for
    for t = 1, ..., m_j. do
        Sample d_jt ~ p(d_jt | c, d^{-jt}, y)
    end for
end for
```

3.1. Sampling $\boldsymbol{c}$

The conditional distribution of c_{jn} given all the other variables is proportional to the product of the prior over c_{jn} associated to the likelihood of y_{jn}. Due to the exchangeability of the region assignments, we can consider c_{jn} to be the last one sampled. Therefore, either $c_{jn} = t \le m_{j\cdot}$ or $c_{jn} = t^{\text{new}}$. As defined in the previous section, c_{jn} takes value t or t^{new} proportionally to ν_{jt}^{-jn} and α_0, respectively. In the first case, the corresponding likelihood of y_{jn} is $f(y_{jn}\,|\,\boldsymbol{y}_{A_{d_{jt}}^{-jn}})$ where $f(y_{jn}\,|\,\boldsymbol{y}_{A_{d_{jt}}^{-jn}}) = f(y_{jn}, \boldsymbol{y}_{A_{d_{jt}}^{-jn}})/f(\boldsymbol{y}_{A_{d_{jt}}^{-jn}})$. The likelihood of y_{jn} corresponding to $c_{jn} = t^{\text{new}}$, can be deduced integrating out with respect to the different possibilities of $d_{jt^{\text{new}}}$,

$$p(y_{jn}\,|\,c_{jn} = t^{\text{new}}, \boldsymbol{c}^{-jn}, \boldsymbol{d}, \boldsymbol{y}^{-jn}) \tag{5}$$
$$\propto \left\{ \sum_k m_{\cdot k} \exp\left(\sum_{n\sim q} \beta\delta(d_{jc_{jq}}, k) \right) + \gamma \right\}^{-1}$$
$$\left\{ \sum_{k=1}^{K} m_{\cdot k} \exp\left(\sum_{n\sim q} \beta\delta(d_{jc_{jq}}, k) \right) f\left(y_{jn}\,|\,\boldsymbol{y}_{A_k^{-jn}}\right) + \gamma f(y_{jn}) \right\}$$

with $f(y_{jn}) = \int f(y_{jn}|\phi^{\text{new}})h(\phi^{\text{new}})\mathrm{d}\phi^{\text{new}}$. Then,

$$Pr(c_{jn} = t\,|\,\boldsymbol{c}^{-jn}, \boldsymbol{d}, \boldsymbol{y}) \tag{6}$$
$$\propto \begin{cases} \nu_{jt}^{-jn}\, f(y_{jn}\,|\,\boldsymbol{y}_{A_{d_{jt}}^{-jn}}) & \text{if } t \le m_{j\cdot} \\ \alpha_0\, p(y_{jn}\,|\,c_{jn} = t^{\text{new}}, \boldsymbol{c}^{-jn}, \boldsymbol{d}, \boldsymbol{y}^{-jn}) & \text{if } t = t^{\text{new}} \end{cases}$$

If a new region is chosen, the assigned class must be chosen,

$$Pr(d_{jt^{\text{new}}} = k\,|\,\boldsymbol{c}, \boldsymbol{d}^{-jt^{\text{new}}}) \tag{7}$$
$$\propto \begin{cases} m_{\cdot k} \exp\left(\sum_{n\sim q} \beta\, \delta(d_{jc_{jq}}, k) \right) f(y_{jn}\,|\,\boldsymbol{y}_{A_k^{-jn}}) & \text{if } k \le K \\ \gamma\, f(y_{jn}) & \text{if } k = k^{\text{new}} \end{cases}$$

3.2. Sampling $\boldsymbol{d}$

As for the region assignments, the probability of d_{jt} conditionally to all the other variables is proportional to the prior times the likelihood of $\boldsymbol{y}_{jt} = \{y_{jq}\,|\,c_{jq} = t\}$. Moreover, either $d_{jt} = k \le K$ or $d_{jt} = k^{\text{new}}$. Thus, d_{jt} takes the number of an existing class k proportionally to the number of regions that are assigned to class k, $m_{\cdot k}$. It also depends, through the Potts model, on the number of neighboring pixels of the region also labeled k. Conversely, d_{jt} takes a new value proportionally to γ.

$$Pr(d_{jt} = k \,|\, \boldsymbol{c}, \boldsymbol{d}^{-jt}, \boldsymbol{y}) \tag{8}$$
$$\propto \begin{cases} m_{\cdot k}^{-jt} \exp\!\left(\sum\limits_{n \sim q} \beta\, \delta(d_{jc_{jq}}, k) \right) f(\boldsymbol{y}_{jt} \,|\, \boldsymbol{y}_{A_k^{-jt}}) & \text{if } k \leq K \\ \gamma\, f(\boldsymbol{y}_{jt}) & \text{if } k = k^{\text{new}} \end{cases}$$

4. FIRST EXPERIMENTAL RESULTS

The algorithm is tested on three ($J = 3$) piecewise constant images $\boldsymbol{y}_j^\star$ of size 50×50 ($N_j = 2500$, $j = 1, \ldots, J$), shown in Fig. 1. They are composed of squares with gray levels: -50, -25, 25, 50 and the background is set to 0. The observed images $\boldsymbol{y}_j$ are defined as the true ones $\boldsymbol{y}_j^\star$ corrupted by an additive noise: $\boldsymbol{y}_j = \boldsymbol{y}_j^\star + \boldsymbol{\epsilon}_j$. The $\boldsymbol{\epsilon}_j$ are independent and spatially white with $\epsilon_{jn} \sim \mathcal{N}(0, \sigma_y^2)$. The likelihood then reads $f(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{j=1}^{J} \prod_{n=1}^{N_j} \mathcal{N}(y_{jn}; \theta_{jn}, \sigma_y^2)$.

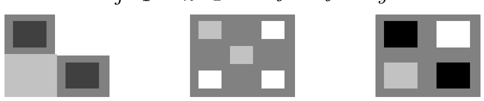

Fig. 1. Synthetic true images

As for the prior model, h is chosen as a Gaussian density: $\mathcal{N}(\mu_H, \sigma_H^2)$ where $\mu_H = 0$ and $\sigma_H = 100$ so that h is conjugate to the likelihood. This leads to the closed-form calculation of the integrated likelihood (4). The scalar hyperparameters are chosen as: $\gamma = 1$ and $\alpha_0 = 1$. The Potts model parameter is set to $\beta = 0.25$.

In addition to the proposed method, two different algorithms have been implemented as a comparison: 1/ the one in [7] which does not account for shared classes, 2/ an algorithm solely based on an HDP prior model, without interactions between the neighboring pixels.

The pixel classes are estimated based on the marginal maximum a posteriori of each pixel. They are numerically computed as the maximizer of the empirical histogram of the simulated $d_{jc_{jn}}$ (except the ones of the burn-in period).

The results of the segmentation are represented in Fig. 2 and 3 for different values of σ_y. For $\sigma_y = 1$, the rate of correct affectation is equal to 100% for the algorithm presented in [7] and for the proposed one. The partition induced using an HDP as prior is also quite similar to the true one. The observed images are not very noisy, that makes the inference easier. While $\sigma_y = 5$, the partition induced for more than 99% of the pixels is good, using the proposed algorithm or the one described in [7]. Here, it can be seen that the method in [7] gives a correct classification for each image but does not allow the pixels sharing same characteristics within the images to be assigned the same class label contrary to the proposed method. Indeed, the background and the squares with same/different gray levels are well recognized but not jointly in the images. If only a classification image-by-image is needed, the method in [7] can be sufficient. On the opposite, our method is well appropriate for a joint classification of the set of images. When only the HDP is put as prior, the number of classes is overestimated since the images are noisy and the neighboring information is not taken into account to favor compact regions. As expected, the proposed algorithm gives a correct joint segmentation of the set of synthetic images.

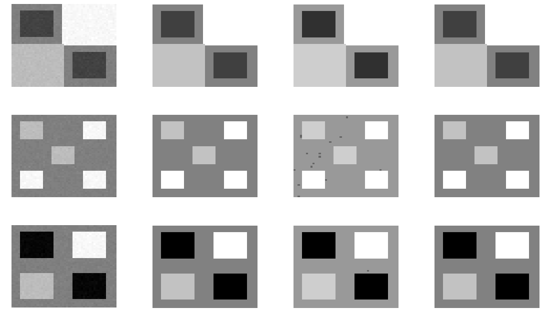

Fig. 2. Segmentation obtained for $\sigma_y = 1$. From left to right: Observed, DP + Potts [7], HDP and HDP + Potts

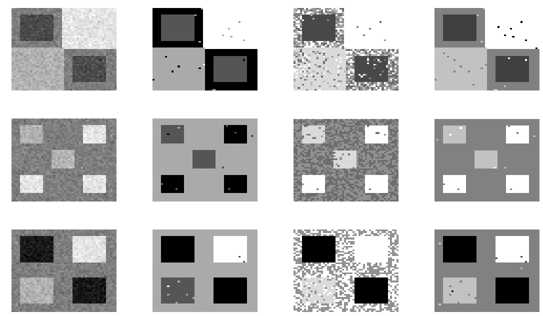

Fig. 3. Segmentation obtained for $\sigma_y = 5$. From left to right: Observed, DP + Potts [7], HDP and HDP + Potts

5. CONCLUSION

A new model as well as a new algorithm have been proposed for the segmentation of a set of images with shared classes based on the hierarchical Dirichlet Processes and the Potts model. A first experimental validation has been done on synthetic images.

Future works will consist of investigating the influence of the various parameters inherent to the model. Images of various types will be also considered, like textured images or images with areas that are not piecewise constant. The model will then be applied on real images, e.g. encountered in remote sensing applications. Using these images, it will be possible to test the algorithm on likewise images presented in the introduction: city, countryside... On the algorithmic side, for a given configuration, the complexity is $O(\prod_{j=1}^{J}[N_j(m_{j\cdot} + K)])$, which strongly depends on the images to be segmented. Thus, we aim at developing a method which allows to update the region label of a set of pixels conjointly, using for example the Swendsen-Wang algorithm [12], as for the Dirichlet Process [13].

6. REFERENCES

[1] D.-S. Lee, "Effective Gaussian mixture learning for video background subtraction," in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, May 2005, pp. 827–832.

[2] P. Szwarc, J. Kawa, M. Rudzki, and E. Pietka, "Automatic brain tumour detection and neovasculature assessment with multiseries MRI analysis," *Computerized Medical Imaging and Graphics*, vol. 46, pp. 178–190, June 2015.

[3] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, pp. 721–741, November 1984.

[4] S. Richardson and P. J. Green, "On Bayesian analysis of mixtures with unknown number of components," *J. Roy. Stat. Soc. Ser. B*, vol. 59, no. 4, pp. 731–792, 1997.

[5] F. Murtagh, A. E. Raftery, and J.-L. Starck, "Bayesian inference for multiband image segmentation via moel-based cluster trees," in *Image and Vision Computing*, vol. 23, June 2005, pp. 587–596.

[6] E. B. Sudderth and M. I. Jordan, "Shared segmentation of natural scenes using dependent Pitman-Yor processes," in *Advances in Neural Information Processing Systems 21*, vol. 1, 2008, pp. 1585–1592.

[7] P. Orbanz and J. M. Buhmann, "Nonparametric Bayesian image segmentation," in *Int. J. Computer Vision*, no. 77, 2007, pp. 25–45.

[8] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical Dirichlet Processes," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1566–1581, December 2006.

[9] D. Blackwell and J. B. MacQueen, "Ferguson distributions via Pólya urn schemes," *The Annals of Statistics*, vol. 1, no. 2, pp. 353–355, 1973.

[10] M. Pereyra, N. Dobigeon, H. Batatia, and J.-Y. Tourneret, "Estimating the granularity coefficient of a Potts-Markov random field within an MCMC algorithm," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2385–2397, June 2013.

[11] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York, NY, USA: Springer, 2004.

[12] A. Barbu and S.-C. Zhu, "Generalizing Swendsen-Wang to sampling arbitrary posterior probabilities," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1239–1253, August 2005.

[13] R. Xu, F. Caron, and A. Doucet, "Bayesian nonparametric image segmentation using a generalized Swendsen-Wang algorithm," *Arxiv*, no. 1602.03048.